



# Bootstrap bei Faktorenanalyse

P30.6

Kurt Holm

Almo Statistik-System

[www.almo-statistik.de](http://www.almo-statistik.de)

[holm@almo-statistik.de](mailto:holm@almo-statistik.de)

[kurt.holm@jku.at](mailto:kurt.holm@jku.at)

2022

## Weitere Almo-Dokumente

Die folgenden Dokumente können alle von der Handbuchseite in <http://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)  
Diskriminanzanalyse.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
- 9b. Bootstrap bei Logit- und Probitanalyse.pdf
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 III.PDF
14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer, Cox-Regression.PDF (1,5 MB)
15. Faktorenanalyse.PDF (1,6 MB)
- 15a Bootstrap bei Faktorenanalyse.PDF
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 Mehrfachantworten
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

<b>P30.6.0 Was mit Bootstrap möglich wird .....</b>	<b>4</b>
<b>P30.6.1 Vorgehensweise beim Bootstrap .....</b>	<b>4</b>
P30.6.1.1 Standardfehler .....	5
P30.6.1.2 Konfidenzintervall, Perzentil-Verfahren .....	5
P30.6.1.3 Signifikanz, p-Wert .....	5
P30.6.1.4 Mittelwert und Verzerrung.....	7
P30.6.1.5 Der Forschungsbericht .....	7
P30.6.1.6 Die Zahl der Bootstrap-Stichproben .....	7
P30.6.1.7 Daten lesen und verarbeiten beim Bootstrap.....	8
<b>P30.6.2 Vorgehensweise beim Bootstrap der Faktorenanalyse .....</b>	<b>13</b>
P30.6.2.0 Sonderfall: Faktorenanalyse mit 1 Faktor .....	13
P30.6.2.0.1 Vorzeichen-Umkehr bei Analyse mit 1 Faktor .....	14
P30.6.2.1 Bootstrap der Eigenwerte .....	21
P30.6.2.2 Unsere Beispieldaten: Die Holzinger-Swineford-Daten .....	23
P30.6.2.3 Rotation und "gemeinsamer Faktoren-Raum" .....	24
P30.6.2.4 Zielmatrix mit rechtwinkligen Koordinatenachsen.....	31
P30.6.2.5 Zielmatrix mit schiefwinkligen Koordinatenachsen .....	35
P30.6.2.6 Gesamte Varimax- bzw. Quartimin-Faktorladungsmatrix als Zielmatrix ..	36
P30.6.2.7 Normierung der Faktorladungen .....	37
<b>P30.6.3 Bootstrap-Programm Prog30ml.Msk .....</b>	<b>39</b>
P30.6.3.1 Zu faktorisierende Variable.....	42
P30.6.3.2 Spezielle Kein-Wert-Behandlung .....	43
P30.6.3.3 Optionen für Faktorenanalyse .....	44
P30.6.3.4 Faktorenanalytisches Modell und Eigenwert-Verfahren.....	45
P30.6.3.5 Faktorenzahl und Kommunalitätenschätzung .....	46
P30.6.3.6 Distanzmatrix ermitteln und Zwischenergebnisse ausgeben .....	50
P30.6.3.7 Das "Aussehen" der Ergebnisse .....	51
P30.6.3.8 Die Bootstrap-Optionsbox.....	52
<b>P30.6.4 Bootstrap-Ergebnisse aus Programm Prog30ml.Msk.....</b>	<b>60</b>
P30.6.4.1 Ergebnisse aus Originalstichprobe bei rechtwinkliger Zielmatrix .....	60
P30.6.4.2 Ergebnisse aus Bootstrap bei rechtwinkliger Zielmatrix .....	61
P30.6.4.2.1 Bootstrap der Eigenwerte.....	61
P30.6.4.2.2 Bootsrap der Faktorladungen .....	62
P30.6.4.2.3 Bootstrap der reproduzierten Kommunalitäten.....	65
P30.6.4.3 Ergebnisse aus Bootstrap bei schiefwinkliger Zielmatrix.....	68
P30.6.4.4 Ergebnisse aus Bootstrap bei Faktorenanalyse mit 1 Faktor.....	70
<b>Anhang: Normierung der Faktorladungen</b>	
<b>Vergleich der Studie von Zientek &amp; Thompson mit Almo .....</b>	<b>71</b>
<b>Literatur zu Bootstrap.....</b>	<b>80</b>

## P30.6 Bootstrap bei Faktorenanalyse

Die Faktorenanalyse wird im Almo-Dokument 15 „Faktorenanalyse“ ausführlich dargestellt. Auf das Bootstrap-Verfahren wird dabei nicht eingegangen. Das soll hier nachgeholt werden.

Das Bootstrap-Verfahrens wurde bereits detailliert im Almo-Dokument 13b „Bootstrap beim *Allgemeinen Linearen Modell*“ erläutert. Wir werden deswegen hier die Vorgehensweise und den Kalkül beim Bootstrapping nicht in voller Ausführlichkeit darstellen und dabei immer wieder auf das Almo-Dokument 13b verweisen, jedoch ausführlich die Besonderheiten des Bootstrappings der Faktorenanalyse und die Eingabemasken für das Bootstrap-Programm „Prog30ml.Msk“ für die Faktorenanalyse erläutern.

### P30.6.0 Was mit Bootstrap möglich wird

Die Faktorenanalyse ist ein mathematisches, auf dem Eigenwert-Eigenvektor-Kalkül beruhendes Verfahren. Das bedeutet, dass der Forscher keine Antwort auf diese zentrale Frage erhält

**Ist die für die Variable  $i$  auf dem Faktor  $j$  errechnete Faktorladung  $f(ij)$  signifikant von 0 verschieden ?**

Eine gewisse Ausnahme bildet die Maximum-Likelihood-Faktorenanalyse. Sie kann als statistisches Verfahren bezeichnet werden. Aus ihrem Rechengang ergibt sich ein Chi-Quadrat-Wert, der pauschal die Güte ausdrückt, mit der die gewonnene Faktorladungsmatrix an die Korrelationsmatrix angepasst ist. Die oben formulierte zentrale Frage, kann sie jedoch nicht beantworten.

Durch Bootstrap kann sie beantwortet werden. Aus dem Bootstrap-Verfahren erhält der Forscher für jede einzelne Faktorladung (1) deren Standardfehler, (2) ihre Signifikanz ( $p$ -Wert) und (3) das Konfidenzintervall - ohne dass dafür eine Verteilungsannahme (z.B. Normalverteilung) getroffen werden muss.

### P30.6.1 Vorgehensweise beim Bootstrap

Aus einer vorliegenden Stichprobe (wir nennen sie "Originalstichprobe") der Größe  $n$  werden zufällig  $n$  (also gleich viele) Datensätze mit *Zurücklegen* ausgewählt. Dadurch entsteht die Bootstrap-Stichprobe Nr. 1. Das Zurücklegen bewirkt, dass manche Datensätze mehrfach ausgewählt werden und dass manche Datensätze der originalen Stichprobe nicht in die Bootstrap-Stichprobe geraten. Auf diese Weise werden viele, etwa 1000 Bootstrap-Stichproben erzeugt. Die Analogie ist offenkundig: Die Originalstichprobe entspricht dem "Universum" der Daten, die Bootstrapstichproben entsprechen den Stichproben, die aus dem Universum gezogen werden. Die theoretische Begründung des Bootstrappings wird bei Efron/ Tibshirani (1994) und relativ übersichtlich bei Sue Man Fan (2010, S. 30-39) ausgeführt.

Für alle (z.B. 1000) Stichproben werden die Ergebnisse errechnet. In Almo werden zuerst die Ergebnisse für die Original-Stichprobe ausgegeben, danach die aus allen Bootstrap-

stichproben zusammengefassten Ergebnisse. Das wird noch detailliert gezeigt. Besonders bedeutsam ist, dass aus dem Bootstrapping *empirische* Verteilungen für die verschiedenen Koeffizienten gewonnen werden. Dadurch ist es möglich, *Standardfehler*, *Signifikanzen* (*p*-Werte) und *Konfidenzintervalle* für diese Koeffizienten zu ermitteln, die keine Verteilungsannahmen erfordern. Das ist der primäre Zweck des Bootstrap-Verfahrens. Es erzeugt "verteilungsfreie" Schätzer und löst somit manches statistische Problem.

### ***P30.6.1.1 Standardfehler***

Betrachten wir als Beispiel den Regressionskoeffizienten  $\beta_1$  für eine unabhängige quantitative Variable  $x_1$  gegenüber der abhängigen Variablen  $y$ . Aus den 1000 Bootstrap-Stichproben erhalten wir 1000 Werte für  $\beta_1$ . Wir berechnen deren Mittelwert und ihre Standardabweichung. Die Standardabweichung wird dann als der "Standardfehler" des Regressionskoeffizienten  $\beta_1$  aus der Originalstichprobe interpretiert.

### ***P30.6.1.2 Konfidenzintervall, Perzentil-Verfahren***

Die obere und untere Grenze des Konfidenzintervalls für beispielsweise ein Konfidenzniveau von 95% erhalten wir sehr einfach in folgender Weise: Die 1000  $\beta_1$ -Werte werden der Größe nach (aufsteigend) sortiert. Vom maximalen  $\beta_1$ -Wert werden absteigend 2,5% von 1000 also 25 Werte heruntergezählt. Der dort in Position 975 stehende  $\beta_1$ -Wert ist die obere Intervallgrenze. Entsprechend wird vom minimalen Wert ausgehend 25 Werte hinaufgezählt. So wird der untere Grenzwert gefunden. Zwischen den beiden Grenzwerten befinden sich dann 95% aller Werte und außerhalb der Grenzwerte 5% aller Werte. Damit ist das 95%-Konfidenzintervall für den Koeffizienten  $\beta_1$  aus der Originalstichprobe gefunden.

Diese sehr einfache Berechnungsweise wird als "Perzentil-Verfahren" bezeichnet. Als alternatives Verfahren wird das PCA-Verfahren empfohlen, das versucht die "Verzerrung" (siehe nachfolgend) und eine eventuelle Schiefe der Verteilung der 1000  $\beta_1$ -Werte auszugleichen. Es gibt noch weitere Verfahren. Einen knappen Überblick findet man im englischen Wikipedia. Almo verwendet das beschriebene "einfache Perzentil-Verfahren" und optional das asymmetrische und symmetrische Perzentil-t -Verfahren. Bei den beiden Perzentil-t -Verfahren wird in gewisser Weise auch die "Verzerrung" ausgeglichen. Siehe dazu das Almo-Dokument 13b „Bootstrap beim *Allgemeinen Linearen Modell*“, Abschnitt P20.25.6. Das Perzentil-t -Verfahren benötigt allerdings einen Schätzwert für den Standardfehler des Koeffizienten  $\beta_1$  für die Originalstichprobe und jede einzelne Bootstrapstichprobe. Bei der Faktorenanalyse steht dieser Schätzwert nicht zur Verfügung, so dass nur das "einfache Perzentil-Verfahren" einsetzbar ist.

### ***P30.6.1.3 Signifikanz, p-Wert***

Die zweiseitige Signifikanz (der *p*-Wert) ergibt sich beim einfachen Perzentil-Verfahren in folgender Weise: Almo ermittelt zuerst das Konfidenzniveau, bei dem sich der Wert 0 aus den 1000 aufsteigend sortierten Bootstrap-Koeffizienten  $\beta_1$  gerade noch außerhalb des Konfidenzintervalls befindet, d.h. bei dem 0 direkt unterhalb der unteren Konfidenzgrenze oder direkt oberhalb der oberen Konfidenzgrenze plaziert ist. Dieses Konfidenzintervall wird in Almo auch als "optimales Konfidenzintervall" bezeichnet. Es kann so interpretiert werden: "Gerade noch" mit der Wahrscheinlichkeit (der Sicherheit) von  $k$  Prozent, die für das "optimale" Konfidenzintervall gefunden wurde, ist der Bootstrap-Koeffizient  $\beta_1$  von 0 verschieden. Die Gegenwahrscheinlichkeit, die Irrtumswahrscheinlichkeit, der *p*-Wert ist dann  $p = 1 - k/100$ .

Der aus dem Perzentil-Verfahren gewonnene *p*-Wert ist dann so zu interpretieren: Er drückt die Wahrscheinlichkeit aus, mit welcher der Bootstrap-Koeffizient  $\beta_1$  gleich 0 sein könnte.

Möglich wäre es auch als "Testwert" anstelle von 0 einen anderen Wert zu verwenden. Almo setzte beispielsweise beim Bootstrap der Eigenwerte als Testwert den Wert 1.0 ein, um den p-Wert für die Eigenwerte zu berechnen. Dieser Testwert wird in den aufsteigend sortierten Koeffizienten gesucht. Der Koeffizient über bzw. unter ihm ist dann der obere bzw. untere Grenzwert des "optimalen Konfidenzintervalls". Diese Vorgehensweise ist allerdings nur zulässig, wenn 1.0 als Kommunalität eingesetzt wurde (Kaiser-Kriterium).

Beim Perzentil-t -Verfahren (das für die Faktorenanalyse nicht einsetzbar ist) ist die Berechnung von p komplexer. Wie der p-Wert beim einfachen Perzentil-Verfahren und beim komplexeren Perzentil-t -Verfahren gewonnen wird, wird im Almo-Dokument 13b „Bootstrap beim Allgemeinen Linearen Modell“, Abschnitt P20.25.5 und P20.25.6 ausführlich und an Hand eines Beispiels beschrieben. Wir werden nachfolgend dieses Beispiel verkürzt vortragen. Die p-Werte aus dem einfachen Perzentil- und dem Perzentil-t -Verfahren sind nicht gleich, in der Regel aber erst an der 3. Kommastelle verschieden.

Wie aber soll verfahren werden, wenn die 1000 aufsteigend sortierten Werte von  $\beta_1$  keinen Wert 0 aufweisen? Der Fall tritt vor allem dann auf, wenn  $\beta_1$  (absolut) sehr groß ist. In dieser Situation muss der ungünstigste Fall unterstellt werden, dass gerade unterhalb bzw. oberhalb der aufsteigend sortierten Werte, der Wert 0 folgen würde - hätte man eine weitere Stichprobe gerechnet. Almo berechnet somit das "optimale" Konfidenzniveau für ein Konfidenzintervall, dessen unterer Grenzwert der erste bzw. niedrigste Wert in der Sortierfolge ist und dessen oberer Grenzwert der letzte bzw. höchste Wert ist. Die zweiseitige Signifikanz ist dann sehr einfach  $p=1/(Stichprobenzahl+1)$ . Die tatsächliche Signifikanz der Variablen kann dann nur gleich diesem p-Wert oder besser (d.h. kleiner) sein. Sie ist nur durch die Stichprobenzahl bestimmt.

Der p-Wert aus dem einfachen Perzentil-Verfahren ist nur dann berechenbar, wenn der dem Bootstrapping unterworfenen Koeffizient positive *und* negative Werte annehmen kann. Das ist der Fall beim Regressionskoeffizient  $\beta_1$ , dem Haupt- und Interaktions-Effekt beim ALM, dem Korrelationskoeffizienten und der Faktorladung - aber z.B. nicht beim multiplen Korrelationskoeffizient, der nur  $\geq 0$  sein kann.

*Ein Beispiel*

Betrachten wir ein Beispiel, das wir aus dem Almo-Handbuch 13b "Bootstrap bei Allgemeinen Linearen Modell", Abschnitt P20.25.5.1 entnommen haben. Aus einer Regressionsanalyse mit 10 000 Bootstrap-Stichproben werden diese 10 000  $\beta_1$ -Koeffizienten gefunden. Sie sind aufsteigend sortiert. Das Beispiel ist problemlos übertragbar auf die Faktorenanalyse und die Faktorladung als Bootstrap-Koeffizient.

Bootsrap Stichproben aufsteigend sortiert	β1-Koeffizient aus 10 000 Bootstrap-Stichproben
1	-0.367582
2	-0.267036
.	.
.	.
208	-0.000266097
209	-0.000042759
----- <--- Wert 0.0	
210	0.000173536 <--- untere "optimale" Konfidenzgrenze
211	0.000316841 <--- bei "optimalen" Konf.niveau 0.9582
.	.
.	.
249	0.012127
250	0.0126888
-----	
untere Konf.grenze ---> 251	0.013056
bei Konf.niv. 0.95	.

	.	.	
	4992	0.371055	
	4993	0.371063	
	4994	0.371097	
	.	.	
	.	.	
obere Konf.grenze	9749	0.720293	
bei Konf.niv. 0.95 ---->	9750	0.720791	
-----			
	9751	0.720834	
	.	.	
	.	.	
	9791	0.73489	obere "optimale" Konfidenzgrenze
			<--- bei "optimalen" Konf.niveau 0.9582
-----			
	9792	0.73550	
	.	.	
	.	.	
	9999	1.00896	
	10000	1.06902	

Bei einem vorgegebenen Konfidenzniveau von 95% liegen bei 10 000 Bootstrap-Stichproben 250 Stichproben unterhalb und 250 oberhalb des Konfidenzintervalls. Die Konfidenzgrenzen in unserem Beispiel sind die  $\beta_1$ -Werte **0.013056** und **0.720791**. Unterhalb von  $\beta_1=0$  liegen in unserem Beispiel 209 Werte. Dem entspricht ein zweiseitiger p-Wert von  $p=(209+209)/10000=0.0418$  und entsprechend ein "optimales" Konfidenzniveau von  $\alpha_{Kn}=(10000-418)/10000=0.9582$

#### P30.6.1.4 Mittelwert und Verzerrung

Wird der Mittelwert aus den  $\beta_1$ -Werten der Bootstrapstichproben mit dem  $\beta_1$ -Wert aus der Original-Stichprobe verglichen, so muss man in aller Regel eine kleine "Verzerrung" zur Kenntnis nehmen. In unserem Beispiel ist der  $\beta_1$ -Wert aus der Originalstichprobe **0.369082**. Das arithmetische Mittel aus dem Bootstrapping für  $\beta_1$  ist **0.371074**. Die Verzerrung ist dann  $0.371074-0.369082=0.001992$ , also vernachlässigbar gering. Das ist die Regel. Beim bereits erwähnten PCA-Verfahren wird versucht, diese Verzerrung zu korrigieren. SPSS verwendet dieses Verfahren zusätzlich zum Perzentil-t -Verfahren (das standardmäßig verwendet wird). Allerdings wird es im SPSS-ALM nur für die "Parameter" eingesetzt. Die Unterschiede gegenüber dem Perzentil-t -Verfahren sind minimal.

Beim auch in Almo vorhandenen Perzentil-t -Verfahren wird in gewisser Weise auch die "Verzerrung" ausgeglichen. Siehe dazu Almo-Dokument 13b „Bootstrap beim Allgemeinen Linearen Modell“, Abschnitt P20.25.6.

#### P30.6.1.5 Der Forschungsbericht

Der Mittelwert hat sonst keine Bedeutung. Als bester Schätzer für  $\beta_1$  wird **nicht** der Mittelwert aus den Bootstrapstichproben verwendet. Der  $\beta_1$ -Wert aus der Original-Stichprobe wird für den Forschungsbericht verwendet. Als seinen Standardfehler wird die aus dem Bootstrap gewonnene Standardabweichung eingesetzt, als seine Signifikanz p und sein Konfidenzintervall wird die aus dem Perzentil-Verfahren errechneten Werte eingesetzt. Alle diese Koeffizienten sind "parameterfrei".

#### P30.6.1.6 Die Zahl der Bootstrap-Stichproben

Wie wirkt sich eine Erhöhung der Stichprobenzahl auf den Kalkül für die Signifikanz (p-Wert) und das Konfidenzintervall aus. Wir vergleichen die aufsteigend sortierten Effekte des Regressionskoeffizienten  $\beta_1$  aus 1000 und 10 000 Stichproben

1000 Stichproben		10 000 Stichproben	
=====		=====	
1	-0.2598930	1	-0.3675820
2	-0.1551420	2	-0.2670360
.	.	.	.

23	-0.0010547	209	-0.0000428	
-----		-----		<--- 0.0
24	0.0033431	210	0.0001735	
		.	.	
25	0.0052912	250	0.0126888	
-----		-----		
26	0.0134753	251	0.0130560	<--- untere Konfidenzgrenze
.	.	.	.	
1000	0.7319800	10000	0.7359790	

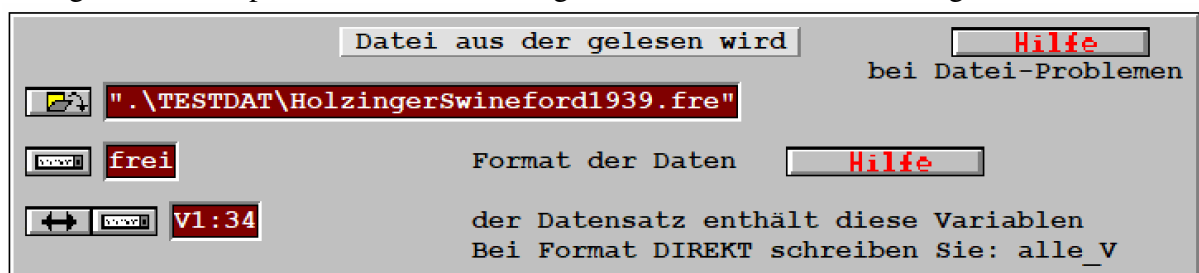
Die Werte aus den beiden Analysen überstreichen einen Bereich von grob -0.26 bis 0.73. Bei 10 000 Stichproben kann der 1. Wert mit  $-0.3675820$  als Ausreisser betrachtet werden. Bei der Analyse mit 1000 Stichproben ist dieser Bereich durch 1000 Werte unterteilt, bei der Analyse mit 10 000 Stichproben durch 10 000 Werte also "feiner". Der Wert 0 wird bei 1000 Stichproben zwischen Position 23 und 24 überschritten. Dem würde bei 10 000 Stichproben die Positionen von 230 bis 240 entsprechen. Tatsächlich wird er jedoch zwischen Position 209 und 210 überschritten. Durch die Position von 0 wird das "optimale" Konfidenzniveau und die Signifikanz p festgelegt. Die Signifikanz beträgt für 1000 Stichproben 0.046 und für 10000 Stichproben 0.042. Der Unterschied von 0.004 ist allerdings vernachlässigbar. Aber wir können behaupten, er ist messgenauer. Also gilt:  
*Je mehr Bootstrapstichproben umso besser. Man sollte mit 10 000 Bootstrapstichproben rechnen, wenn man die Rechenzeit noch als akzeptable empfindet.*

### P30.6.1.7 Daten lesen und verarbeiten beim Bootstrap

Dieser Abschnitt beschreibt, wie in Almo die Daten für das Bootstrapping aufbereitet werden. Der weniger "programmier-technisch" interessierte Leser kann diesen Abschnitt übergehen.

Das Einlesen der Daten und ihre Verarbeitung im Almo-Programm Prog30ml zum Bootstrap der Faktorenanalyse geschieht in folgenden Schritten.

Die Daten der *Original-Stichprobe* werden - ein Datensatz nach dem anderen - eingelesen. Dies geschieht entsprechend der Anweisungen in der Box "Datei, aus der gelesen wird".



In unserem Beispiel in Prog30ml werden die Daten einer Studie zur Begabung von Schülern von Holzinger/Swineford eingelesen. Siehe dazu nachfolgenden Abschnitt P30.6.2.1.

### Datenmanipulationen in der Originalstichprobe

Jeder für die Originalstichprobe eingelesene Datensatz wird den Manipulationen unterworfen, die in den folgenden 4 Optionsboxen angefordert werden. Wenn der Benutzer diese Boxen nicht öffnet, dann bleiben die Daten so wie sie eingelesen wurden.

	Option: Ein- und Ausschliessen von Untersuchungseinheiten	Box 1
	Option: Umkodierungen und Kein-Wert-Angaben	Box 2
	Option: Spezielle Kein-Wert-Behandlung	Box 3
	Option: Untersuchungseinheiten gewichten	Box 4

Die Optionsbox 3 "Spezielle Kein-Wert-Behandlung" spielt eine Sonderrolle. Wir werden weiter unten im Unter-Abschnitt "Spezielle Kein-Wert-Behandlung" und auch in P30.6.3.2 noch darauf eingehen.

Im Detail geschieht folgendes

1. Der gesamte Datensatz des Probanden wird eingelesen. In unserem Beispiel sind das 34 Variablenwerte.
2. Entsprechend der Anweisungen in der 1. Box wird überprüft ob, der Proband in die Analyse aufgenommen wird oder ob er ausgeschlossen werden muss (und somit der nächste Datensatz einzulesen ist)
3. Die Analysevariablen (die zu faktorisierten Variablen) werden herausgegriffen. Dadurch entsteht ein auf die Analysevariablen *reduzierter* Datensatz. Er besteht in unserem 2-faktoriellen Beispiel aus 6 Variablenwerten, den Variablen V14,15,17, 22,23,25.
4. Die Umkodierungs-Anweisungen aus der 2. Box und die Kein-Wert-Angaben werden ausgeführt.
5. Die Gewichtungvariable wird entsprechend der Anweisungen in Box 4 gebildet. Beispiel:  
Die Anweisung lautet "**Wenn Alter >50 dann Gewicht1=1.5; EndeWenn**". Die Gewichtungvariable **Gewicht1** wird in den *reduzierten* Datensatz des Probanden als 7. Variable noch aufgenommen. Der reduzierte Datensatz in unserem Beispiel umfasst nunmehr 7 Werte.

Mit diesen 5 Schritten entsteht ein auf die Analysevariablen *reduzierter* Datensatz mit angehängter Gewichtungvariable, der gemäß der Anweisungen in Box 1, 2 und 4 verändert ist.

Betrachten wir ein weiteres kürzeres Beispiel mit einem von uns konstruierten Datensatz von 13 Variablen und 5 und mehr Probanden. Die Daten seien folgende:

v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13
			Alter					x1	x2	x3	x4	x5
1	2	7	66	1	157	13.0833	2	8	31	12	0	40
2	1	7	73	7	163	13.5833	2	32	21	12	17	34
3	1	7	26	1	157	13.0833	4	27	21	12	15	20
4	2	7	38	2	158	13.1667	2	32	31	16	24	42
5	1	7	29	2	146	12.1667	2	29	19	12	7	37
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.

Ein vollständiger Datensatz eines Probanden umfasst die 13 Variable V1 bis V13. Zum besseren Verständnis haben wir der Variablen V4 den Namen "Alter" und den zu

faktorisierenden Variablen V9 bis 13 die Namen x1 bis x5 gegeben. Der auf die Analysevariable reduzierte Datensatz umfasst also diese 5 Variable plus eine GewichtungsvARIABLE.

Die Ein- bzw. Ausschluss-Anweisung fordert, dass ein Datensatz ausgeschlossen wird, wenn V5 größer 5 ist. Es muss also der 2. Datensatz ausgeschlossen werden

Die Kein-Wert-Deklaration für x4 sei **x4 (0=KeinWert)**.

Im ersten Datensatz wird die 0 in **x4** auf "kw" umkodiert

Das ist eine riesige negative Zahl, die von Almo als Kein-Wert-Code interpretiert wird und durch "kw" symbolisiert wird

Die Unkodierungs-Anweisung sei: Werte von x1 kleinergleich 9 werden auf 10 gesetzt

**x1(1 bis 9 = 10)**. Im ersten Datensatz wird also x1=8 auf 10 gesetzt

Durch die Gewichtungs-Anweisung sollen die unterrepräsentierten Alten mit 1.5 gewichtet werden, alle Jüngeren mit Gewicht1=1

**Wenn Alter >50 dann Gewicht1=1.5; EndeWenn**

Der Proband im 1. Datensatz ist 66. Er erhält also Gewicht1=1.5

Der reduzierte, um die GewichtungsvARIABLE erweiterte Datensatz des Probanden 1 ist dann:

x1	x2	x3	x4	x5	G1
10	31	12	kw	40	1.5

Beachte: x4 ist =kw. Der 2. Datensatz wird ausgeschlossen. Der 3. ist

27	21	12	15	20	1.0
----	----	----	----	----	-----

usw.

Mit dem so entstandenen *reduzierten* Datensatz werden nun zwei Operationen durchgeführt:

1. Es wird die Faktorenanalyse für die Originalstichprobe gerechnet und
2. es wird eine interne Datenmatrix aufgebaut, aus der die Daten für die Bootstrapstichproben entnommen werden

Der jeweils eingelesene i-te *reduzierte* Datensatz wird in eine interne rechteckige Matrix als Zeile i gespeichert. Die Zeilen dieser Datenmatrix werden von den Probanden gebildet und die Spalten von den Analysevariablen plus der GewichtungsvARIABLEN. Aus der Datenmatrix werden dann, wie oben in P30.6.1 beschrieben, durch einen Zufallsgenerator die Probanden für die Bootstrapstichproben gezogen. Die Originaldaten werden nicht mehr verwendet.

In Almo wird die Datenmatrix gegenwärtig im Arbeitsspeicher des Rechners angelegt. Damit wird eine maximale Rechengeschwindigkeit ermöglicht, die es erlaubt mehrere 10 000 Bootstrapstichproben in einer angemessenen Zeit zu rechnen.

Es scheint die Gefahr zu bestehen, dass eine sehr große Datenmatrix im Arbeitsspeicher des Rechners keinen Platz mehr findet. Ein solches Problem wird so gut wie nie auftreten. Man beachte, dass nicht der gesamte Datensatz eines Probanden, sondern nur der auf die

zu faktorisierenden Variablen plus GewichtungsvARIABLE *reduzierte* Datensatz in die Datenmatrix eingeschrieben wird. Betrachten wir ein Beispiel für eine riesige Datenmenge: Eine Datenmatrix mit 100.000 Probanden und 25 zu faktorisierenden Variablen besitzt 2.500.000 Zellen, die ca. 19 MB Speicher belegen. Das ist nur ein Bruchteil des verfügbaren Arbeitsspeichers eines Standard-Computers. 1 Million Probanden werden ungefähr 1900 MB (1,9 GB) Speicherplatz benötigen. Das ist zwar sehr viel, ist aber für ein großzügig ausgestatteten Rechner noch kein Problem.

### **Datenmanipulationen in den Bootstrapsstichproben, das "Analogie-Modell"**

Das Analogie-Modells des Bootstrappings lautet: Die Originalstichprobe entspricht dem "Universum" der Daten. Die Bootstrapsstichproben entsprechen den Stichproben, die aus dem Universum gezogen werden. Diesem Analogie-Modell entspricht es, dass die Daten in den Bootstrapsstichproben nicht durch Datenmanipulationen nachträglich verändert werden dürfen. In ihnen sind nur Daten enthalten, die aus der Originalstichprobe entnommen wurden. Nur die Daten der Originalstichprobe können durch Umkodierungen, Kein-Wert-Behandlungen und Ein- und Ausschluss-Bedingungen bearbeitet werden. Also folgt weitgehend diesem Modell. Im nachfolgendem Unterabschnitt wird gezeigt, was diese Entscheidung für die Kein-Wert-Behandlung bedeutet und dass für das Umkodieren eine Ausnahme gemacht werden muss.

### **Die zwangsweise Umkodierung durch Almo**

Umkodierungen werden nur in der Originalstichprobe vorgenommen. Die umkodierten Daten werden dann von Almo in die Datenmatrix gegeben, aus der die Bootstrapsstichproben zufällig (mit Zurücklegen) entnommen werden. In den Bootstrapsstichproben wird dann nicht mehr umkodiert. Das entspricht, wie oben ausgeführt dem Analogie-Modell und spart außerdem Rechenzeit. Ein unangenehmes Problem gibt es jedoch. Wenn Variable vom Benutzer mit Werten kodiert wurden, die Almo in den verschiedenen Verfahren nicht so erwartet, dann werden diese Werte zwangsweise so umkodiert, wie Almo sie benötigt. Diese zwangsweise Umkodierung wird in der Regel detailliert dem Benutzer von Almo gemeldet. Überwiegend geht es dabei um die Kodierung nominaler und ordinaler Variablen. Almo erwartet (mit wenigen Ausnahmen), dass nominale und auch ordinale Variable ganzzahlig mit 1 beginnen und mit Schrittweite 1 aufsteigend kodiert sind. Durch die zwangsweise Umkodierung wird so ein korrekter, absturzfrierer Rechengang ermöglicht. Beim Bootstrap funktioniert das häufig nicht. Almo kann in den Bootstrapsstichproben nicht zwangsweise umkodieren. Daraus folgt, dass der Benutzer selbst umkodieren muss. Betrachten wir ein Beispiel.

Die Variable "Beruf" wurde durch Qualifikationspunkte in den Daten so kodiert

Beruf	Codeziffer
-----	-----
Arbeiter	10.5
Angestellter	12
Beamter	13.99
Bauer	8.1
.	.
.	.
.	.

Wurde die Variable als nominale oder ordinale durch den Benutzer deklariert, dann muss er in folgender Weise umkodieren

alte	umkodierte
Codeziffer	Codeziffer

-----	-----
8.1	1
10.5	2
12	3
13.99	4
.	.
.	.
.	.

Es wird bei 1 begonnen und mit Schrittweite 1 aufsteigend umkodiert. In der Options-Box "Kein-Wert-Angabe und Umkodierungen" muss diese Anweisung geschrieben werden

**Beruf (8.1=1; 10.5=2; 12=3; 13.9=4)**

Wird kein Bootstrap angefordert, dann nimmt Almo diese Umkodierung selbst vor. Der Benutzer kann sich die Umkodierung ersparen. Beim Bootstrap jedoch würde das Programm abstürzen.

### **Optionsbox "Spezielle Kein-Wert-Behandlung"**

Diese Optionsbox wird ausführlich in Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.3.2. erläutert.

In unserem kurzen Beispiel in Abschnitt P30.6.1.1 wurde im 1. Datensatz die Variable x4 =0 auf "kw" (Almo-intern: eine riesige negative Zahl) umkodiert. In die interne Datenmatrix, wird somit "kw" eingesetzt - aber nur dann, wenn die Kein-Wert-Behandlung 1 oder 2 oder 3 vom Benutzer eingestellt wurden - nicht bei den KW-Behandlungen 4 bis 7.

KW-Behandlung 1 ist das "paarweise Ausscheiden". Es ist in Almo voreingestellt. D.h. bleibt die Optionsbox geschlossen, dann ist KW-Behandlung1 aktiv. KW-Behandlung 2 ist eine Variante des "paarweisen Ausscheidens". Die KW-Behandlung 3 ist das "vollständige Ausscheiden". In diesen drei KW-Behandlungen wird also festgelegt, wie Almo in der Originalstichprobe und in jeder Bootstraptstichprobe im Rechenverlauf verfahren soll, wenn es auf den Wert "kw" trifft. Die Art und Weise, wie mit "kw" verfahren wird, ist in Originalstichprobe und Bootstraptstichprobe dann gleich.

Anders ist es, wenn der Benutzer die KW-Behandlungen 4 bis 7 vorschreibt. Dies sind Varianten der Mittelwert-Einsetzung. Besitzt eine Variable in der Originalstichprobe den Wert "kw", dann wird "kw" durch den Variablenmittelwert ersetzt - und der Datensatz des betreffenden Probanden mit dem Variablenmittelwert anstelle von "kw" in die interne Datenmatrix eingetragen. Aus der Datenmatrix werden dann, wie oben beschrieben, die Datensätze für die Bootstraptstichproben entnommen. Die Bootstraptstichprobe übernimmt also den Variablenmittelwert der Originalstichprobe als Kein-Wert-Ersatz - obwohl ihr eigener Variablenmittelwert sicherlich ein anderer ist. Die Alternative dazu wäre es, den Mittelwert aus den Variablenwerten der Bootstraptstichprobe zu bilden und diesen als Ersatzwert in die Bootstraptstichprobe einzusetzen. Das geschieht in Almo nicht. Aus folgenden Gründen: (1) Almo orientiert sich am bereits oben beschriebenen Analogie-Modell, bei dem die Originalstichprobe dem "Universum der Daten" entspricht und die Bootstraptstichprobe einer Zufallsstichprobe aus diesem entspricht. (2) Dann ist der Mittelwert aus der Originalstichprobe der bessere Schätzwert für den fehlenden Variablenwert als der Mittelwert aus der Bootstraptstichprobe.

Wenn ein Benutzer sich dieser Argumentation nicht anschließen kann, dann sollte er die KW-Behandlungen 4 bis 7 nicht benutzen. Almo bringt eine Warnung, wenn eine KW-

Behandlung von 4 bis 7 gewählt wurde. Wir werden in Abschnitt P30.6.3.2 auf dieses Thema nochmals zurückkommen.

### **P30.6.2 Vorgehensweise beim Bootstrap der Faktorenanalyse**

In der statistischen Literatur ist das Thema, den Standardfehler, den p-Wert und das Konfidenzintervall von Faktorladungen durch Bootstrapping zu finden, mehrfach behandelt worden. Wir orientieren uns im Folgenden ungefähr an der Arbeit von Zientek und Thompson (2007), die auch ein kleines Literaturverzeichnis zu diesem Thema anbieten. Insbesondere werden wir dasselbe Datenbeispiel, die Holzinger-Swineford-Daten, übernehmen. Im Anhang werden wir dann die Vorgehensweise von Zientek/Thompson mit der von Almo vergleichen.

#### ***P30.6.2.0 Sonderfall: Faktorenanalyse mit 1 Faktor***

Dass aus einer Korrelationsmatrix nur 1 Faktor extrahiert wird, ist nur möglich, wenn der Forscher entscheidet, weitere Faktoren, die sich aus dem Eigenwert-Eigenvektor-Kalkül prinzipiell immer ergeben, als irrelevant zu betrachten oder wenn er sich einem spezifischen Auswahl-Kriterium unterwirft, wie z.B. dem Kaiser-Kriterium, bei dem nur Eigenwerte größer 1 akzeptiert werden.

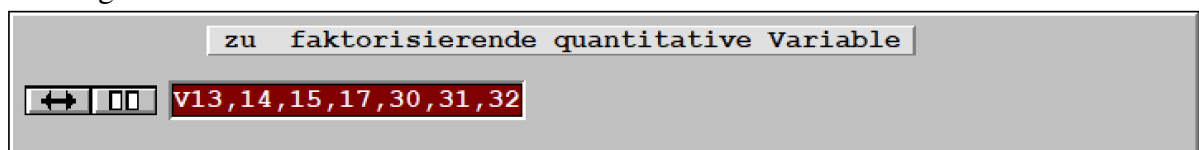
Entsteht aus der Faktorenanalyse der Originalstichprobe nur 1 Faktor, dann wird genau so vorgegangen, wie oben in Abschnitt 30.6.1 für den Regressionskoeffizienten  $\beta_1$  beschrieben. Es wird die Originalstichprobe faktorisiert und die Ergebnisse ausgegeben. Danach werden viele, etwa 1000 Bootstrapstichproben erzeugt, faktorisiert, die Ergebnisse zusammengefasst und Standardfehler, Signifikanz p und Konfidenz-grenzen, wie oben beschrieben, errechnet und ausgegeben.

Wir müssen allerdings im nachfolgenden Abschnitt P30.6.2.0.1 diese Darstellung für eine spezielle Konstellation erweitern.

#### ***Beispiel einer 1-faktoriellen Faktorenanalyse***

Der Benutzer rechne eine Faktorenanalyse mit dem Beispielprogramm **Prog30ml\_1\_Faktor Alm**. Der Benutzer findet das Programm nach Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Das Programm verwendet die Holzinger-Swineford-Daten, die im nächsten Abschnitt vorgestellt werden. Die Optionsbox "Faktoren und Kommunalitäten" in der Programm-Maske bleibt geschlossen. Damit ist das Kaiser-Kriterium gültig, d.h. es werden nur Faktoren, deren Eigenwerte  $\geq 1.0$  sind, verwendet. Der Forscher unterstellt dabei, dass durch das Kaiser-Kriterium die substantiell bedeutsamen Faktoren gefunden werden.

Die Eingabebox für die zu faktorisierenden Variablen lautet:



The image shows a screenshot of a software interface. At the top, there is a label "zu faktorisierende quantitative Variable" in a light gray box. Below it is an input field containing the text "v13,14,15,17,30,31,32". To the left of the input field are two small icons: a double-headed arrow and a square with a diagonal line.

Es entsteht nur 1 Faktor. In Abschnitt P30.6.4.3 werden die Ergebnisse ausgegeben. In der Ergebnisliste gibt Almo für die Originalstichprobe u.a. aus:

```
Eigenwerte (Varianz je Faktor)
-----
alle Eigenwerte > 0
  4.1536  0.9294  0.6140  0.4572  0.3515  0.2590  0.2353
fuer Analyse verwendete Eigenwerte
```

Man erkennt, dass 7 Eigenwerte größer 0 sind, aber nur 1 Eigenwert größergleich 1.0 ist - der dann auch, dem Kaiser-Kriterium entsprechend, für die Analyse verwendet wird. Durch Bootstrap der Eigenwerte wird dies auch bestätigt. In einem nachfolgenden Abschnitt wird das noch gezeigt.

### *Mehrere Faktoren*

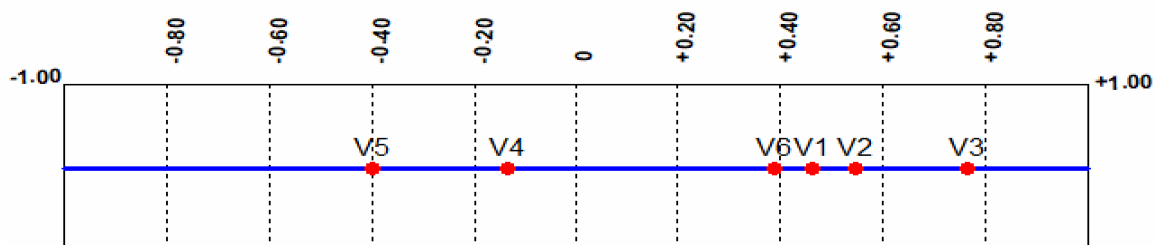
Für das Bootstrap-Verfahren treten Probleme auf, wenn die Faktorenanalyse der Originalstichprobe 2 oder mehr Faktoren liefert. Dann muss die Faktorladungsmatrix der Originalstichprobe in eine inhaltlich interpretierbare Position rotiert werden. Weiterhin ist eine vorausgehende Sonderbehandlung notwendig, die darin besteht, die Faktorladungsmatrizen aus der Originalstichprobe und aus den Bootstrapstichproben in einen "gemeinsamen Raum" zu stellen. Das wird im folgenden Text ausführlich dargestellt. Es gilt also festzuhalten, dass Almo zwischen einem "1-Faktor-Bootstrapping" und einem "Mehr-Faktor-Bootstrapping" unterscheidet. Es wäre also nicht korrekt, eine signifikant mehr-faktorielle Lösung für nur einen Faktor dem Bootstrapping zu unterziehen. Der Benutzer könnte dies jedoch erzwingen, indem er in der Optionsbox "Faktoren und Kommunalitäten" als Faktorenzahl "1" vorgibt.

In Psychologie und Sozialwissenschaften wird häufig versucht Testbatterien bzw. Fragebatterien so zu gestalten, dass sie eine gemeinsame Hintergrunddimension, d.h. einen einzelnen Faktor messen. Das wird dann über eine Faktorenanalyse (mit Kaiser-Kriterium) überprüft. Im Erfolgsfall hat der Forscher dann eine 1-faktorielle Lösung gefunden. Ansonsten dürfte es eher selten sein, dass eine Faktorenanalyse aus der Originalstichprobe nur einen Faktor liefert. Werden 2 oder mehrere Faktoren ausgegeben, dann liegt es am Forscher zu entscheiden, ob nur der 1. Faktor bedeutsam ist und die weiteren Faktoren Artefakte darstellen. Eine gewisse Entscheidungshilfe ist ihm durch das Kaiser-Kriterium gegeben, ebenso durch das Guttman-Kriterium, sowie durch den Scree-Test und den Faktoren-Signifikanztest nach Rippe oder Lawley. Der Benutzer muss zu diesem Zweck die Optionsbox "Faktoren und Kommunalitäten" öffnen.

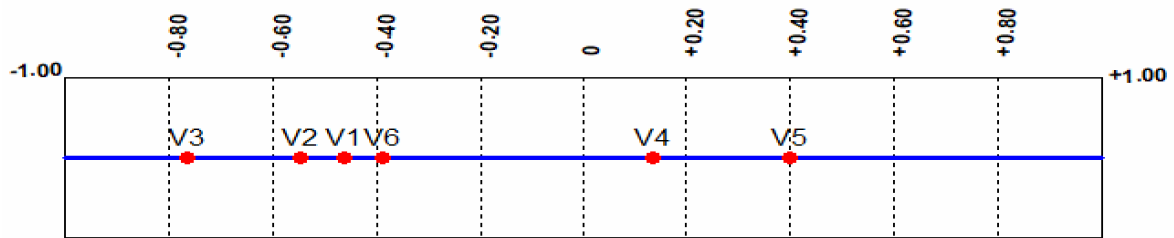
#### ***P30.6.2.0.1 Vorzeichen-Umkehr bei Analyse mit 1 Faktor***

Bei einer Faktorladungsmatrix mit nur 1 Faktor ist eine Vorzeichen-Umkehr für alle Variablen ohne Belang. Es kann z.B. geschehen, dass durch einen Wechsel des Eigenwert-Verfahrens vom Tridiagonal-QR-Verfahren auf das Jacobi-Verfahren die Vorzeichen der Ladungen umgekehrt werden. Grafisch dargestellt:

Ergebnis aus Tridiagonal-QR-Verfahren



Ergebnis aus Jacobi-Verfahren



Beim Bootstrap-Verfahren hingegen darf sich eine Vorzeichen-Umkehr bei einigen Bootstrapstichproben bezüglich der anderen Bootstrapstichproben, insbesondere bezüglich der Originalstichprobe, nicht ereignen. Der Mittelwert und die Standardabweichung der Ladungen würden falsch berechnet - mit der Folge, dass die Verzerrung und der Standardfehler zu groß würden.

Wir müssen also bei jeder Bootstrapstichprobe überprüfen, ob ihre Ladungen auf Faktor 1 vorzeichen-invers zu denen der Originalstichprobe sind.. Dieser Fall tritt real nur auf, wenn die zu faktorisierten Variablen sehr schwach miteinander korrelieren.

Wir rechnen das Beispiel-Programm **Prog20ml\_Zufalldat.Alm**. Man findet es durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Es ist ein sehr extremes Beispiel. Das Programm rechnet eine Faktorenanalyse für 6 Zufallsvariable. Die Korrelationen sind erwartungsgemäß außerordentlich klein.

Korrelations-Matrix

	V1	V2	V3	V4	V5	V6
V1	1.0000	0.0627	0.0335	-0.0195	-0.0520	0.0099
V2	0.0627	1.0000	0.0361	-0.0048	-0.0831	0.0413
V3	0.0335	0.0361	1.0000	-0.0301	-0.0340	0.0570
V4	-0.0195	-0.0048	-0.0301	1.0000	0.0108	-0.0923
V5	-0.0520	-0.0831	-0.0340	0.0108	1.0000	0.0476
V6	0.0099	0.0413	0.0570	-0.0923	0.0476	1.0000

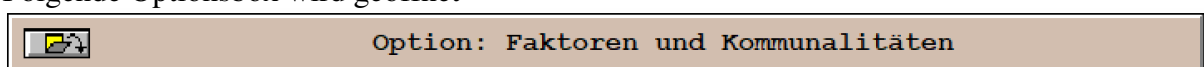
Nach dem Kaiser-Kriterium erhalten wir 2 Faktoren mit Eigenwerten größer 1.0.

Matrix der Faktorladungen

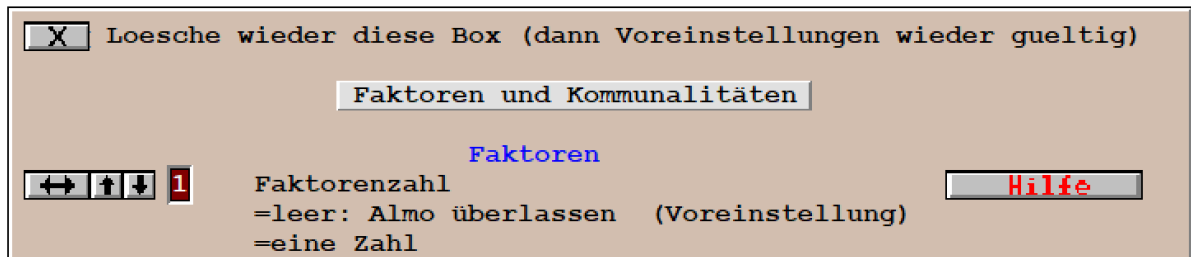
	Faktor 1	Faktor 2
V1	0.4617	0.2531
V2	0.5467	0.2776
V3	0.4642	-0.1314
V4	-0.3719	0.4889
V5	-0.3996	-0.5519
V6	0.3892	-0.6329

Wir wollen jedoch eine 1-faktorielle Lösung erzwingen.

Folgende Optionsbox wird geöffnet



Wir zeigen nur den oberen Teil der geöffneten Box



Wir rechnen ein Bootstrap-Verfahren mit 1000 Bootstrapstichproben, wobei wir das voreingestellte Tridiagonal-QR- Eigenwertverfahren verwenden.

Almo findet 18 Bootstrapstichproben, deren Ladungen auf dem 1. Faktor für alle 6 Variablen vorzeichen-invers zur Originalstichprobe sind, beispielsweise Bootstrapstichprobe 90.

Faktorladungen auf Faktor 1

	Original- stichprobe	Bootstrap- stichprobe 90
V1	0.462	-0.766
V2	0.547	-0.354
V3	0.464	-0.204
V4	-0.372	0.558
V5	-0.400	0.241
V6	0.389	-0.236

Almo bringt zuerst folgende Warnung

\*\*\*\*\* WARNUNG

Bei einigen oder auch allen Variablen der Bootstrapstichproben sind die Ladungen von Faktor 1 vorzeichen-invers zu denen der Originalstichprobe Faktor 1 wird gespiegelt. Alle Vorzeichen werden umgekehrt Hilfe

und meldet dann

---- alle Var. sind vorzeichen-invers in B'stichprobe 90

Bei *allen* Ladungen des Faktors 1 aus der Bootstrapstichprobe 90 wird dann das Vorzeichen umgekehrt, d.h. Faktor 1 wird gespiegelt. Das ist notwendig, damit der Mittelwert über alle Bootstrapstichproben und der Standardfehler korrekt berechnet werden kann.

Verwenden wir anstelle des Tridiagonal-QR-Eigenwertverfahrens das Jacobi-Verfahren, dann sind bei Bootstrapstichprobe 90 alle Ladungen auf Faktor 1 vorzeichen-konform zu denen der Originalstichprobe. Eine Spiegelung wäre dann nicht notwendig gewesen. Insgesamt findet Almo mit dem Jacobi-Verfahren 32 Bootstrapstichproben, bei denen alle 6 Ladungen vorzeichen-invers zu denen der Originalstichprobe sind

Sehr viel häufiger wird man Bootstrap-Stichproben finden, bei denen nur einige Ladungen auf Faktor 1 vorzeichen-invers zu denen der Originalstichprobe sind. In unserem Beispiel ist das 158 mal der Fall wenn wir das voreingestellte Tridiagonal-QR- Eigenwertverfahren rechnen.

Betrachten wir Bootstrapstichprobe 30.

Faktorladungen auf Faktor 1

	A Original- stichprobe	B Bootstrap- stichprobe 30	(A-B) <sup>2</sup>
V1	0.462	0.193	0.072
V2	0.547	-0.331 *	0.770
V3	0.464	0.248	0.047
V4	-0.372	0.611 *	0.966
V5	-0.400	-0.469	0.005
V6	0.389	-0.698 *	1.181
			-----
			3.042 Qsum (gesamte Abweichungs- Quadratsumme)

Man erkennt: Die Ladungen der 3 Variablen V2,4,6 sind vorzeichen-invers zu denen der Originalstichprobe. In obiger Tabelle wurden sie durch einen Stern markiert. Also bringt folgende Meldung:

```
**** einige Var. sind vorzeichen-invers in B'stichprobe 30      3 Var. von 6

                                Qsum          =3.0421
                                Qsum_spiegel=1.8893
```

Nun stellt sich die Frage:

Muss der Faktor 1 aus Bootstrap-Stichprobe 30 gespiegelt werden, damit im weiteren Verlauf des Bootstrap-Verfahrens der Mittelwert für alle Bootstrap-Stichproben und die Standardabweichung (Standardfehler) korrekt ermittelt werden können.

In Also wird folgendermaßen vorgegangen:

- (1) Es wird die Summe der Abweichungsquadrate der 6 Bootstrap-Ladungen von den 6 Ladungen der Originalstichprobe ermittelt

$$Q_{sum} = \text{Summe } (A_i - B_i)^2 = 3.0421$$

$Q_{sum}$  = Summe der Abweichungs-Quadrate aus *allen* Ladungen

$A_i$  = Faktorladung der Variablen i aus Originalstichprobe auf Faktor 1

$B_i$  = Faktorladung der Variablen i aus Bootstrapstichprobe auf Faktor 1

Die Bootstrapstichprobe ist eine Zufallsauswahl aus den originalen Daten. Wir können also  $Q_{sum}$  als "Fehler" der Bootstrapstichprobe begreifen

- (2) Die Vorzeichen aller Bootstrap-Ladungen werden umgekehrt und die Abweichungs-Quadratsumme  $Q_{sum\_spiegel}$  ermittelt. Das Ergebnis ist dann folgendes:

	A Original- stichprobe	B Vorzeichen- umgekehrte Bootstrap- stichprobe 30	(A-B) <sup>2</sup>
V1	0.462	-0.193 *	0.428
V2	0.547	0.331	0.046
V3	0.464	-0.248 *	0.507
V4	-0.372	-0.611	0.057
V5	-0.400	0.469 *	0.755
V6	0.389	0.698	0.095
			-----
			1.889 Qsum_spiegel

=gesamte Abweichungs-  
 Quadratsumme  
 wenn Vorzeichen für alle  
 Ladungen  
 umgekehrt sind)

Zwar sind 3 Ladungen wieder vorzeichen-invers zu denen der Originalstichprobe. Der "Fehler" der vorzeichen-umgekehrten Ladungen der Bootstrapstichprobe 30 ist mit  $Qsum\_spiegel=1.889$  kleiner als der der nicht-umgekehrten Ladungen der Bootstrapstichprobe 30 mit  $Qsum=3.042$ .

Der Benutzer kann nun festlegen, wie in einer solchen Situation reagiert werden soll. Er kann Almo anweisen, den Faktor 1 der Bootstrap-Stichprobe zu spiegeln. Die Folge davon wird sein, dass die schlussendlichen Standardfehler der Variablen kleiner sein werden, allerdings nicht notwendigerweise bei allen Variablen.

In unserem Beispiel ergaben sich als Ergebnis des Bootstrap-Verfahrens folgende Standardfehler bei 3 verschiedenen Methoden der Vorzeichen-Umkehrung (Spiegelung) aller Variablen in Faktor 1.

	Standardfehler		
	I	II	III
V1	0.289243	0.273139	0.280454
V2	0.382197	0.371827	0.290810
V3	0.371806	0.361919	0.272845
V4	0.414685	0.408516	0.354837
V5	0.396379	0.389956	0.371845
V6	0.457500	0.451433	0.393072

Spalte I = Keine Umkehrung bei vorzeichen-inversen Ladungen  
 Spalte II = Umkehrung nur wenn alle Variable vorzeichen-invers sind  
 Spalte III = Umkehrung wenn alle Variable vorzeichen-invers sind und auch Umkehrung wenn nur einige Variable vorzeichen-invers sind und  $Qsum\_spiegel$  kleiner ist als  $Qsum$

Für unser Beispiel meldet Almo, dass es folgende Spiegelungen bei 176 Bootstrapstichproben vorgenommen hat. Bei 18 Stichproben waren alle 6 Ladungen vorzeichen-invers zu denen der Originalstichprobe. Bei 158 Bootstrapstichproben waren nur einige Variable vorzeichen-invers. (Ausgabe verkürzt.  $Qsum$  und  $Qsum\_spiegel$  wegen Platzmangel weggelassen)

Vorzeichen-Umkehrungen wenn mit Tridiagonal-QR- Eigenwertverfahren gerechnet wird

```

1 **** einige Var. sind vorzeichen-invers in B'stichprobe 16 (4 Var. von 6)
2 **** einige Var. sind vorzeichen-invers in B'stichprobe 18 (3 Var. von 6)
3 **** einige Var. sind vorzeichen-invers in B'stichprobe 24 (5 Var. von 6)
4 **** einige Var. sind vorzeichen-invers in B'stichprobe 30 (3 Var. von 6)
5 **** einige Var. sind vorzeichen-invers in B'stichprobe 45 (3 Var. von 6)
6 **** einige Var. sind vorzeichen-invers in B'stichprobe 54 (4 Var. von 6)
7 **** einige Var. sind vorzeichen-invers in B'stichprobe 58 (5 Var. von 6)
8 **** einige Var. sind vorzeichen-invers in B'stichprobe 65 (5 Var. von 6)
9 **** einige Var. sind vorzeichen-invers in B'stichprobe 84 (5 Var. von 6)
10 **** einige Var. sind vorzeichen-invers in B'stichprobe 86 (3 Var. von 6)
11 ---- alle Var. sind vorzeichen-invers in B'stichprobe 90
12 **** einige Var. sind vorzeichen-invers in B'stichprobe 91 (3 Var. von 6)
.      .      .      .      .      .
.      .      .      .      .      .
.      .      .      .      .      .
176 **** einige Var. sind vorzeichen-invers in B'stichprobe 979 (3 Var. von 6)

```

Wird mit dem Jacobi-Eigenwertverfahren gerechnet (mit identischer Startzahl für den Zufallsgenerator), dann werden teilweise andere Bootstrapstichproben im Vorzeichen umgekehrt. Insgesamt werden die Faktoren 1 aus 167 Bootstrapstichproben umgekehrt.

Wir zeigen hier nur die ersten 8 Umkehrungen

Vorzeichen-Umkehrungen wenn mit Jacobi-Eigenwertverfahren gerechnet wird

```

1 **** einige Var. sind vorzeichen-invers in B'stichprobe 16 (4 Var. von 6)
2 **** einige Var. sind vorzeichen-invers in B'stichprobe 18 (3 Var. von 6)
3 **** einige Var. sind vorzeichen-invers in B'stichprobe 24 (5 Var. von 6)
4 **** einige Var. sind vorzeichen-invers in B'stichprobe 28 (4 Var. von 6)
5 **** einige Var. sind vorzeichen-invers in B'stichprobe 30 (3 Var. von 6)
6 **** einige Var. sind vorzeichen-invers in B'stichprobe 31 (5 Var. von 6)
7 ---- alle Var. sind vorzeichen-invers in B'stichprobe 33
8 **** einige Var. sind vorzeichen-invers in B'stichprobe 43 (5 Var. von 6)
. . . . .
. . . . .
. . . . .

```

Von den ersten 8 gemeldeten Vorzeichen-Umkehrungen werden nur die vier Bootstrapstichproben 16, 18, 24, 30 bei beiden Verfahren gemeldet.

Betrachten wir Bootstrapstichprobe 31, die nur beim Jacobi-Verfahren gemeldet wird, genauer

Matrix der Faktorladungen

	Originalstichprobe		Bootstrapstichproben 31	
	Tridiagonal	Jacobi	Tridiagonal	Jacobi
V1	0.461740	0.461740	0.278002	-0.278002
V2	0.546670	0.546670	0.030270	-0.030270
V3	0.464216	0.464216	0.478287	-0.478287
V4	-0.371901	-0.371901	-0.446313	0.446313
V5	-0.399556	-0.399556	0.505644	-0.505644
V6	0.389207	0.389207	0.696904	-0.696904

Die absoluten Werte sind bei Tridiagonal-QR- und Jacobi-Verfahren exakt identisch – sowohl bei der Originalstichprobe als auch bei Bootstrapstichprobe 31. Die Werte der Originalstichprobe unterscheiden sich jedoch erheblich von denen der Bootstrapstichprobe. Das ist dadurch zu erklären, dass die Bootstrapstichprobe eine *Zufalls*-Stichprobe aus einer Datenmatrix ist, die ihrerseits aus *Zufalls*-Werten besteht. Die mit dem Tridiagonal-QR-Verfahren errechneten Bootstrap-Ladungen sind vorzeichenkonform zu denen der Originalstichprobe, die aus dem Jacobi-Verfahren vorzeicheninvers.

Trotz der dargestellten Unterschiede sind alle Bootstrap-Ergebnisse (Mittelwert, Standardfehler, p-Wert, Konfidenzintervall) aus dem Tridiagonal-QR- und dem Jacobi-Verfahren für alle 6 Variable exakt identisch.

Wir wollen jetzt noch betrachten wie sich die Bootstrap-Ergebnisse unterscheiden, wenn wir die Vorzeich-Umkehrung abschalten.

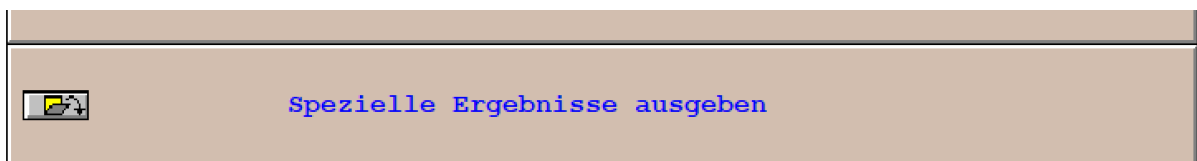
	Standardfehler			Verzerrung		
	I	II	III	I	II	III
V1	0.289243	0.273139	0.280454	-0.112309	-0.099595	-0.105224
V2	0.382197	0.371827	0.290810	-0.260665	-0.247320	-0.168197
V3	0.371806	0.361919	0.272845	-0.224509	-0.209840	-0.116091
V4	0.414685	0.408516	0.354837	0.231050	0.214064	0.115294
V5	0.396379	0.389956	0.371845	0.164114	0.153634	0.127047

V6 0.457500 0.451433 0.393072 | -0.245136 -0.227138 -0.114430

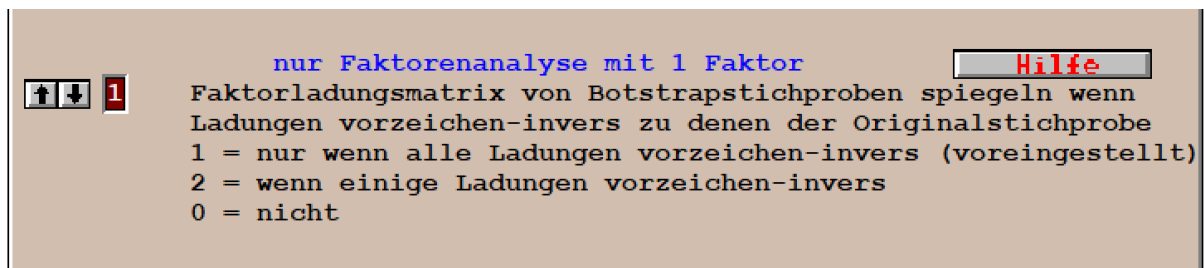
Spalte I = Keine Korrektur bei vorzeichen-inversen Ladungen  
Spalte II = Korrektur nur wenn alle Variable vorzeichen-invers sind  
Spalte III = Korrektur wenn alle Variable vorzeichen-invers sind und auch  
Korrektur wenn nur einige Variable vorzeichen-invers sind  
und Qsum\_spiegel kleiner ist als Qsum

Verzerrung und Standardfehler werden von Spalte I über II zu Spalte III bei allen Variablen kleiner – mit Ausnahme von V1. Die Werte von V1 in Spalte II und III sind zwar kleiner als die nicht korrigierten Werte in Spalte I, jedoch ist der Wert in Spalte III größer als in Spalte II.

In der Bootstrap-Box der Almo-Programm-Maske kann der Benutzer in der Sub-Box "Spezielle Ergebnisse ausgeben" zwischen 3 Möglichkeiten wählen, den durch vorzeichen-inverse Ladungen verursachten Fehler zu korrigieren.



Wird die Sub-Box geöffnet, dann sieht man folgendes. Wir zeigen nur einen Ausschnitt



**Möglichkeit 1** ist voreingestellt. Das bedeutet: Wenn die Sub-Box in der Bootstrap-Box geschlossen bleibt, dann werden die Ladungsmatrizen aus Bootstrapstichproben, bei denen **alle** Ladungen vorzeichen-invers zu denen der Originalstichprobe sind, gespiegelt.

**Möglichkeit 2** inkludiert 1, d.h. bei 2 werden gespiegelt die Ladungsmatrizen der Bootstrapstichproben, bei denen **alle** Ladungen vorzeichen-invers sind

und

zusätzlich auch die Bootstrapstichproben, bei denen nur einige Ladungen vorzeichen-invers sind, sofern die Abweichungs-Quadratsumme der gespiegelten Ladungsmatrix **Qsum\_spiegel** kleiner ist als **Qsum** der Ladungsmatrix mit vorzeichen-inversen Ladungen. Beachte: Die gesamte Ladungsmatrix wird gespiegelt. Bei allen Ladungen wird das Vorzeichen umgedreht - nicht nur bei den vorzeichen-inversen.

**Möglichkeit 3** Almo überprüft nicht, ob vorzeichen-inverse Ladungen existieren.

### P30.6.2.1 Bootstrap der Eigenwerte

Eine weitere Möglichkeit, die Faktorenzahl zu bestimmen, besteht darin, die *Eigenwerte* dem Bootstrap-Verfahren zu unterwerfen. Wir rechnen wieder mit dem Beispielprogramm **Prog30ml\_1\_Faktor.Alm** mit 1000 Bootstrap-Stichproben und erhalten dabei folgendes Ergebnis:

#### Bootstrap der Eigenwerte

Beachte: In Spalte \*g wird getestet, ob der Eigenwert > 1.0 ist  
 p ist die Wahrscheinlichkeit, dass die Annahme, der Eigenwert sei > 1.0 falsch ist  
 (1-p)\*100 ist der Prozentanteil der Bootstrapstichproben, deren Eigenwert > 1.0 ist  
 Spalte \*g ist nur sinnvoll interpretierbar bei Kommunalitätschätzung 1 (Diagonale=1.0 und keine Iterationen)

	Original- Stichprobe	Ergebnisse aus 1000 Bootstrapstichproben der Eigenwerte					
		*a Eigenwert	*d	*e	*f	*g	*h
			Mitt.wert Eigenwert	Verzerr. Eigenwert	Standard fehler	Signifik. p	Konfidenzintervall Konf.niv=0.950 unten oben
Eigenwert 1	4.153619	4.151614	-0.002005	0.145854	0.000999	3.873642	4.420299
Eigenwert 2	0.929403	0.937657	0.008254	0.079946	0.793207	0.785945	1.101808
Eigenwert 3	0.614010	0.621479	0.007468	0.048696	0.999001	0.526597	0.722532
Eigenwert 4	0.457152	0.460589	0.003437	0.040095	0.999001	0.387634	0.542691
Eigenwert 5	0.351473	0.348520	-0.002953	0.029374	0.999001	0.293257	0.407730
Eigenwert 6	0.258996	0.259549	0.000553	0.021588	0.999001	0.220789	0.303890
Eigenwert 7	0.235347	0.220593	-0.014755	0.019850	0.999001	0.183565	0.258663

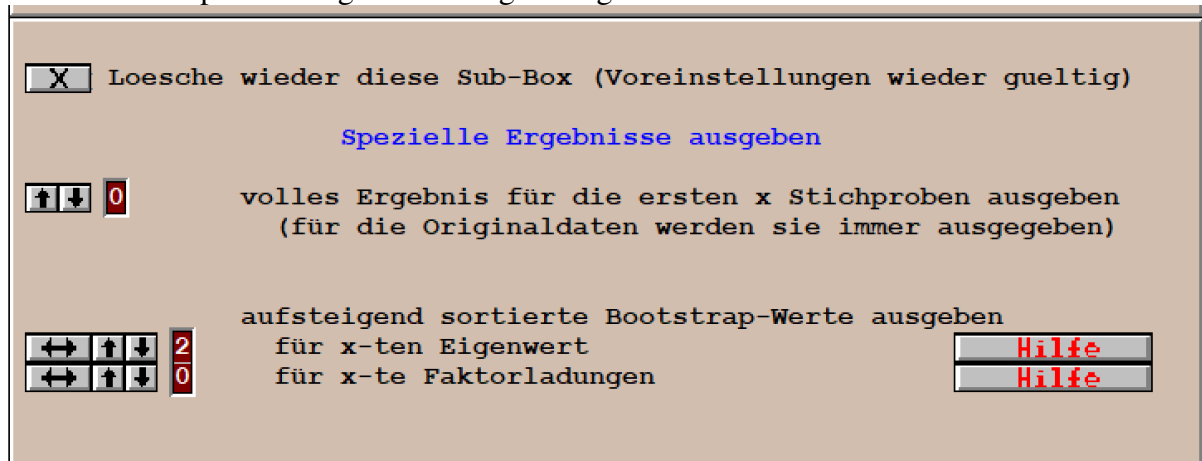
- \*a Eigenwerte aus Originalstichprobe
- \*d Mittelwert der Eigenwerte aus allen Bootstrap-Stichproben
- \*e mit "Verzerrung" wird die Differenz zwischen dem Mittelwert aus allen Bootstrap-Stichproben minus dem Wert aus der Originalstichprobe bezeichnet
- \*f Der Standardfehler ist gleich der Standardabweichung der Eigenwerte aus allen Bootstrap-Stichproben
- \*g Getestet wird, ob der Eigenwert grösser 1.0 sein könnte  
 1-p ist der Anteil der aufsteigend sortierten Eigenwerte grösser 1.0 an der Zahl von 1000 Bootstrapstichproben
- \*h Konfidenzintervall - nach dem vom Benutzer vorgegebenen Konfidenz-Niveau von 95.00%  
 Beim "einfachen" Perzentil-Verfahren bedeutet das:  
 Von den aufsteigend sortierten 1000 Werten aus den Bootstrap-Stichproben befinden sich 95.00 % der Werte zwischen den Konfidenzgrenzen und je 2.50% oberhalb und unterhalb der Konfidenzgrenzen

Bei 7 Variablen erhalten wir 7 Eigenwerte, sofern in der Optionsbox "Faktoren und Kommunalitäten" die Kommunalitätenschätzung 1 eingesetzt wurde. Das ist auch die Voreinstellung. Die Korrelationsmatrix wird dann mit 1.0 in der Diagonalen faktorisiert. Entsprechend dem Kaiser-Kriterium werden dann alle Faktoren > 1.0 für substantiell erachtet. Die Frage, die sich nun in unserem Beispiel stellt, ist ob der 2. Eigenwert, der in der Originalstichprobe mit 0.929 sehr nahe bei 1.0 liegt, mit einer gewissen Wahrscheinlichkeit doch >1.0 sein könnte. Das Konfidenzintervall des 2. Eigenwerts enthält von **0.785945 bis 1.101808** Werte >1.0. Daraus würde folgen, dass der 2. Eigenwert noch dem Kaiser-Kriterium entsprechen würde. Diese Folgerung ist falsch. Wir erfahren mit einer Irrtumswahrscheinlichkeit von 5% nur, dass (1) der Wert 0 außerhalb des Intervalls liegt, aber das ist für unsere Fragestellung irrelevant, und (2) dass der "wahre" Eigenwert sich im angegebenen Intervall befindet, also auch unter 1.0 liegen kann. Das Konfidenzintervall liefert keine brauchbare Information.


Was wir jedoch wissen wollen, ist ob der "wahre" Wert *größer* 1.0 ist. Das wird uns durch den p-Wert=0.7922 angezeigt und das erfahren wir durch folgende Berechnung, die uns einen *einseitigen* p-Wert für die *gerichtete* Hypothese liefert, dass der Eigenwert grösser 1

ist. Zu den Begriffen *einseitig* und *gerichtet* siehe etwa Bortz (1993, Abschnitt 4.4 und 4.5).

Wir haben uns die aufsteigend sortierten 2. Eigenwerte der 1000 Bootstrapstichproben ausgeben lassen. Dazu wird in der Bootstrap-Optionsbox von **Prog30ml\_1\_Faktor.Alm** die Sub-Box "Spezielle Ergebnisse ausgeben" geöffnet.



Im zweiten Eingabefeld wird Almo aufgefordert, die Werte der 1000 Bootstrapstichproben des 2. Eigenwerts auszugeben. In der Ergebnisliste erscheint oberhalb der Tabelle der Eigenwerte diese Ausgabe

 Zeige Ausgabe:      Aufsteigend sortierte Werte fuer 2.Eigenwert

Nach Klick auf den Öffne-Knopf werden dann die 1000 Werte präsentiert. Wir zeigen die Ausgabe hier stark verkürzt

Stichprobe	2.Eigenwert
1	0.713760
2	0.721502
3	0.727416
4	0.741014
.	.
.	.
.	.
790	0.999459
791	0.999605
792	0.999694
793	1.000344
794	1.000426
795	1.000603
.	.
.	.
.	.
998	1.187031
999	1.203081
1000	1.235440

Ab der Stichprobe 793 sind die Eigenwerte *grösser* 1.0. Von den 1000 Eigenwerten sind somit 792 kleiner als 1.0 und 208 *grösser* 1.0. Damit ist schon klar entschieden, dass der 2. Eigenwert nicht das Kaiser-Kriterium erfüllt.

In unserem Beispiel wird der Anteil, bzw. die Wahrscheinlichkeit, dass der 2. Eigenwert *grösser* 1.0 in Formel (1) ausgedrückt

$$(1) \quad \text{Anteil} = 208/1000 = 0.208$$

Die Irrtumswahrscheinlichkeit, der p-Wert ist dann

$$(2) p = 1.0 - 208 / (1000 + 1) = 0.7922$$

Die Schlussfolgerung für unser Beispiel ist: Der 2. Eigenwert ist mit  $p=0.7922$  eindeutig nicht signifikant. Er erfüllt nicht das Kaiser-Kriterium.

Der p-Wert, den Almo ausgibt ist *einseitig* und *gerichtet*

Aus (2) folgt  $1-p = 0.2078$ . Die Differenz zum Wert aus (1) ist so gering, dass  $1-p$  deswegen auch als Anteilswert interpretiert werden darf. Je größer die Stichprobenzahl umso geringer wird der Unterschied.

Auch die beiden Beispiel-Programme „Prog30ml\_Schwachkorrdat.Alm“ und „Prog30ml\_Sport.Alm“ zeigen in eindrucksvoller Weise, wie der Bootstrap der Eigenwerte dem Forscher helfen kann, die signifikante Faktorenzahl zu finden. Beide Programme können nach Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters geöffnet werden.

### ***P30.6.2.2 Unsere Beispieldaten: Die Holzinger-Swineford-Daten***

Wir werden in unserer Darstellung immer wieder auf Daten von Holzinger und Swineford zurückgreifen. Der Benutzer findet sie im Almo-Ordner TESTDAT. Durch Doppelklick auf den Namen "HolzingerSwineford1939.fre" kann er sie laden und anschauen. Dazu gehört noch die Datei der Variablennamen "HolzingerSwineford1939.nam".

Diese aus dem Jahr 1939 stammende Datei wird in vielen Arbeiten zur Faktorenanalyse als Beispieldatei verwendet. Die Daten entstanden aus einer Studie zur Begabung (Intelligenz) von 301 Schülern der 7. und 8. Klassen aus 2 verschiedenen Schulen. Die ersten 8 Variablen sind demographische Variable, darauf folgen die 24 "Begabungsfragen" t1 bis t24. Aus den von verschiedenen Forschern durchgeführten Faktorenanalysen ergaben sich 3 bis 5 inhaltlich sinnvoll interpretierbare Faktoren. Thurstone kam auf die 3 Faktoren (1) verbal reasoning, (2) numerical reasoning und (3) space reasoning.

Wir werden (in Anlehnung an die Studie von Zientek/Thompson, 2007) einmal 6 Begabungsfragen für eine 2-faktorielle Analyse und einmal 9 für eine 3-faktorielle Analyse herausgreifen. Die 9 Variablen sind:

Variablen- nummer	Variablennamen
V14	t6_paragraph_comprehension
V15	t7_sentence
V17	t9_word_meaning
V18	t10_addition
V20	t12_counting_groups_of_dots
V21	t13_straight_and_curved_capitals
V22	t14_word_recognition
V23	t15_number_recognition
V25	t17_object_number

Die erste und die letzte 3er-Gruppe werden für die 2-faktorielle Analyse verwendet. Sollen alle Variablen zusammen faktorisiert werden, dann muss in der Programm-Maske **v9:32** als Analysevariable eingetragen werden. Die letzten 2 Variablen V33 und V34 wurden nur in einer der beiden Schulklassen erhoben

### ***P30.6.2.3 Rotation und "gemeinsamer Faktoren-Raum"***

Das Bootstrapping liefert, wie oben in P30.6.1 bereits ausgeführt, den (1) Standardfehler der Faktorladungen, ihren (2) p-Wert und (3) den unteren und (4) oberen Grenzwert ihres Konfidenzintervalls. Die Matrix der Faktorladungen wird grafisch dargestellt als Punktekonfiguration in einem Koordinatensystem (siehe nachfolgende Grafik1). Die Werte für die vier genannten Koeffizient sind nun abhängig von der Lage der "Punktekonfiguration" im Koordinatensystem. Werden die Koordinatenachsen gedreht (rotiert), wobei die Punktekonfiguration stehen bleibt, dann entstehen andere Koordinatenwerte für die Punkte und andere Werte für die vier Koeffizienten. Es muss also entschieden werden, in welche Position das Koordinatensystem letztendlich rotiert werden soll. Erst für diese werden dann die oben genannten vier Koeffizienten des Bootstrappings ermittelt.

#### ***P30.6.2.3.1 Rotation***

Das mathematische Eigenwert-Eigenvektor-Verfahren liefert uns die "orthogonale unrotierte Faktorladungsmatrix". Damit enthüllt uns dieses Verfahren zwar die "Gestalt" der Punktekonfiguration, positioniert diese Punktekonfiguration jedoch im Koordinatensystem nach einem mathematischen Maximierungsprinzip, das keinen Bezug zum Forschungsgegenstand besitzt. Die Punktwolke muss im Koordinatensystem noch in eine sinnvolle Position rotiert werden,

Eine sinnvolle Position besteht, wenn

1. die Punktekonfiguration in so viele "eigenständige Punktwolken" aufgelöst werden kann wie Faktoren extrahiert wurden - anders formuliert, wenn den extrahierten Faktoren "eigenständige Punktwolken" entsprechen und die (den Faktoren entsprechenden) Koordinatenachsen mitten durch oder möglichst nahe bei den Punktwolken verlaufen. Siehe dazu die nachfolgende Graphik2. Diese Position wird in Almo dadurch erreicht, dass die "orthogonale unrotierte Faktorladungsmatrix" (rechtwinklig) nach dem Varimax-Verfahren oder (schiefwinklig) nach dem Quartimin-Verfahren rotiert wird. Selbstverständlich könnten auch andere bewährte Rotationsverfahren eingesetzt werden.
2. Die zunächst durch den Rotations-Algorithmus gefundene Position muss hinsichtlich des Forschungsgegenstands inhaltlich interpretiert werden können. Den gefundenen Faktoren müssen Begriffe zugeordnet werden können. Siehe zu dieser Thematik auch Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.1.9.

Durch das Bootstrap-Verfahren werden dem Forscher für jede einzelne Faktorladung die 4 Koeffizienten "Standardfehler", p-Wert und die beiden Grenzwerte des Konfidenzintervall geliefert. Darin besteht ja die besondere Leistung des Bootstrap-Verfahrens bei der Faktorenanalyse. Diese 4 Koeffizienten sind nun in ihren Werten, davon abhängig wie die Punktwolke im Koordinatensystem positioniert ist. Bei zwei unterschiedlichen Lagen derselben Punktwolke können je Faktorladung völlig verschiedene Werte entstehen. Deswegen ist es wichtig die Punktwolke in eine "sinnvolle" Position zu rotieren

#### ***P30.6.2.3.2 Gemeinsamer Faktorenraum***

Eine weitere Bedingung muss erfüllt werden: Die Faktorladungen der Originalstichprobe und aller Bootstrapstichproben müssen sich im gleichen "Faktoren-Raum" befinden. Was

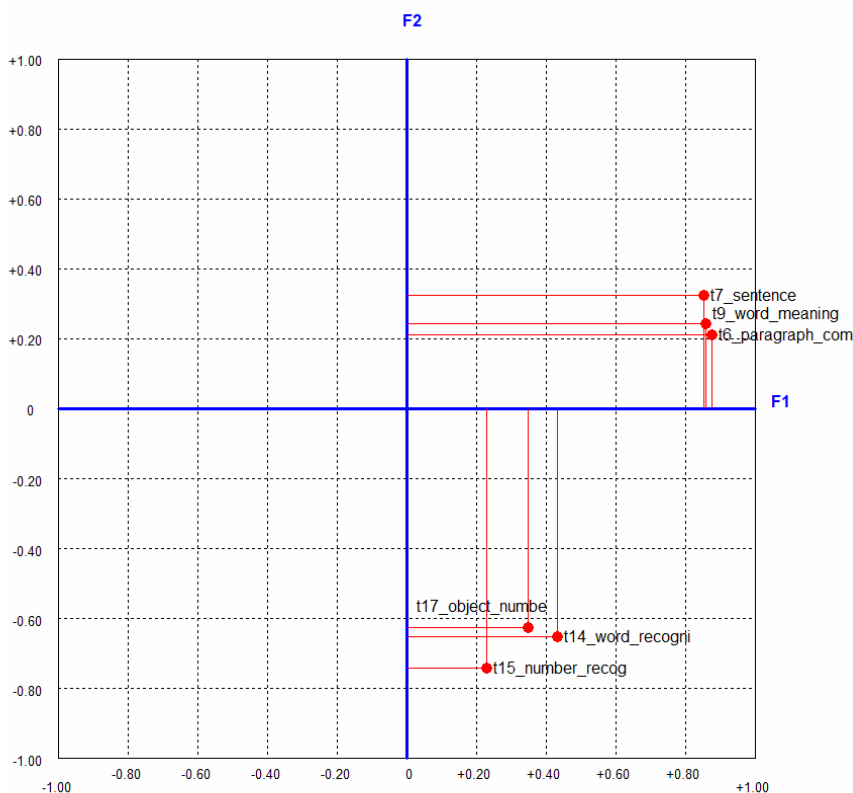
damit gemeint ist, wollen wir am Beispiel der Holzinger-Swineford-Daten erläutern. Für die Originalstichprobe haben wir folgende orthogonale unrotierte Faktorladungsmatrix gefunden.

Matrix der nicht-rotierten orthogonalen Faktorladungen der Originalstichprobe

		Faktor 1	Faktor 2
t6_parag	V14	0.8752	0.2121
t7_sente	V15	0.8513	0.3246
t9_word_	V17	0.8591	0.2444
t14_word	V22	0.4328	-0.6527
t15_numb	V23	0.2310	-0.7418
t17_obje	V25	0.3487	-0.6257

Die Matrix wird graphisch in einem 2-dimensionalen Koordinatensystem abgebildet.

Graphik 1 Faktorladungen der Originalstichprobe



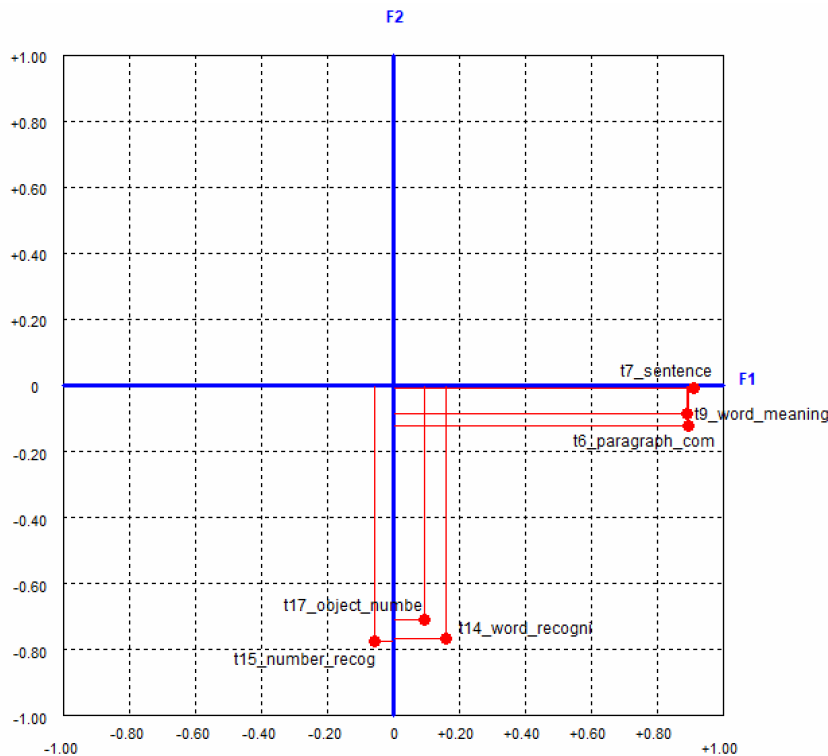
Diese Matrix wird nach dem Varimax-Prinzip orthogonal rotiert.

Varimax-rotierte Faktorladungen der Originalstichprobe

		Faktor 1	Faktor 2
t6_parag	V14	0.8918	-0.1251
t7_sente	V15	0.9110	-0.0117
t9_word_	V17	0.8887	-0.0891
t14_word	V22	0.1620	-0.7662
t15_numb	V23	-0.0584	-0.7747
t17_obje	V25	0.0938	-0.7101

Die höchsten Ladungen der ersten 3 Variablen liegen auf Faktor 1, die der zweiten 3 Variablen auf Faktor 2. Damit haben wir eine dem Forschungsgegenstand (der Begabung von Schülern) angemessene inhaltliche Interpretation ermöglicht. Graphisch dargestellt:

Graphik 2 Varimax-rotierte Faktorladungen der Originalstichprobe



Aus der Varimax-Ladungsmatrix der Originalstichprobe wird eine *Zielmatrix* gewonnen. Sie definiert den "Raum", in den die Faktorladungsmatrizen der Originalstichprobe und aller Bootstraptstichproben "hinein gezwungen" werden. Dieser Vorgang wird "Prokrustes-Rotation" genannt. Da die Ausgangs-Faktorladungsmatrix aber nicht nur rotiert wird, sondern (wie noch gezeigt wird) auch "gespiegelt" und "vertauscht" wird, ziehen wir es vor, von "Prokrustes-Anpassung" zu sprechen.

Als *Zielmatrix* für die Prokrustes-Anpassung bietet Almo u. a. die "-1, 0, 1" -Zielmatrix der Original-Stichprobe an. Sie entsteht in folgender Weise aus der Varimax-Faktorladungsmatrix der Originalstichprobe. Betrachten wir unser Beispiel

		Varimax-rotierte Faktorladungen		Zielmatrix	
		Faktor 1	Faktor 2	Faktor 1	Faktor 2
t6_parag	V14	<u>0.8918</u>	-0.1251	1.0000	0
t7_sente	V15	<u>0.9110</u>	-0.0117	1.0000	0
t9_word_	V17	<u>0.8887</u>	-0.0891	1.0000	0
t14_word	V22	0.1620	<u>-0.7662</u>	0	-1.0000
t15_numb	V23	-0.0584	<u>-0.7747</u>	0	-1.0000
t17_obje	V25	0.0938	<u>-0.7101</u>	0	-1.0000

In jeder Zeile wird der absolut größte Wert gesucht. Ist er positiv wird 1.0 gesetzt, ist er negativ dann -1.0. Für die anderen Werte in der Zeile wird 0 eingesetzt. Die Zielmatrix ist also eine Faktorladungsmatrix, bei der die ersten 3 Variablen mit 1.0 auf dem 1. Faktor und mit 0 auf dem 2. Faktor laden. Die Variablen der zweiten Gruppe laden alle mit 0 auf dem 1. Faktor und mit -1.0 auf dem 2. Faktor.

Im nächsten Abschnitt werden wir zeigen, dass noch weitere orthogonale und auch schiefwinklige Zielmatrizen entworfen werden können. Naheliegend ist es z.B., die rechtwinklige Varimax-Matrix *selbst* als Zielmatrix für die Bootstraptstichproben einzusetzen. Wir werden in Abschnitt P30.6.2.6 darauf zurückkommen

Wird die -1,0,1- Matrix als Zielmatrix verwendet, dann muss auch die nicht-rotierte Faktorladungsmatrix der Originalstichprobe noch an die Zielmatrix "heran gezwungen" werden. Dabei entsteht folgendes Ergebnis

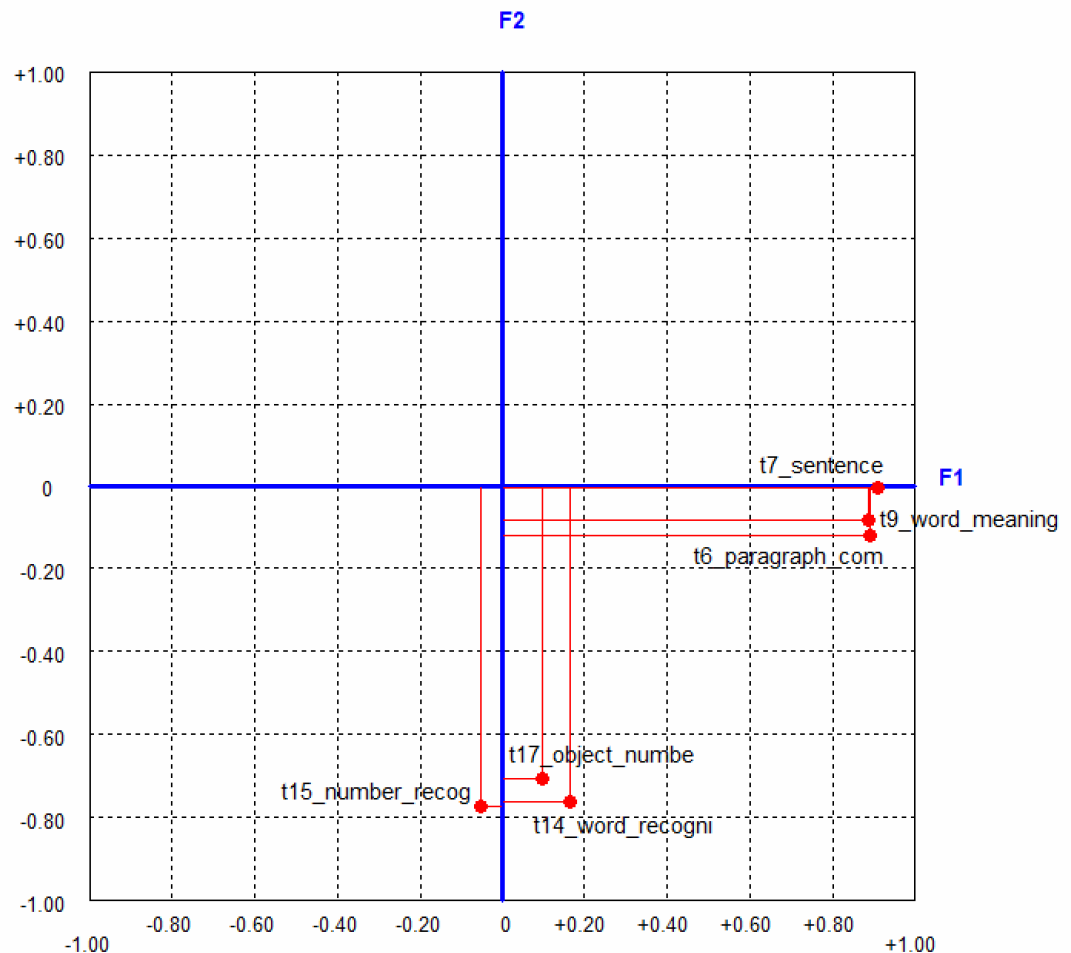
Matrix der orthogonalen konfirmatorischen Faktorladungen der Originalstichprobe

		Faktor 1	Faktor 2
t6_parag	V14	0.8925	-0.1199
t7_sente	V15	0.9111	-0.0065
t9_word_	V17	0.8892	-0.0839
t14_word	V22	0.1664	-0.7652
t15_num	V23	-0.0540	-0.7750
t17_obje	V25	0.0979	-0.7096

Wir bezeichnen die so gewonnene Matrix als *konfirmatorische* Faktorladungsmatrix - dies deswegen, weil das "Heranzwingen" an eine Zielmatrix, der Gegenstand der "konfirmatorischen Faktorenanalyse" (nach dem Kleinste-Quadrate-Kalkül) ist. Graphisch dargestellt

Graphik 3

orthogonale konfirmatorische Faktorladungen



Auch die vielen Bootstrapstichproben müssen mit den für sie errechneten orthogonalen nicht-rotierten Faktorladungsmatrizen an die -1,0,1 -Zielmatrix "heran gezwungen" werden, damit sie in demselben "Raum" liegen wie die der Originalstichprobe. Für einige der Bootstrapstichproben ist dies auch dringend notwendig - wie in folgendem Beispiel ersichtlich wird.

### ***P30.6.2.3.3 Gespiegelte Faktorladungsmatrizen***

Für die Bootstrap-Stichprobe Nr. 29 erhalten wir folgende nicht-rotierte orthogonale Faktorladungsmatrix:

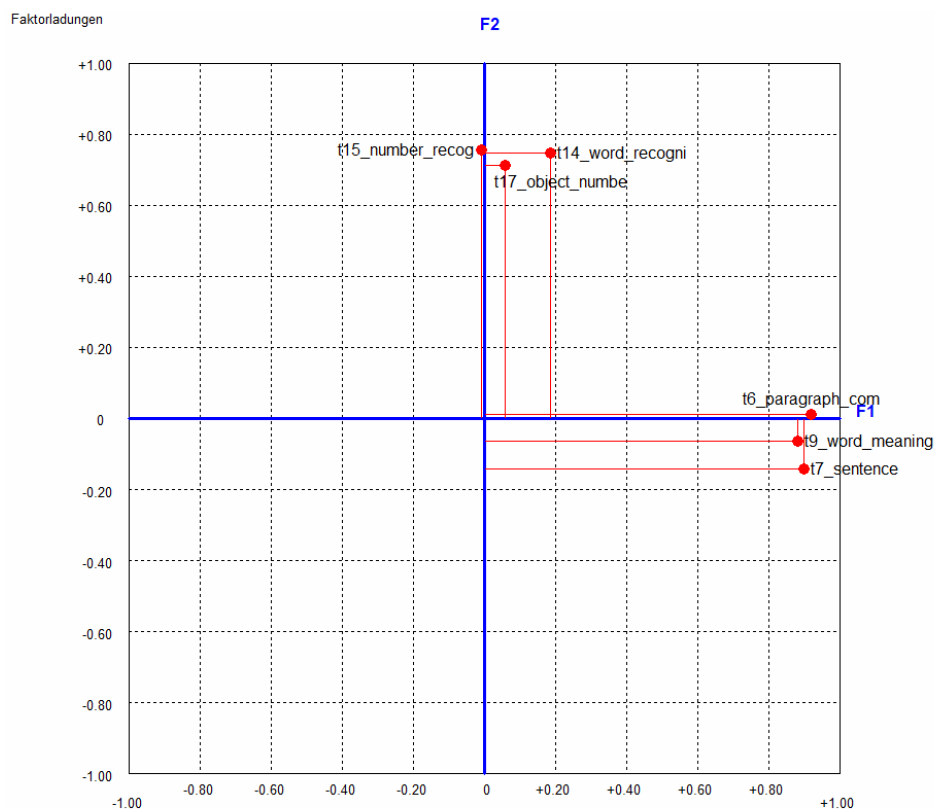
Matrix der nicht-rotierten orthogonalen Faktorladungen der Bootstrapstichprobe 29

		Faktor 1	Faktor 2
t6_parag	V14	0.9186	0.0128
t7_sente	V15	0.8983	-0.1419
t9_word_	V17	0.8810	-0.0650
t14_word	V22	0.1856	0.7477
t15_numb	V23	-0.0113	0.7560
t17_obje	V25	0.0599	0.7117

Auch hier, wie bei der Originalstichprobe liegen die höchsten Ladungen der ersten Variablen-gruppe auf Faktor 1, die der zweiten auf Faktor 2. Die Vorzeichen der zweiten sind jedoch auf Faktor 2 umgedreht. Graphisch bedeutet dies, dass die Variablen-Punkte im Vergleich zur Originalstichprobe entlang der Achsen F2 von unten nach oben *gespiegelt* sind. Das wird in der nachfolgenden Abbildung der Ladungsmatrix gut sichtbar.

#### Graphik 4

nicht-rotierte Faktorladungen der Bootstrapstichprobe 29



Wenn wir nun den Mittelwert aus allen Bootstrap-Faktorladungen berechnen wollen, dann müssen wir zuvor die Spiegelung rückgängig machen. Wir müssen gewissermaßen der Bootstrapstichprobe Nr. 29 zeigen, wie sie ihre Punktekonfiguration drehen oder kippen muss, damit diese in dieselbe Position gerät, wie die der Originalstichprobe. Die Grundvoraussetzung dabei ist:

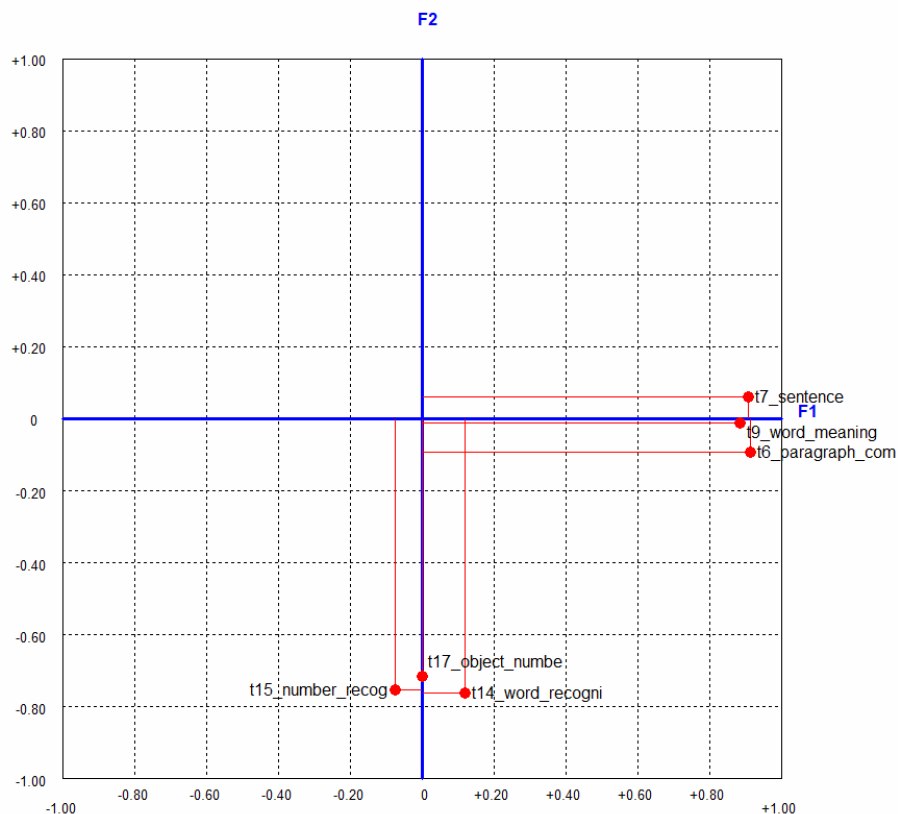
Die Punktekonfiguration muss unverändert bleiben, d.h. die euklidischen Distanzen zwischen allen Punkten müssen gleich bleiben. Und die aus der so veränderten Faktorladungsmatrix ableitbare "reproduzierte Korrelationsmatrix" muss identisch sein mit der aus der ursprünglichen, unrotierten Faktorladungsmatrix.

Die Lösung für diese Aufgabe ist: Wir geben die *Zielmatrix* vor (die aus der Varimax-Ladungsmatrix der Originalstichprobe hervorgegangen ist) und zwingen den Faktorladungen der Bootstrapstichprobe 29 eine "Prokrustes-Anpassung" auf. Das Verfahren, das dies leistet ist die **konfirmatorische Faktorenanalyse**. Als Ergebnis entsteht dann die "konfirmatorische Faktorladungsmatrix":

Matrix der orthogonalen konfirmatorischen Faktorladungen der Bootstrap-Stichprobe 29

		Faktor 1	Faktor 2
t6_parag	V14	0.9140	-0.0926
t7_sente	V15	0.9072	0.0633
t9_word_	V17	0.8833	-0.0117
t14_word	V22	0.1200	-0.7610
t15_numb	V23	-0.0769	-0.7521
t17_obje	V25	-0.0021	-0.7142

Graphisch dargestellt:  
Graphik 5



Wenn man Graphik 5 mit Graphik 3 (von der Originalstichprobe) vergleicht, dann erkennt man, dass die Punktekonfiguration zwar nicht gleich aber in demselben Raum platziert ist. In unserem Datenbeispiel ist es durch die "Prokrustes-Anpassung" zu einer Spiegelung bei 17 Bootstrapstichproben gekommen. Der Benutzer kann dies nachrechnen, wenn er als Zufallszahl 578125 in Prog30ml einsetzt.

Wir errechnen jetzt noch die Distanzen zwischen allen Ladungspunkten, zuerst aus der nicht-rotierten Faktorladungsmatrix der Bootstrapstichprobe 29 und dann die aus ihrer konfirmatorischen Ladungsmatrix. Die beiden Distanzmatrizen und auch die Matrizen der aus den beiden Ladungsmatrizen *reproduzierten Korrelationen* stimmen exakt überein. In der Programmaske Prog30ml muss zu diesem Zweck die Optionsbox "weitere Optionen" geöffnet werden. Siehe dazu den folgenden Abschnitt P30.6.36.

Die Distanzmatrix ist

vom Modell prognostizierte Matrix der euklidischen Distanzen zwischen den Variablen für die nicht-rotierte und die konfirmatorische Faktorladungsmatrix aus Bootstrapstichprobe 29 (nach 2 extrahierten Faktoren)

	t6_parag	t7_sente	t9_word_	t14_word	t15_numb	t17_obje
t6_parag	0	0.1560	0.0865	1.0380	1.1904	1.1072
t7_sente	0.1560	0	0.0788	1.1399	1.2781	1.1964
t9_word_	0.0865	0.0788	0	1.0696	1.2125	1.1302
t14_word	1.0380	1.1399	1.0696	0	0.1971	0.1308
t15_numb	1.1904	1.2781	1.2125	0.1971	0	0.0838
t17_obje	1.1072	1.1964	1.1302	0.1308	0.0838	0

Zu beachten ist, dass das "Spiegeln" durch das Verfahren der *konfirmatorischen Faktorenanalyse* kein Rotieren ist. Die gesamte Punktwolke in Graphik 3 kann nicht so im 2-dimensionalen Koordinatensystem gedreht werden, dass Graphik 4 entsteht. Räumlich kann man sich das eher als ein Kippen nach vorne um die Achse F1 vorstellen. Das wäre eine Rotation durch die nicht vorhandene 3. Dimension. Das Anpassen der nicht-rotierten Faktorladungsmatrix an die Zielmatrix ist kein *Rotieren*, in dem Sinne, in dem der Begriff der "Rotation" in der Theorie der Faktorenanalyse verwendet wird.

#### **P30.6.2.3.4 Konfirmatorische Faktorenanalyse**

Wenn wir im Folgenden hin und wieder trotzdem den Begriff der "Prokustes-Rotation" verwenden, dann meinen wir eigentlich "Prokrustes-Anpassung". Der Kalkül dafür ist die "konfirmatorische Faktorenanalyse". Verwendet wird dabei eine von Roppert/Fischer (1965) und Tarnai (1978) entwickelte "Kleinste-Quadrate-Schätzung". Der Kalkül wird in aller Ausführlichkeit beschrieben im

Almo-Dokument Nr 16, "Konfirmatorische Faktorenanalyse", Abschnitt P30.5.2

Wir verwenden folgende Notation

$F[i,k]$  = das ist die Ladung der Variablen  $i$  auf dem Faktor  $k$  der nicht-rotierten Ladungsmatrix aus der jeweils untersuchten Stichprobe. Dabei kann diese die Originalstichprobe oder eine der vielen Bootstrapstichproben sein  
 $Z[i,k]$  = das ist die Ladung der Variablen  $i$  auf dem Faktor  $k$  aus der Zielmatrix

Der Vorgang ist folgender:

1. Die Punktekonfiguration aus der nicht-rotierten Faktorladungsmatrix wird an die der Zielmatrix "heran gezwungen". Aus der Kleinste-Quadrate-Lösung nach Roppert/Fischer/Tarnai wird dazu eine Transformationsmatrix geliefert, die folgendes leistet.
2. Die quadrierte Distanz zwischen dem Punktepaar  $F[i,k]$  ---  $Z[i,k]$  wird für die untersuchte Stichprobe ermittelt. Die aufsummierte quadrierte Distanz aus allen

Paaren zueinander gehörender Punkte aus der Ladungsmatrix  $F$  und der Ladungsmatrix  $Z$  wird ein Minimum.

3. Dabei bleibt die Punktekonfiguration der nicht-rotierten Faktormatrix unverändert erhalten. D.h. die Distanzmatrizen der Variablenpunkte aus der nicht-rotierten Faktorladungsmatrix und der "konfirmatorischen" Faktorladungsmatrix sind identisch.

#### ***P30.6.2.3.5 Vertauschte Faktoren***

Die Bootstrapsstichproben können nicht nur auf Grund der oben erwähnten Spiegelung anders im Koordinaten-Raum liegen, sondern auch weil die Faktoren vertauscht sind. Es ist durchaus möglich, dass unter den vielen Bootstrapsstichproben einige wenige dabei sind, bei denen beispielsweise die 2. Variablengruppe (t14, t15, t17) dem 1. Faktor zugeordnet werden kann und nicht dem 2. Faktor, wie bei der Mehrzahl der Bootstrapsstichproben, und entsprechend die erste Variablengruppe dem 2. Faktor. Durch die Werte -1,0,1 in der Zielmatrix wird den Variablen "*gezeigt wo sie hin gehören*". Durch den Kalkül der konfirmatorischen Faktorenanalyse wird diese Faktoren-Vertauschung dann wieder rückgängig gemacht. In unserem Datenbeispiel (mit 9 Items aus den Holzinger-Swineford-Daten und 3 Faktoren) tritt der Fall der Faktoren-Vertauschung 209 mal auf. In Anhang 2 werden wir nochmals ausführlich auf das Thema der Faktoren-Vertauschung zurückkommen.

#### ***Zusammenfassung***

Die nicht-rotierte orthogonale Faktorladungsmatrix aus der Originalstichprobe und alle nicht-rotierten orthogonalen Faktorladungsmatrizen aus den Bootstrapsstichproben werden an eine vorgegebenen Zielmatrix angepasst. Dies geschieht gemäß dem Kalkül der konfirmatorischen Faktorenanalyse. Die Zielmatrix wird dabei von der Originalstichprobe vorgegeben. Sie wird abgeleitet aus der Rotation der originalen nicht-rotierten Faktorladungsmatrix in eine inhaltlich sinnvoll interpretierbare Position. Die Zielmatrix kann rechtwinklig (orthogonal) sein oder, wie wir im folgenden Abschnitt noch zeigen werden, auch "schiefwinklig".

#### ***P30.6.2.4 Zielmatrix mit rechtwinkligen Koordinatenachsen***

##### ***Abgeleitet aus der rotierten Faktorladungsmatrix***

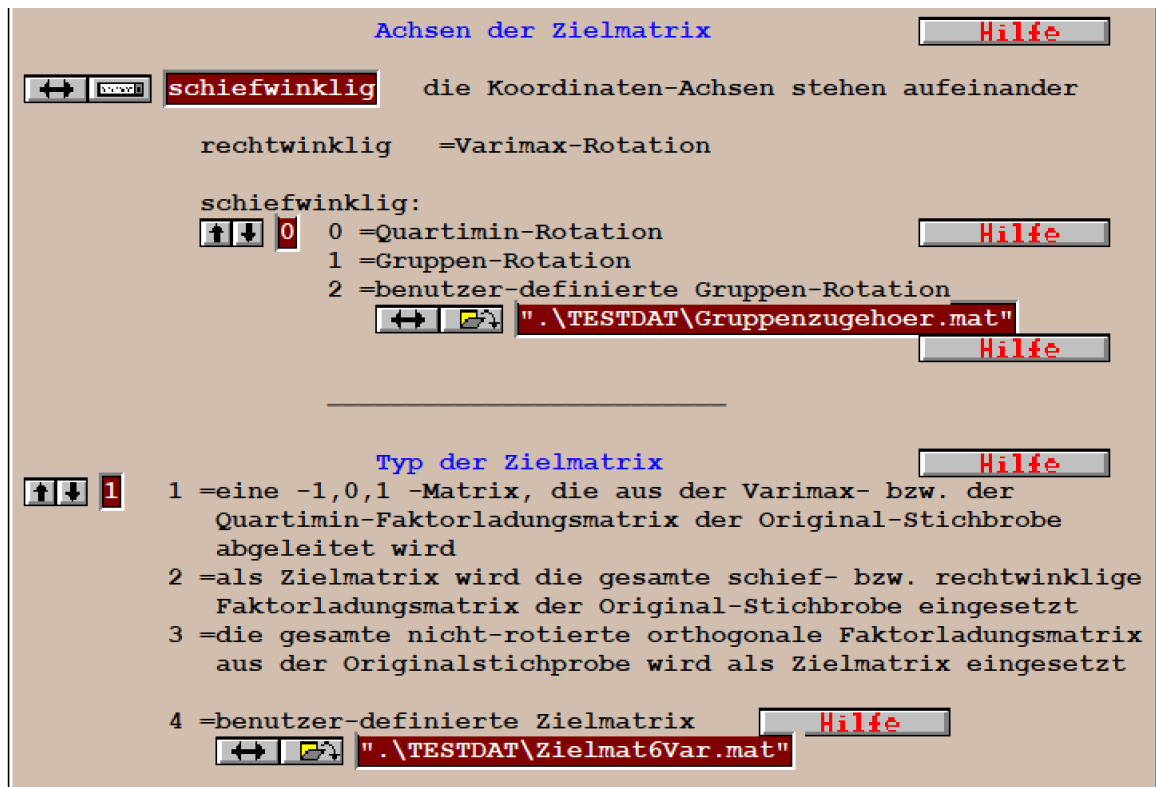
Die Zielmatrix beeinflusst die Bootstrap-Ergebnisse. D.h. der Standardfehler, der p-Wert und das Konfidenzintervall werden durch die Form und die Werte der Zielmatrix mitbestimmt.

Die Zielmatrix kann deswegen nicht beliebige Werte besitzen. Sie muss aus der Originalstichprobe abgeleitet werden. Das folgt aus diesen beiden Gründen: (1) Die Bootstrapsstichproben sind Zufallsstichproben aus den Originaldaten, d.h. jede von ihnen ist ein Repräsentant der Originaldaten. (2) Die Zielmatrix ist dieselbe für alle Bootstrapsstichproben.

Almo erlaubt es dem Benutzer eine Zielmatrix selbst zu entwickeln. Er muss sich dabei jedoch an der Faktorenstruktur der Originalmatrix orientieren. Wir empfehlen dem Forscher, bevor er sich für eine bestimmte Zielmatrix entscheidet, zuerst eine Faktorenanalyse ohne Bootstrap für die Originalstichprobe zu rechnen. Deren Ergebnisse, insbesondere die rechtwinklige varimax-rotierte bzw. schiefwinklig quartimin-rotierte Faktorladungsmatrix gilt es dann zu studieren.:

##### ***Eingabe-Box für die Zielmatrix***

Der Benutzer entscheidet sich für eine Zielmatrix durch zwei Klicks in der Bootstrap-Optionsbox. Wir zeigen hier den zutreffenden Ausschnitt aus der Optionsbox.



Die Eingabefelder der Box werden in Abschnitt P30.6.3.8 ausführlich erläutert.

### Die rechtwinklige Zielmatrix

Die Zielmatrix definiert das Koordinatenkreuz, in dessen Richtung die gegebene Faktorladungsmatrix gedreht, gespiegelt, "vertauscht" werden muss. Dazu muss eine Transformationsmatrix errechnet werden, die die eigentliche Aufgabe des Drehens etc. der gegebenen Matrix leistet.

Als Zielmatrix für die "Prokrustes-Anpassung" bietet Almo an

#### 1. Die "-1, 0, 1"-Zielmatrix

Sie entsteht in folgender Weise aus der Varimax-Faktorladungsmatrix der Originalstichprobe. Betrachten wir unser Beispiel

		Varimax-rotierte Faktorladungen		Zielmatrix	
		Faktor 1	Faktor 2	Faktor 1	Faktor 2
t6_parag	V14	0.8918	-0.1251	1.0000	0
t7_sente	V15	0.9110	-0.0117	1.0000	0
t9_word_	V17	0.8887	-0.0891	1.0000	0
t14_word	V22	0.1620	-0.7662	0	-1.0000
t15_num	V23	-0.0584	-0.7747	0	-1.0000
t17_obje	V25	0.0938	-0.7101	0	-1.0000

In jeder Zeile wird der absolut größte Wert gesucht. Ist er positiv wird 1.0 gesetzt, ist er negativ dann -1.0. Für die anderen Werte in der Zeile wird 0 eingesetzt. Die Zielmatrix ist also ein Faktorladungsmatrix, bei der die ersten 3 Variablen mit 1.0 auf dem 1. Faktor und mit 0 auf dem 2. Faktor laden. Die Variablen der zweiten Gruppe laden alle mit 0 auf dem

1. Faktor und mit -1.0 auf dem 2. Faktor. Die Zielmatrix ist gewissermaßen eine "perfekte Rotationsmatrix".

Bei der "selbst-definierten" Zielmatrix (die in Punkt 4 beschrieben wird) kann der Forscher nach eigenen Überlegungen, Werte einsetzen, z.B. für die Variable V22 eine Teilung in 0.200 und -0.800. Durch den Kalkül der konfirmatorischen Faktorenanalyse ist gewährleistet, dass die Struktur der "Ladungspunktwolke" nicht verändert würde, wenn die ursprüngliche Faktorladungsmatrix an die Zielmatrix heran gezwungen wird.

**2. Die gesamte Varimax-rotierte Faktorladungsmatrix der Original-Stichprobe wird selbst als Zielmatrix vorgegeben.**

Betrachten wir die Bootstrap-Ergebnisse, die bei den beiden Zielmatrizen entstehen. Wir zeigen nur die (gekürzten) Ergebnisse für den 1. Faktor.

Mit "orthogonal konfirmatorischer Faktorladung" bezeichnen wir in den folgenden Tabellen

- (a) bei der Originalstichprobe: die Faktorladungsmatrix (in Spalte \*c), die durch das "Heranzwingen" der unrotierten Faktorladungsmatrix an die Zielmatrix entstand. Bei Zielmatrix 2 ist sie identisch mit der Varimax-Faktorladungsmatrix.
- (b) In Spalt \*d ist der Mittelwert aus den orthogonal konfirmatorischen Faktorladungsmatrizen" der vielen Bootstrapstichproben. Die einzelnen Matrizen entstehen durch das "Heranzwingen" der unrotierten Faktorladungsmatrizen der Bootstrapstichproben an die Zielmatrix.

Ergebnisse für Faktor 1

identisch mit Varimax-rotierter  
Faktorladungsmatrix aus  
Originalstichprobe

**Zielmatrix: Varimax-Matrix**

	Ergebnisse: Bootstrap der orthogonal konfirmatorischen Faktorladungen Zielmatrix: Varimax-Faktorladung *i)						
	Original- stichprobe *c	Mitt.wert *d					
	orthogonal konfirmator. Faktorladung	orthogonal konfirmator. Faktorladung	*e Verzerrung Faktorladung	*f Standard fehler	*g Signifik. p	*h Konfidenzintervall unten oben	
t6	0.891791	0.890547	-0.001243	0.012172	0.000999	0.865511	0.912086
t7	0.911040	0.910527	-0.000513	0.009252	0.000999	0.890992	0.927371
t9	0.888700	0.887898	-0.000802	0.011227	0.000999	0.863386	0.908064
t14	0.162007	0.164313	0.002306	0.048695	0.002000	0.069456	0.260130
t15	-0.058437	-0.061201	-0.002764	0.060310	0.322000	-0.179785	0.055141
t17	0.093774	0.093609	-0.000165	0.062185	0.146000	-0.033762	0.215882

Ergebnisse für Faktor 1

**Zielmatrix: 1,0,-1 -Matrix**

	Ergebnisse: Bootstrap der orthogonal konfirmatorischen Faktorladungen Zielmatrix: -1,0,1 -Matrix aus Varimax *i)						
	Original- stichprobe *c	Mitt.wert *d					
	orthogonal konfirmator. Faktorladung	orthogonal konfirmator. Faktorladung	*e Verzerrung Faktorladung	*f Standard fehler	*g Signifik. p	*h Konfidenzintervall unten oben	
t6	0.892498	0.891228	-0.001269	0.012202	0.000999	0.865715	0.912623
t7	0.911092	0.910511	-0.000581	0.009238	0.000999	0.890842	0.927471
t9	0.889200	0.888348	-0.000852	0.011229	0.000999	0.864000	0.908476
t14	0.166428	0.168763	0.002335	0.048147	0.002000	0.074789	0.263801
t15	-0.053963	-0.056686	-0.002723	0.059856	0.342000	-0.175103	0.057562
t17	0.097873	0.097735	-0.000138	0.060882	0.122000	-0.026843	0.218787

Die Unterschiede sind minimal. Die Verzerrung (Spalte \*d minus Spalte \*c) und der Standardfehler sind geringfügig kleiner, also besser, wenn die gesamte Varimax-Matrix aus der Originalstichprobe als Zielmatrix vorgegeben wird. Das ist vermutlich deswegen der Fall, weil die "Punktwolken" der beiden Variablengruppen in unserem Beispiel deutlich voneinander getrennt sind und dies auch noch beinahe rechtwinklig.

### 3. Die nicht-rotierte orthogonale Faktorladungsmatrix der Original-Stichprobe wird selbst als Zielmatrix vorgegeben.

#### Ergebnisse für Faktor 1

Original-Stichprobe		Ergebnisse: Bootstrap der orthogonal konfirmatorischen Faktorladungen Zielmatrix: orthogonale Fakt.lad.matrix *i)					
*a	Mitt.wert *d	*e		*f	*g	*h	
unrotierte orthogonale Faktorladung	orthogonal konfirmator. Faktorladung	Verzerrung Faktorladung	Standard fehler	Signifik. p	Konfidenzintervall unten oben		
t6	0.875175	0.874221	-0.000954	0.013878	0.001000	0.843906	0.900316
t7	0.851345	0.850693	-0.000652	0.015909	0.001000	0.817106	0.881123
t9	0.859052	0.858153	-0.000899	0.015529	0.001000	0.826697	0.885510
t14	0.432765	0.434191	0.001426	0.040942	0.001000	0.355474	0.510214
t15	0.230952	0.228143	-0.002810	0.058766	0.001000	0.110313	0.338361
t17	0.348691	0.348002	-0.000689	0.055981	0.001000	0.230481	0.450823

Die aus den Bootstrapstichproben gewonnenen unrotierten orthogonalen Faktorladungsmatrizen werden an die unrotierte Originalstichprobe "heran gezwungen". Der "gemeinsame Faktorenraum" (siehe Abschnitt P30.6.2.2) ist also der der unrotierten Faktorenanalyse der Originalstichprobe. Das ist eine eher ungewöhnliche "Prokrustes-Anpassung", zu der noch keine Erfahrungen vorliegen. Sie sollte deswegen nur versuchsweise und nur "für den eigenen Gebrauch" verwendet werden.

### 4. Die selbst definierte orthogonale Zielmatrix

Betrachten wir folgende Situation: Die rechtwinklige Varimax-Rotation liefert kein befriedigendes Ergebnis. Der Forscher verfügt jedoch über eine Theorie (oder schlichter formuliert: eine Vermutung) bezüglich der Faktorstruktur der eingesetzten Variablen. In dieser Situation wird er versuchen, die Zielmatrix, die ihm der Varimax-Kalkül vorgelegt hat, zu verbessern. Er konstruiert selbst eine Zielmatrix.

Beispiel für eine benutzer-definierte Zielmatrix für 9 Variable und 3 Faktoren

0.8	0	0.2
0.7	0	0.3
1.0000	0	0
0	-1.0000	0
0	-1.0000	0
0	-1.0000	0
0	0	1.0000
0	0	1.0000
0	0	1.0000

Der Benutzer hat in diesem Beispiel eine -1,0,1-Matrix in den ersten beiden Zeilen verändert. Er muss diese Matrix mit einem beliebigen Namen in eine Datei speichern. Dazu erzeugt er ein neues Fenster über das Menü "Datei / Neue Datei anlegen" oder durch Klick auf das Symbol ganz links in der Knopfleiste am Oberrand des Almo-Fensters. Dann muss er noch einen Ordner auswählen und einen Dateinamen festlegen. In der Bootstrap-Optionsbox in der Sub-Box "Zielmatrix" (siehe oben) muss dann dieser Name eingetragen werden.

### P30.6.2.5 Zielmatrix mit schiefwinkligen Koordinatenachsen

#### 1. Die "-1, 0, 1"-Matrix

Sie wird mit derselben Vorgehensweise, wie oben für die orthogonale Zielmatrix gezeigt, aus der

quartimin-rotierten,  
achsparell projizierten Faktorladungsmatrix  
der Originalstichprobe

abgeleitet. Die Quartimin-Rotation ist weitgehend identisch mit dem Rotationsverfahren "Oblimin 0".

Zusätzlich wird den Bootstrapstichproben noch vorgegeben:

Die Matrix der Korrelationen  
zwischen den schiefwinkligen Achsen  
aus der Quartimin-Rotation der Originalstichprobe.

Das geschieht Algo-intern und automatisch.

		Quartimin-Matrix: Matrix der auf die schiefwinkligen Achsen achsparell projizierten Faktorladungen (Ladungsmatrix)		Zielmatrix -1, 0, 1 -Matrix	
		Faktor 1	Faktor 2	Faktor 1	Faktor 2
t6_parag	V14	0.8908	-0.0497	1.0000	0
t7_sente	V15	0.9200	0.0665	1.0000	0
t9_word_	V17	0.8908	-0.0136	1.0000	0
t14_word	V22	0.0979	-0.7606	0	-1.0000
t15_num	V23	-0.1257	-0.7882	0	-1.0000
t17_obje	V25	0.0337	-0.7098	0	-1.0000

Aus Quartimin-Rotation:  
Matrix der Korrelationen zwischen den schiefwinkligen Achsen

	Faktor 1	Faktor 2
Faktor 1	1.0000	-0.1689
Faktor 2	-0.1689	1.0000

Die Korrelation zwischen den Faktoren entspricht dem Kosinus des Winkels zwischen den schiefwinkligen Faktoren. Die Winkel sind:

Matrix der Winkel zwischen den schiefwinkligen Achsen

	Faktor 1	Faktor 2
Faktor 1	0	-80.2735
Faktor 2	-80.2735	0

Die -80.2735 Grad aus dieser Matrix werden durch die Graphik in Abschnitt P30.6.4.3 ersichtlich und das negative Vorzeichen verständlich.

Aus der Korrelationsmatrix der Faktoren und der Zielmatrix entsteht dann die eigentliche "schiefwinklige Zielmatrix", an die die Ladungsmatrizen der Bootstrapstichproben heran gezwungen werden.

**2. Die gesamte Quartimin-rotierte, achsparallel projizierte Faktorladungsmatrix aus der Originalstichprobe wird als Zielmatrix verwendet.**

Siehe oben die Quartimin-Matrix. Zusätzlich wird den Bootstrapsstichproben noch vorgegeben: Die Matrix der Winkel (bzw. Korrelationen) zwischen den schiefwinkligen Achsen aus der Quartimin-Rotation der Originalstichprobe. Das geschieht Almo-intern und automatisch.

Da in unserem Beispiel die schiefwinkligen Achsen für die beiden Variablengruppen, mit - 80 Grad fast orthogonal aufeinander stehen, ist das Ergebnis dem aus der orthogonalen Zielmatrix sehr ähnlich.

**3. Die selbst definierte schiefwinklige Zielmatrix**

Die Situation könnte folgende sein: Die schiefwinklige Quartimin-Rotation liefert dem Forscher kein befriedigendes Ergebnis. Er vermutet jedoch, dass die eingesetzten Variablen eine ganz bestimmte schiefwinklige Faktorenstruktur besitzen. In dieser Situation wird er versuchen, die Zielmatrix und die Korrelationsmatrix der schiefwinkligen Faktoren, die ihm der Quartimin-Kalkül vorgelegt hat, zu verbessern. Er konstruiert selbst eine Zielmatrix und eine Faktor-Korrelationsmatrix.

Beispiel für eine 9\*3 Zielmatrix mit angeschlossener Faktor-Korrelationsmatrix

0.8	0	0.2	
0.7	0	0.3	
1.0000	0	0	
0	-1.0000	0	<b>Zielmatrix</b>
0	-1.0000	0	
0	-1.0000	0	
0	0	1.0000	
0	0	1.0000	
0	0	1.0000	
1.0000	-0.4	0.1	
-0.4	1.0000	-0.3	<b>Faktor-Korrelationen</b>
0.1	-0.3	1.0000	<b>der 3 Achsen</b>

Der Benutzer erzeugt ein neues Fenster, in das er die beiden Matrizen speichert

Wichtig ist folgendes:

Die Zielmatrix ist orthogonal. Aus ihr und der Matrix der Faktor-Korrelationen wird dann intern, für den Benutzer nicht sichtbar, die schiefwinklige Zielmatrix errechnet.

**P30.6.2.6 Die gesamte Varimax- bzw. Quartimin-Faktorladungsmatrix als Zielmatrix**

Soll eine Faktorladungsmatrix mit 2 oder mehr Faktoren in Hinblick auf den untersuchten Gegenstand inhaltlich interpretierbar sein, so muss sie in aller Regel rotiert werden. In Frage kommt dafür in Almo die rechtwinklige Varimax-Rotation oder die schiefwinklige Quartimin-Rotation. Werden deren Ladungsmatrizen unmittelbar auch als Zielmatrizen für die unrotierten Ladungsmatrizen aus den Bootstrapsstichproben eingesetzt, dann hat das folgende Wirkung: Die inhaltlich interpretierbare varimax- bzw. quartimin-rotierte Ladungsmatrix aus der Originalstichprobe "gewinnt" durch das Bootstrapping einen Standardfehler, einen p-Wert und ein Konfidenzintervall - was ohne Bootstrapping seither nicht möglich war.

Wird die recht- bzw. schiefwinklige -1,0,1 - Matrix als Zielmatrix verwendet, dann muss auch die Faktorladungsmatrix aus der **Originalstichprobe** an diese Zielmatrix "herangezungen" werden. Das Ergebnis bezeichnen wir dann als *konfirmatorische*

Faktorladungs-matrix. Varimax- bzw. quartimin-rotierte einerseits und konfirmatorische Ladungsmatrix andererseits sind nicht mehr identisch. Die durch das Bootstrapping "gewonnenen" Standardfehler, p-Wert und Konfidenzintervall-Grenzen beziehen sich dann auf die *konfirmatorische* Faktorladungsmatrix aus der Originalstichprobe. Das ist nur dann unproblematisch, wenn die konfirmatorische Faktorladungsmatrix der Originalstichprobe hinsichtlich des Forschungsgegenstandes, sinnvoll interpretiert werden kann. In der Regel ist das der Fall, wenn - wie in unserem Beispiel - die Variablen den Faktoren eindeutig zugeordnet werden können.

### P30.6.2.7 Normierung der Faktorladungen

Anstelle „Normierung“ wird in der Literatur auch der Begriff „*Normalisierung*“ verwendet.

Almo bietet zwar eine Option an, die Faktorladungen zu normieren. Die Normierung hat jedoch einen ernst zu nehmenden Defekt, weswegen wir empfehlen, diese Option nicht standardmäßig zu nutzen. Diesen Defekt werden wir anschließend und besonders ausführlich im Anhang beschreiben. Die Normierung geschieht in folgender Weise:

In der orthogonalen, nicht-rotierten Faktorladungsmatrix der Original-Stichprobe und auch der Bootstrap-Stichproben wird für jede Matrixzeile *i* die Summe *s(i)* der quadrierten Ladungen ermittelt. Die Summe *s(i)* ist gleich der "reproduzierten Kommunalität" der Variablen *i*, die sich von der durch den Benutzer vorgegebenen unterscheidet. Jede Ladung der Zeile *i* wird dann durch die Wurzel aus *s(i)* dividiert. Die Summe der quadrierten normierten Faktorladung in einer Matrixzeile ist nunmehr 1.0. Das Ergebnis ist die zeilenweise normierte, orthogonale, nicht-rotierte Faktorladungsmatrix. Diese wird dann der "Prokrustes-Anpassung" unterzogen - woraus die konfirmatorische Faktorladung entsteht.

Alle *x* Faktorladungen für die Variable *i* werden durch die Wurzel aus *s(i)*, also durch ein Konstante gewichtet. Die Konstante hat jedoch für jede Zeile/Variable wieder einen anderen Wert. Der Ladungspunkt *i* wird also im Koordinatensystem um einen anderen Betrag verschoben als der Ladungspunkt *j*. Daraus folgt, dass die Normierung die "Konfiguration der Ladungspunkte" verändert. Die Distanzen zwischen den Ladungspunkten werden verändert.

Wir zeigen die Faktorladungsmatrizen und Distanzmatrizen für unser 2-faktorielles Beispiel der Holzinger-Swineford-Daten.

Faktorladungen aus Originalstichprobe						
	A Ausgangsmatrix unrotiert nicht-normiert		B konfirmatorisch prokrustes rotiert nicht-normiert		C konfirmatorisch prokrustes rotiert <i>normiert</i>	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2	Faktor 1	Faktor 2
t6_parag	0.8752	0.2121	0.8925	-0.1199	0.9901	-0.1407
t7_sente	0.8513	0.3246	0.9111	-0.0065	0.9999	-0.0148
t9_word_	0.8591	0.2444	0.8892	-0.0839	0.9948	-0.1016
t14_word	0.4328	-0.6527	0.1664	-0.7652	0.2051	-0.9788
t15_numb	0.2310	-0.7418	-0.0540	-0.7750	-0.0771	-0.9970
t17_obje	0.3487	-0.6257	0.0979	-0.7096	0.1291	-0.9916

Die Distanzmatrizen für unser Beispiel sind 6\*6 – Matrizen. Wir zeigen von Ihnen nur die 1. Spalte.

Distanzmatrizen aus Originalstichprobe (1. Spalte)

- 
- 1 Distanzmatrix der Ausgangsmatrix A  
unrotiert und nicht-nnormiert
  - 2 Distanzmatrix der konfirmatorischen Faktorladungsmatrix B  
prokrustes-rotiert und nicht-nnormiert
  - 3 Distanzmatrix der konfirmatorischen Faktorladungsmatrix C  
prokrustes-rotiert und *normiert*

	1	2	3
	t6_parag V14	t6_parag V14	t6_parag V14
t6_parag	0	0	0
t7_sente	0.1149	0.1149	0.1263
t9_word_	0.0361	0.0361	0.0394
t14_word	0.9714	0.9714	1.1483
t15_numb	1.1511	1.1511	1.3682
t17_obje	0.9895	0.9895	1.2105

Die Distanzen in der Matrix 3 müssten dieselben sein wie die in der Ausgangsmatrix 1. Das ist nicht der Fall. Wird nicht normiert dann stimmen die Distanzen in 1 und 2 überein.

Die Normierung verletzt die zentrale Bedingung des Rotierens, dass die Konfiguration der Ladungspunkte nicht verändert werden darf, d.h. genauer formuliert, dass die aus der "prokrustes-angepassten" konfirmatorischen Faktorladungsmatrix errechnete *Distanzmatrix* der Ladungspunkte (sowie auch die *reproduzierte Korrelationsmatrix*) identisch sein muss mit der, die aus der ursprünglichen nicht-rotierten orthogonalen Faktorladungsmatrix hervorgegangen ist. Durch die Normierung werden die Distanzmatrix und die reproduzierte Korrelationsmatrix aus unrotierter Faktorladungsmatrix und "prokrustes-angepasster", konfirmatorischer Faktorladungsmatrix ungleich.

Zientek/Thompson (2007) rechnen prinzipiell und nur mit normierten Faktorladungen. Wir werden das im Anhang zeigen. Der Grund warum normiert wird ist vermutlich folgender. Es wird angenommen, dass die Faktorladungsmatrizen der Originalstichprobe und aller Bootstrapstichproben einheitlicher werden und dass es damit gelingt, sie besser in einen "gemeinsamen Faktorenraum" zu stellen. Das mag zwar gelingen, ist aber an unserem Beispiel nicht erkennbar.

### P30.6.3 Bootstrap-Programm Prog30ml.Msk

Prog30ml.Msk  
Faktorenanalyse quantitativer Variablen  
mit Bootstrap

Das Programm rechnet eine Standard-Faktorenanalyse, die durch eine Vielzahl von Optionen abgeändert werden kann. So können die Faktorladungen recht- oder schiefwinklig rotiert werden. Es können Kommunalitäten-Schätzwerte eingesetzt werden und die Faktorenanalyse auch als Alpha- oder Image-analyse oder als kanonische Faktorenanalyse gerechnet werden.

Danach wird ein Bootstrap-Verfahren für x (z.B. 1000) Bootstrapstichproben gerechnet, dem wahlweise die Faktorladungen oder die Faktorwert-Koeffizienten unterworfen werden, so dass für diese beiden Koeffizienten verteilungsfreie Standardfehler, p-Werte und Konfidenzintervalle ermittelt werden können.

---

weitere Beispiel-Programme: laden durch  
Doppelklick

sehr schwach korrelierende Variable  
|

Bootstrap mit 1 Faktor


Siehe Almo-Dokument 15 "Faktorenanalyse" und  
15a "Bootstrap bei Faktorenanalyse"


Programm-Bedienung --->

Vereinbare Variable=


**Variablenamen**







Datei der Variablenamen


 ".\TESTDAT\HolzingerSwineford1939.nam"

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



**Freie Namensfelder**

 Hole alle Eingaben wieder zurueck


   
   
 


 [...]  
erzeuge zusätzliche Namensfelder


**Variablenamen in Datei speichern**   
Eingabefeld leer = nicht speichern


 

**Datei aus der gelesen wird**   
bei Datei-Problemen


 ".\TESTDAT\HolzingerSwineford1939.fre"










 **frei** Format der Daten

 **V1:34** der Datensatz enthält diese Variablen  
Bei Format DIREKT schreiben Sie: alle\_V

 **Wenn Dateiformat FIX oder Nicht-Standard-FREI**

**zu faktorisierende quantitative Variable**

 **V14,15,17, #18,20,21,# 22,23,25**

-  Option: Ein- und Ausschliessen von Untersuchungseinheiten
-  Option: Umkodierungen und Kein-Wert-Angaben
-  Option: Spezielle Kein-Wert-Behandlung
-  Option: Untersuchungseinheiten gewichten
-  Option: Faktorenanalytisches Modell und Verfahren
-  Option: Faktoren und Kommunalitäten
-  Option: weitere Optionen
-  Option: "Aussehen" der auszugebenden Tabelle bzw. Matrix
-  Grafik-Optionen

<input type="checkbox"/> X		Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)	
Option: Bootstrap			
<input type="checkbox"/>	<input type="checkbox"/>	1	1 =Bootstrap ausführen 0 =nicht
<input type="checkbox"/>	<input type="checkbox"/>	1	1 =Bootstrap für Faktorladungen 2 =für Faktorwertkoeffizienten
<input type="checkbox"/>	<input type="checkbox"/>	0	Faktorladungen normieren 0 =nein 1 =ja
<input type="checkbox"/>	<input type="checkbox"/>	1000	wieviele Stichproben sollen gerechnet werden
Konfidenzintervall			
<input type="checkbox"/>	<input type="checkbox"/>	95.00	Konfidenzniveau in %
<input type="checkbox"/>	<input type="checkbox"/>	0	Konfidenzintervall berechnen mit .... 0 = einfachem Perzentil-Verfahren 1 = nicht möglich 2 = nicht möglich 3 = Perzentil-Verf. BC (bias corrected) 4 = Perzentil-Verf. BCa (bias corrected accelerated)
<input type="checkbox"/>	<input type="checkbox"/>	578125	Startzahl Zufallsgenerator
Zielmatrix			
Spezielle Ergebnisse ausgeben			

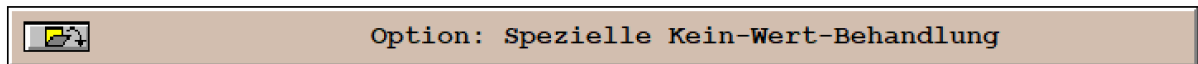
Die Eingabeboxen von Prog30ml.Msk sind weitgehend identisch mit denen des Faktorenanalyse-Standardprogramms Prog30m2.Msk. Sie werden ausführlich im Almo-Dokument 15 "Faktorenanalyse" bzw. Dokument 0 "P0 Arbeiten mit Almo" beschrieben. Wir werden im Folgenden nur einige wenige Boxen erläutern.

### P30.6.3.1 Zu faktorisierende Variable

zu faktorisierende quantitative Variable		
<input type="checkbox"/>	<input type="checkbox"/>	V14,15,17, #18,20,21,# 22,23,25

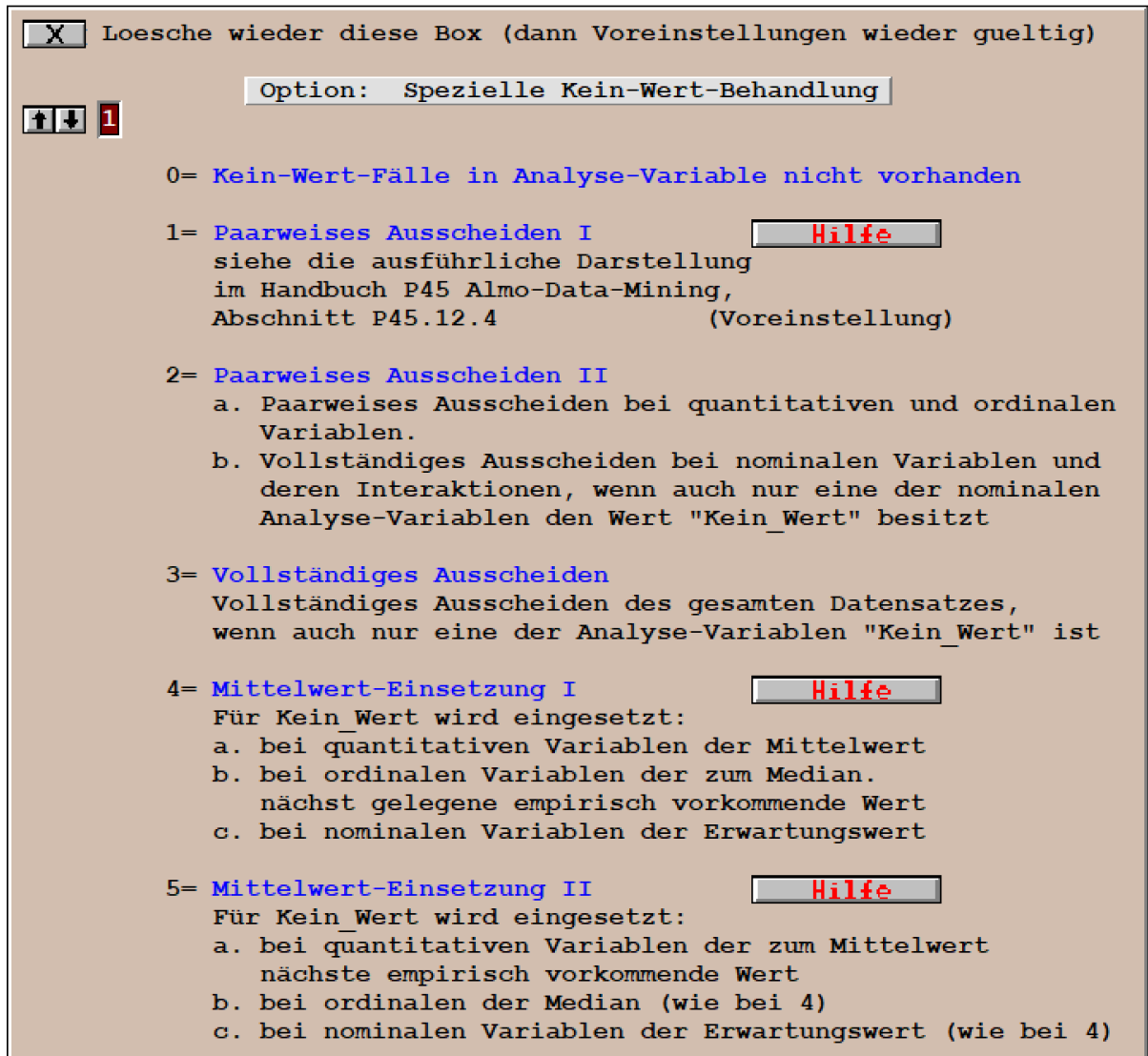
Die Nummern oder Namen der zu faktorisierenden Variablen werden eingegeben. Wir haben hier die 6 Variablen aus unserem Datenbeispiel von Holzinger/Swineford eingegeben. Sie formieren ein deutlich 2-faktorielles System. Das 3-faktorielle System, das wir im Text auch verwenden, entsteht, wenn die 3 Variablen 18, 20, 21 eingefügt werden - indem die Kommentarzeichen #...# gelöscht werden.

### P30.6.3.2 Spezielle Kein-Wert-Behandlung



In Almo ist die Kein-Wert-Behandlung des "paarweisen Ausscheidens" voreingestellt. D.h. wird die Optionsbox nicht geöffnet, dann ist diese KW-Behandlung wirksam.

Wird die Optionsbox geöffnet dann sieht man folgendes. Wir zeigen nur den oberen Teil dieser sehr umfangreichen Optionsbox.



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)

Die Optionsbox wird ausführlich erläutert in Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.3.2. Diese sieben KW-Behandlungen wirken auf die Originalstichprobe ein, aber auch auf die Bootstrapstichproben. Dies in folgender Weise:

*KW-Behandlung 1 bis 3*

Die KW-Behandlungen 1 und 2 sind Varianten des "paarweisen Ausscheidens". Die KW-Behandlung 3 ist das "vollständige Ausscheiden". In die oben in Abschnitt P30.6.1.1 beschriebene Datenmatrix, die aus den Datensätzen der Originalstichprobe aufgebaut ist, wird in die entsprechende Datenmatrix-Zelle "kw" eingesetzt (bzw. der Almo-interne Code für Kein-Wert). Wird im Rechenverlauf die Bootstrapstichprobe i aus der Datenmatrix entnommen, dann wird, wenn Almo auf "kw" trifft, entsprechend der aktivierten KW-Behandlung verfahren. Bei KW-Behandlung 3 beispielsweise wird der gesamte Datensatz nicht ausgewertet.

*KW-Behandlung 4 bis 7*

Dies sind Varianten der Mittelwert-Einsetzung. Besitzt eine Variable in der *Originalstichprobe* den Wert "kw", dann wird "kw" durch den Variablen-Mittelwert ersetzt - und der Datensatz des betreffenden Probanden in die interne Datenmatrix eingetragen. Aus der internen Datenmatrix werden dann Datensätze für die Bootstrapstichproben entnommen. Das hat folgende Konsequenz:

In allen Bootstrapstichprobe, die (nach dem Zufallsprinzip) aus der internen Datenmatrix gezogen wurden, besitzt die Variable, die ursprünglich in der Originalstichprobe "kw" war, den Mittelwert für diese Variable aus der Originalstichprobe - also nicht den Mittelwert aus der Bootstrapstichprobe. Der Mittelwert aus der Originalstichprobe ist sicherlich der bessere Schätzer für den fehlenden Wert als der aus der Bootstrapstichprobe. Wir empfehlen deswegen KW-Behandlung 1 oder 2 oder 3 zu wählen. Almo bringt eine Warnung, wenn eine KW-Behandlung von 4 bis 7 gewählt wurde. Siehe dazu Abschnitt P30.6.1.1, Unter-Abschnitt "Datenmanipulationen in den Bootstrapstichproben, das Analogie-Modell" und "Spezielle Kein-Wert-Behandlung".

**P30.6.3.3 Optionen für Faktorenanalyse**

Werden die im Folgenden beschriebenen Optionsboxen nicht geöffnet, dann wird eine

Standard-Faktorenanalyse gerechnet. Sie ist durch folgende Einstellungen charakterisiert:

Einstellungen für Standard-Faktorenanalyse

Modell: "normale" Faktorenanalyse  
Eigenwert-Kalkuel: Tridiagonal-Qr-Verfahren  
Kommunalitäten: in Diagonale der Korrelationsmatrix steht 1.0  
Kommunalitäten-Iteration: 0 Iterationen  
Faktoren: Faktoren, deren Eigenwert größer 1.0

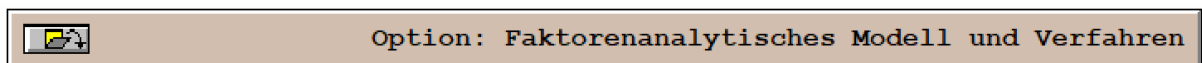
bei Bootstrap

Rotation: rechtwinklige Varimax-Rotation bei der Originalstichprobe, keine bei Bootstrapstichproben  
Zielmatrix: eine -1,0,1 -Matrix, gebildet aus der Varimax-Rotation der Originalstichprobe

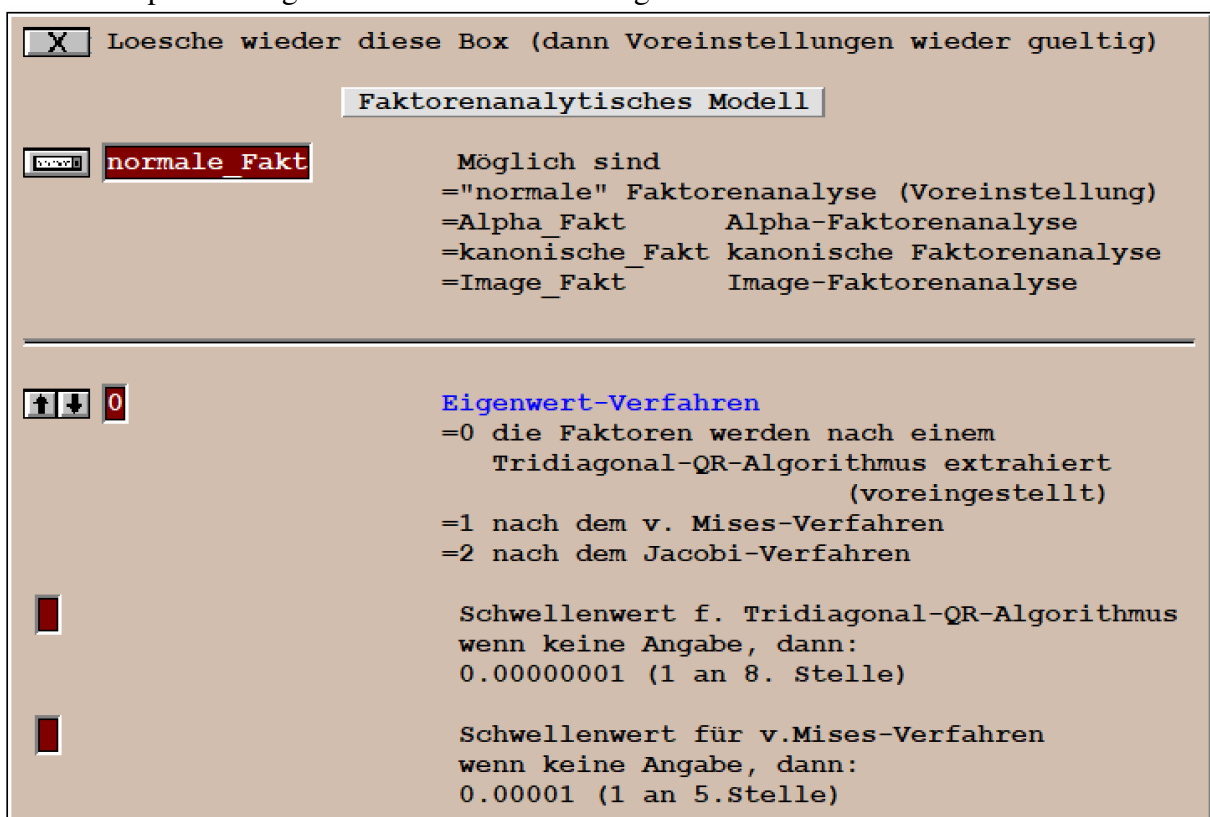
Die für das Bootstrapping notwendigen Einstellungen, *Rotation* und *Zielmatrix*, werden in der Bootstrap-Optionsbox selektiert. Das wird im Folgenden noch ausgeführt.

Das Standardmodell der Faktorenanalyse kann in den folgenden Optionsboxen verändert werden

**P30.6.3.4 Faktorenanalytisches Modell und Eigenwert-Verfahren**



Wird die Optionsbox geöffnet dann sieht man folgendes.



*Eingabefeld 1:*

Neben der "normalen" Faktorenanalyse können in Almo noch folgende 3 Sonder-Modell gerechnet werden

die Alpha-Faktorenanalyse

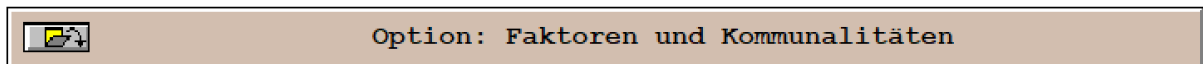
die kanonische Faktorenanalyse  
die Image-Faktorenanalyse

Werden sie eingesetzt, dann sollten die folgenden Optionen (siehe nächste Optionsbox) nicht verwendet werden, d.h. die Faktorenzahl und die Kommunalität sollte vom Benutzer *nicht* vorgegeben werden. Sie können jedoch vorgegeben werden - mit dem Risiko, dass Komplikationen entstehen. Die kanonische Faktorenanalyse erbringt Ergebnisse, die denen der Maximum-Likelihood-Faktorenanalyse, die in Almo nicht enthalten ist, sehr ähnlich sind.

#### *Eingabefeld 2:*

Das mathematische Eigenwert- Eigenvektor-Verfahren kann gewählt werden. Das Tridiagonal-QR-Verfahren und das Jacobi-Verfahren sind gleichwertig. Dabei kann es geschehen, dass das eine Verfahren gegenüber dem anderen einen oder mehrere Faktoren spiegelt (d.h. das Vorzeichen umkehrt). Zum Begriff des "Spiegelns" siehe das Beispiel der Bootstrapstichprobe 30 in P30.6.2.3. Das v.Mises-Verfahren ist sehr zeitintensiv und weniger genau.

#### ***P30.6.3.5 Faktorenzahl und Kommunalitätenschätzung***



Wird die Optionsbox geöffnet dann sieht man folgendes.

X Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

**Faktoren und Kommunalitäten**

**Faktoren**

Faktorenzahl  
 =leer: Almo überlassen (Voreinstellung)  
 =eine Zahl

---

**Faktorensignifikanz-Test**   
 =0 keine Signifikanz-Prüfung  
 =1 Test nach Rippe  
 =2 Test nach Lawley

---

**Kommunalitäten**

Wahl nur bei "normaler" Faktorenanalyse möglich

**1**  
 1= 1.0 In die Diagonale wird 1.0 eingesetzt  
 dabei werden alle Faktoren mit Eigenwert  
 grösser 1.0 extrahiert (Voreinstellung)

2= R Quadrat (multiples Bestimmtheitsmaß)  
  **0** 0 = dabei werden alle Faktoren mit Eigenwert  
 grösser 0 extrahiert  
 1 = grösser 1 extrahiert

3= maximale Korrelation  
 In die Diagonale wird die maximale Korrelation  
 (aus der Spalte) eingesetzt.

---

Zahl der Kommunalitäten-Iterationen   
 wenn keine Angabe, dann  
 0 Iterationen bei normaler und Image-Fakt.analyse  
 10 Iterationen bei Alpha- und kanonischer Fakt.

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

Siehe Almo-Dokument 16 "Faktorenanalyse"  
 zu "Faktorenzahl" Abschnitt P30.3.7.1  
 zu "Faktorensignifikanz" Abschnitt P30.3.7.2  
 zu "Kommunalitäten" Abschnitt P30.1.7

Faktorenzahl und Kommunalitätenschätzung können zwar unabhängig voneinander festgelegt werden. In Almo ist jedoch eine Automatik eingebaut, durch die die beiden verknüpft werden.

Bleibt das Eingabefeld für die Faktorenzahl *leer*, dann verfährt Almo beim Modell der "normalen" Faktorenanalyse so

- a. Wird die Optionsbox nicht geöffnet, dann setzt Almo automatisch "1.0" als Kommunalitätenschätzung ein. Identisch dazu ist natürlich der Fall, dass in das 2. Eingabefeld "1" eingesetzt wird. Es soll also eine "Hauptkomponentenanalyse"

gerechnet werden mit der Besonderheit, dass nur jene Faktoren extrahiert werden, deren Eigenwerte größer 1.0 sind. Dies wird als "**Kaiser-Kriterium**" bezeichnet.

Diese so ermittelte Faktorenzahl wird dann den Faktorenanalysen der Bootstrapstichproben vorgegeben. Wird in einer Bootstrapstichprobe die vorgegebene Faktorenzahl nicht erreicht, dann wird sie ausgeschieden und als Ersatz eine neue Bootstrapstichprobe aus der Datenmatrix (siehe P30.6.1.1) gezogen. Also meldet diesen Fall. Ist die Faktorenzahl der Bootstrapstichprobe größer als die der Originalstichprobe, dann werden die überzähligen Faktoren nicht verwendet.

**b.** Wurde die Optionsbox geöffnet und im 2. Eingabefeld die Diagonale auf *R\_QUADRAT* gesetzt, 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

Erst nachdem die Faktorenzahl festgelegt wurde, werden die multiplen Bestimmtheitsmaße in die Diagonale der Korrelationsmatrix eingesetzt und die Faktorenanalyse gerechnet. Dies ist ein modifiziertes "**Guttman-Kriterium**".

Wie bei Punkt a schon beschrieben, wird diese so ermittelten Faktorenzahl dann den Faktorenanalysen der Bootstrapstichproben vorgegeben. Wird in einer Bootstrapstichprobe die vorgegebene Faktorenzahl nicht erreicht, dann wird sie ausgeschieden und als Ersatz eine neue Bootstrapstichprobe aus der Datenmatrix (siehe P30.6.1.1) gezogen. Also meldet diesen Fall. Ist die Faktorenzahl der Bootstrapstichprobe größer als die der Originalstichprobe, dann werden die überzähligen Faktoren nicht verwendet.

Beim 2-faktoriellen Beispiel mit den Holzinger-Swineford-Daten erhielten wir folgende Bootstrap-Ergebnisse für Faktor 1.

		Standard fehler	Signifik. p	Konfidenzintervall		
				unten	oben	Breite
Diagonale 1.0	t6_paragraph_co	0.012202	<u>0.000999</u>	0.865715	0.912623	0.046908
	t7_sentence	0.009245	<u>0.000999</u>	0.890842	0.927471	0.036630
	t9_word_meaning	0.011234	<u>0.000999</u>	0.864000	0.908476	0.044476
	t14_word_recogn	0.048154	0.002000	0.074789	0.263801	0.189012
	t15_number_reco	0.059857	0.342000	-0.175103	0.057562	0.232666
	t17_object_num	0.060899	0.122000	-0.026843	0.218787	0.245630
-----						
Diagonale R-Quadrat	t6_paragraph_co	0.021423	<u>0.000999</u>	0.773263	0.855600	0.082337
	t7_sentence	0.017368	<u>0.000999</u>	0.805620	0.874595	0.068974
	t9_word_meaning	0.020009	<u>0.000999</u>	0.764696	0.844060	0.079363
	t14_word_recogn	0.041463	0.002000	0.077188	0.237933	0.160744
	t15_number_reco	0.050615	0.848000	-0.111899	0.083597	0.195496
	t17_object_num	0.049034	0.042000	0.004133	0.196353	0.192219

Die von den beiden Methoden gelieferten Ergebnisse sind etwas verschieden.

**c.** Bei der Kommunalitätenschätzung "maximale Korrelation" werden die Faktoren mit Eigenwert größer 1 (aus der Korrelationsmatrix mit 1.0 in der Diagonale) extrahiert. Danach

werden die maximalen Korrelationen aus der jeweiligen Zeile/Spalt in die Diagonale der Korrelationsmatrix eingesetzt und die Faktorenanalyse gerechnet.

Der Benutzer ist nicht gezwungen, die Faktorenzahl nach dem Kaiser-Kriterium bzw. dem Guttman-Kriterium zu übernehmen. Es gilt das Prinzip:

Legt der Benutzer selbst eine Faktorenzahl fest, dann wird dieser vom Programm auch entsprochen

und diese Faktorenzahl für die Originalstichprobe und die Bootstrapstichproben vorgegeben. Selbstverständlich können nur Faktoren mit Eigenwert  $> 0$  extrahiert werden.

**d.** Kommunalitäten-Iteration. Siehe Almo-Dokument 15 "Faktorenanalyse", P30.1.7.

Werden die Kommunalitäten iteriert, dann kann es geschehen, dass sich die Zahl der extrahierten Faktoren gegenüber dem vorherigen Iterationszyklus erhöht oder verringert. Bei den Bootstrap-Stichproben verfährt Almo folgendermaßen

- Werden mehr Faktoren extrahiert als aus der Originalstichprobe vorgegeben wurden, dann werden diese negiert.
- Ist die Faktorenzahl geringer, dann wird die Bootstrapstichprobe verworfen und durch eine neue Bootstrapstichprobe ersetzt. Almo bringt dann eine Warnung. Ereignet sich dieser Fall zu oft (bei der Hälfte aller Bootstrapstichproben), dann wird das Bootstrap-Verfahren abgebrochen

Wir empfehlen auf Iterationen zu verzichten - da ohnehin nicht nachweisbar ist, dass dadurch die Kommunalitäten besser geschätzt werden können

Bei der kanonischen und Alpha-Faktorenanalyse ist die Iteration ein unverzichtbarer Bestandteil des Kalküls.

Bei Alpha-, Image- und kanonische Faktorenanalyse verfährt Almo so:

Bei Alpha-, Image- und kanonische Faktorenanalyse wird die Faktorenzahl und die Kommunalität durch den Kalkül festgelegt. Man sollte für diese also das Eingabefeld leer lassen.

Kanonische Faktorenanalyse und Image-Faktorenanalyse verwenden  $R_{\text{Quadrat}}$  in der Diagonale der Korrelationsmatrix als Kommunalitätenschätzung und die Faktoren, deren Eigenwerte  $> 0$  sind. Die Alpha-Faktorenanalyse verwendet 1.0 in der Diagonale als Kommunalitätenschätzung und die Faktoren, deren Eigenwerte  $> 1.0$  sind.

Wird aber doch eine Faktorenzahl oder Kommunalitätenschätzung vom Benutzer festgelegt, die nicht vom Modell vorgesehen ist, dann wird dieser Anweisung vom Programm entsprochen. Selbstverständlich können nur Faktoren mit Eigenwert  $> 0$  extrahiert werden.

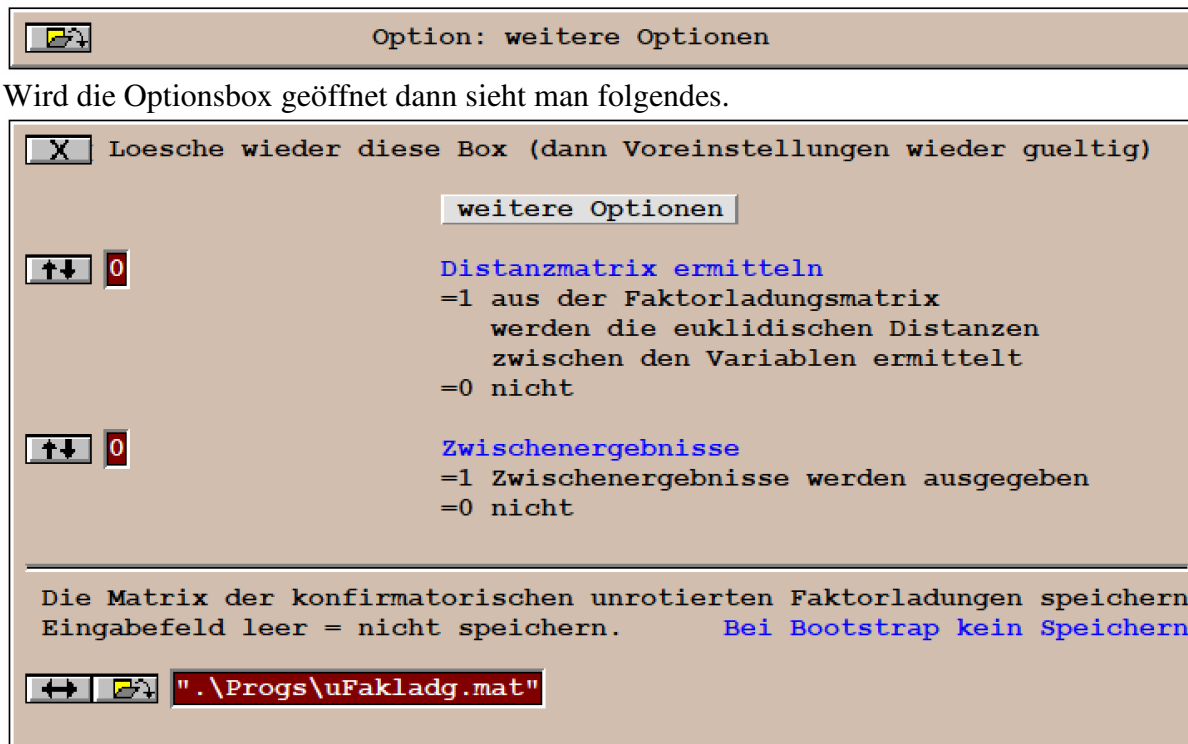
Bei kanonischer und Alpha-Faktorenanalyse werden die eingesetzten Kommunalitäten iteriert bis eine Iterationsdifferenz von 0.001 unterschritten wird. Nach maximal 10 Iterationszyklen wird abgebrochen. In den beiden letzten Eingabefeldern können diese voreingestellten Werte verändert werden.

*Für das Bootstrap-Verfahren gilt prinzipiell:*

Die Zahl der Faktoren, die für die Originalstichprobe extrahiert wurde, wird den Bootstrap-Stichproben vorgegeben. Kann für eine Bootstrapstichprobe nur eine kleinere Faktorenzahl ermittelt werden, dann wird sie aus der Analyse ausgeschieden. Als Ersatz

wird eine neue Stichprobe aus der Datenmatrix (siehe P30.6.1.1) gezogen. Ist die Faktorenzahl der Bootstrapsstichprobe entsprechend dem Kaiser- bzw. modifizierten Guttman-Kriterium größer als die der Originalstichprobe, dann werden die überzähligen Faktoren nicht verwendet.

### P30.6.3.6 Distanzmatrix ermitteln und Zwischenergebnisse ausgeben



#### Eingabefeld 1: Distanzmatrix

Die euklidischen Distanzen zwischen den Ladungspunkten werden berechnet und ausgegeben.

Die Faktorladungen in der nicht-rotierten Ausgangs-Faktorenladungsmatrix können als "Punktwolke" in einem mehrdimensionalen Raum begriffen werden. Sie kann bis zu 3 Dimensionen auch graphisch abgebildet werden. Jeder "Ladungs-Punkt" in der Punktwolke hat zu den anderen Punkten eine bestimmte räumliche Distanz. Durch das Rotieren der Ausgangsmatrix bzw. die "Prokrustes-Anpassung" dürfen diese Distanzen nicht verändert werden. Den Vorgang des Rotierens kann man so begreifen:

- (1) die Punktwolke ist in ein mehrdimensionales Koordinatensystem eingebettet, (2) das Koordinatensystem wird um seinen Ursprung, der beibehalten wird, in allen Dimensionen gedreht, gespiegelt und "vertauscht" - so wie das in P30.6.2.2 beschrieben wurde.
- (2) Die Punktwolke bleibt dabei fix. Die Distanzen zwischen den Punkten bleiben unverändert

#### Eingabefeld 2: Zwischenergebnisse

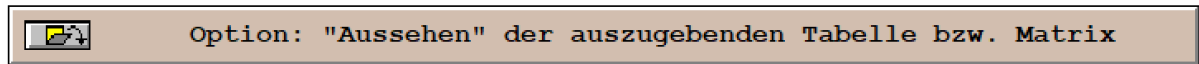
Wird "1" eingegeben, dann werden verschiedene Zwischenergebnisse des Kalküls ausgegeben. Bedeutsam ist, dass die durch die Faktorladungsmatrix "reproduzierte Korrelationsmatrix" ausgegeben wird-

**BEACHTTE:** Die Distanzmatrix und die Zwischenergebnisse werden nur für die Originalstichprobe und diejenigen Bootstrapsstichproben ausgegeben, die der Benutzer in der

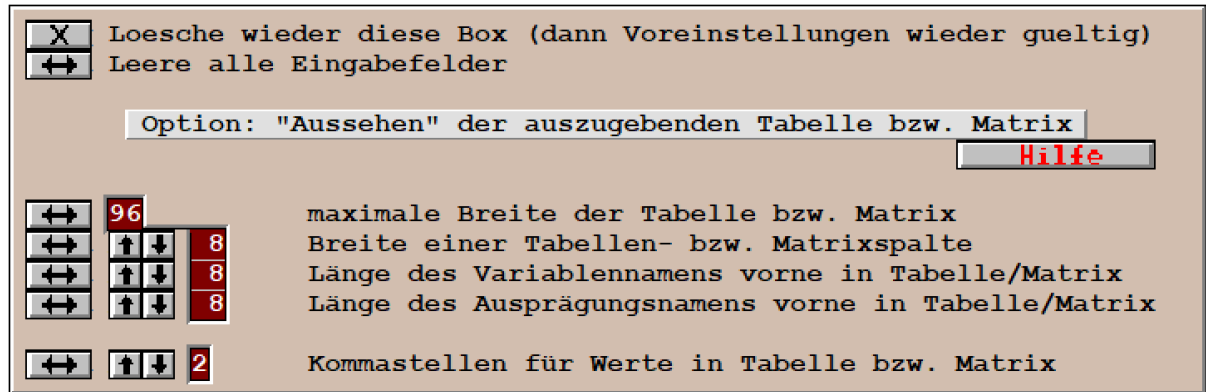
Bootstrap-Optionsbox in der Sub-Box „Spezielle Ergebnisse ausgeben“ als solche bezeichnet, deren Ergebnisse er sehen möchte.

### P30.6.3.7 Das "Aussehen" der Ergebnisse

Das "Aussehen" der Ausgabe kann über die folgende Optionsbox beeinflusst werden



Wird die Box geöffnet, dann sieht man folgendes



Wenn der Benutzer (im Programm, nicht hier) auf den Hilfeknopf klickt, dann werden ihm die verschiedenen Ausgabevarianten erläutert. Die angebotenen Gestaltungsmöglichkeiten gelten vorrangig für die Tabellen der Originalstichprobe.

Die Bootstrap-Ergebnis-Tabelle hingegen ist in ihrem Aufbau nicht beeinflussbar. Lediglich die Zahl der Kommastellen der Tabellenwerte kann verändert werden. Die Voreinstellung (bei geschlossener Optionsbox) ist 4. Maximal möglich sind 9 Kommastellen. Werden mehr angefordert dann gibt es Verschiebungen.

### P30.6.3.8 Die Bootstrap-Optionsbox

<input checked="" type="checkbox"/>	Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)	
Option: Bootstrap		
<input type="checkbox"/>	1	1 =Bootstrap ausführen 0 =nicht
<input type="checkbox"/>	1	1 =Bootstrap für Faktorladungen 2 =für Faktorwertkoeffizienten
<input type="checkbox"/>	0	Faktorladungen normieren 0 =nein 1 =ja
<input type="checkbox"/>	1000	wieviele Stichproben sollen gerechnet werden
Konfidenzintervall		
<input type="checkbox"/>	95.00	Konfidenzniveau in %
<input type="checkbox"/>	0	Konfidenzintervall berechnen mit .... 0 = einfachem Perzentil-Verfahren 1 = nicht möglich 2 = nicht möglich 3 = Perzentil-Verf. BC (bias corrected) 4 = Perzentil-Verf. BCa (bias corrected accelerated)
<input type="checkbox"/>	578125	Startzahl Zufallsgenerator
<input type="checkbox"/>	Zielmatrix	
<input type="checkbox"/>	Spezielle Ergebnisse ausgeben	

*Eingabefeld 1: Bootstrap 1=ausführen, 0=nicht*

Wird 0 eingegeben oder die Optionsbox geschlossen, dann wird eine Analyse nur für die originale Stichprobe gerechnet. Das Programm verhält sich dann so, wie das Standard-Programm für die Faktorenanalyse Prog30m2.Msk

*Eingabefeld 2: Bootstrap der Faktorladungen oder der Faktorwertkoeffizienten*

Bei Eingabe von "1" oder "2" wird dieselbe Analyse gerechnet - mit dem Unterschied, dass

- bei "1" die Faktorladungen dem Bootstrap-Verfahren unterworfen werden und

- bei "2" die aus den Faktorladungen abgeleiteten Faktorwertkoeffizienten - nach dem Hauptkomponenten-Verfahren bzw. dem Regressions-Verfahren. Siehe dazu Almo-Dokument 15 "Faktorenanalyse", Abschnitt P30.3.9. Es werden wie bei "1" die konfirmatorischen Faktorladungsmatrizen für die Original- und die Bootstrap-Stichproben ermittelt. Zusätzlich werden dann noch die Faktorwertkoeffizienten

errechnet und für jede Bootstrapstichproben gespeichert. Aus den gespeicherten Faktorwertkoeffizienten wird dann der Bootstrap-Mittelwert, der Standardfehler, der p-Wert und die Konfidenzgrenzen errechnet.

#### *Eingabefeld 3: Faktorladungen normieren*

0=nein

1=ja

##### Normierung der Faktorladungen

In der orthogonalen, nicht-rotierten Faktorladungsmatrix der Original-Stichprobe und auch

der Bootstrap-Stichproben wird für jede Matrixzeile bzw Variable  $i$  die Summe  $s(i)$  der quadrierten Ladungen ermittelt. Die Summe  $s(i)$  entspricht dann die "reproduzierte Kommunalität" der Variablen  $i$ , die sich von der durch den Benutzer vorgegebenen unterscheidet. Jede Ladung der Zeile  $i$  wird dann durch die Wurzel aus  $s(i)$  dividiert. Das Ergebnis ist die zeilenweise normierte, orthogonale, nicht-rotierte Faktorladungsmatrix. Diese wird dann der "Prokrustes-Anpassung" unterzogen -woraus die konfirmatorische Faktorladung hervorgeht. Die Summe der quadrierten normierten Faktorladung in einer Matrixzeile ist nunmehr 1.0.

Die Normierung verletzt die zentrale Bedingung des Rotierens, dass die euklidischen Distanzen zwischen den Ladungspunkten nicht verändert werden dürfen und dass die aus der "prokrustes-rotierte" Faktorladungsmatrix reproduzierte Korrelationsmatrix identisch sein muss mit der, die aus der ursprünglichen nicht-rotierten orthogonalen Faktorladungsmatrix hervorgegangen ist. Durch die Normierung werden die Distanzmatrix und die reproduzierte Korrelationsmatrix aus unrotierter Faktorladungsmatrix und "prokrustes-rotierter", konformatorischer Faktorladungsmatrix ungleich.

#### *Eingabefeld 4: Zahl der Bootstrap-Stichproben.*

Empfohlen sind mindestens 1000. Die Rechenzeit beträgt für unser Beispiel nur 1 Sekunde. Sie hängt direkt von der gewählten Stichprobenzahl ab. Erhöht man die Stichprobenzahl von beispielsweise 1000 auf 2000, dann wird die Rechenzeit ungefähr verdoppelt und man erzeugt dadurch Bootstrap-Ergebnisse, die vielleicht an der 3. Kommastellen verändert sind. Man darf aber nicht unterstellen, dass sie "besser" sind, d.h. dass sie den "wahren" Werten in der Grundgesamtheit näher kommen. Entscheidend ist die Qualität der Originalstichprobe. Ist sie verzerrt, dann wird sie durch Bootstrapping nicht repräsentativ. Trotzdem gilt das Prinzip: Je mehr Bootstrapstichproben umso "feiner" messen wir die gesuchten Bootstrapkoeffizienten. In Abschnitt P30.6.1.6 haben wir mit Hilfe eines Beispiels dieses Prinzip begründet. Wichtig ist besonders, dass man die Stichprobenzahl an das Konfidenzniveau anpasst. Wir werden bei der Erläuterung zu Eingabefeld 5 (Konfidenzniveau) darauf zurückkommen. Siehe auch Abschnitt P20.25.9. im Almo-Dokument 13b "Bootstrap beim Allgemeinen Linearen Modell". Dort wird gezeigt welche Auswirkungen eintreten, wenn die Stichprobenzahl erhöht wird.

#### *Eingabefeld 5: Konfidenzniveau für Konfidenzintervall*

Der Benutzer bestimmt das Konfidenzniveau. Üblich ist ein Niveau von 95%. Werden die Faktorladungen je Variable aus z.B. 1000 Bootstrap-Stichproben der Größe nach (aufsteigend) sortiert, dann liegen die Grenzwerte des Konfidenzintervalls 2.5%, also 25 Werte unterhalb bzw. oberhalb des maximalen bzw. minimalen Werts. 950 Werte liegen zwischen den Intervallgrenzen. 25 Werte sind wenig, besser wäre es, 2000

Stichproben zu rechnen. Dann liegen 50 Werte ober- und unterhalb der Grenzwerte. Wird ein Konfidenzniveau von 99% gewählt, dann liegen bei 1000 Stichproben nur 5 Werte ausserhalb der Intervallgrenzen. Erst mit 5000 Stichproben werden 25 Werte und mit 10 000 Stichproben 50 Werte außerhalb der Intervallgrenzen erreicht. Der Benutzer sollte die Stichprobenzahl an das Konfidenzniveau (oder umgekehrt) anpassen. Ziel muss sein, möglichst viele Werte unter- bzw. oberhalb der Grenzwerte des Konfidenzintervalls zu erhalten. Je mehr aufsteigend sortierte Stichprobenwerte vorliegen umso feiner sind die Differenzen von einem Wert zum nächsten, umso genauer können die Grenzwerte bestimmt werden. Je höher auch der Benutzer das Konfidenzniveau ansetzt, umso mehr nähern sich der obere und untere Grenzwert dem maximalen bzw. minimalen Wert an, umso breiter wird das dazwischen liegende Konfidenzintervall.

#### *Eingabefeld 6: Konfidenzintervall*

Almo bietet drei Methoden für die Berechnung des Konfidenzintervalls an. Dies sind  
0 = das einfache Perzentil-Verfahren  
1 = --- ist nicht möglich ---  
2 = --- ist nicht möglich ---  
3 = das Perzentil BC-Verfahren (bias corrected)  
4 = das Perzentil BCa-Verfahren (bias corrected and accelerated)

Das Verfahren 1 (das asymmetrische Perzentil-t -Verfahren) und 2 (das symmetrische Perzentil-t -Verfahren) können bei der Faktorenanalyse nicht eingesetzt werden. Die Verfahren wurden oben in Abschnitt P30.6.1.2 erläutert

Vom gewählten Verfahren hängt auch ab, wie die Signifikanz, genauer der p-Wert des untersuchten Koeffizienten (z.B. des Regressionskoeffizienten) ermittelt wird. Bei den beiden Perzentil-t -Verfahren wird p anders berechnet als bei den übrigen Verfahren.

#### *Eingabefeld 7: Startzahl des Zufallsgenerators.*

Der Benutzer kann die Startzahl beliebig verändern. Wird eine zweite Analyse mit der gleichen Startzahl gerechnet, dann entsteht exakt dieselbe Folge von Zufallszahlen, wodurch aus der Datenmatrix dieselben Probanden für die Bootstrap-Stichprobe ausgewählt werden, wie für die erste Analyse. Damit sind auch die Ergebnisse identisch.

Die Sub-Box "Zielmatrix"



Wird die Sub-Box geöffnet dann sieht man folgendes

X Loesche wieder diese Sub-Box (Voreinstellungen wieder gueltig)

**Zielmatrix**  
 Faktorladungen aus Original- u. Bootstrap-Stichproben werden gemäß einer Zielmatrix in einen gemeinsamen Raum "gezwungen"

**Achsen der Zielmatrix**

**schiefwinklig** die Koordinaten-Achsen stehen aufeinander  
 rechteckig =Varimax-Rotation

**schiefwinklig:**

**0** 0 =Quartimin-Rotation   
 1 =Gruppen-Rotation  
 2 =benutzer-definierte Gruppen-Rotation  
   **".\TESTDAT\Gruppenzugehoer.mat"**

---

**Typ der Zielmatrix**

**1** 1 =eine -1,0,1 -Matrix, die aus der Varimax- bzw. der Quartimin-Faktorladungsmatrix der Original-Stichprobe abgeleitet wird  
 2 =als Zielmatrix wird die gesamte schief- bzw. rechteckige Faktorladungsmatrix der Original-Stichprobe eingesetzt  
 3 =die gesamte nicht-rotierte orthogonale Faktorladungsmatrix aus der Originalstichprobe wird als Zielmatrix eingesetzt  
 4 =benutzer-definierte Zielmatrix   
   **".\TESTDAT\Zielmat6Var.mat"**

---

**wenn schiefwinkligen Achsen für Zielmatrix**   
 dann

**1** 1 =Bootstrap für Faktorladungen mit **achsparalleler** Projektion der Variablenpunkte auf die schiefwinkligen Achsen  
 2 =Bootstrap für Faktorladungen mit **rechteckiger** Projektion der Variablenpunkte auf die schiefwinkligen Achsen

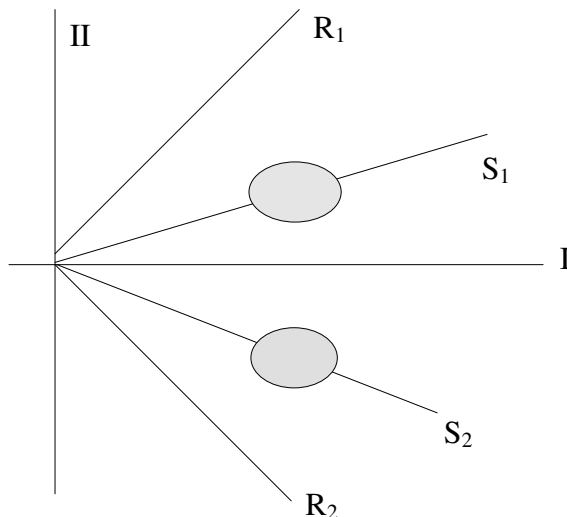
Eingabefeld 1 der Sub-Box: Achsen der Zielmatrix

Die nicht-rotierte orthogonale Faktorladungsmatrix aus der Originalstichprobe und alle nicht-rotierten orthogonalen Faktorladungsmatrizen aus den Bootstrapstichproben werden an eine vorgegebenen Zielmatrix angepasst. Die Zielmatrix sollte dabei aus der rechteckig oder schiefwinklig rotierten Faktorladungsmatrix der Originalstichprobe abgeleitet werden. Sie erzeugt ein für alle Stichproben gemeinsames Koordinatensystem, das, nach Wahl des Benutzers, rechteckig oder "schiefwinklig" sein kann. Der Kalkül, der die neue angepasste Faktorladungsmatrix generiert ist der der "konfirmatorischen Faktorenanalyse".

**rechteckig** =die Faktorladungsmatrix aus der Originalstichprobe wird orthogonal nach

dem Varimax-Verfahren rotiert. Daraus wird eine Zielmatrix abgeleitet, an die die Ladungsmatrizen der Bootstrapsstichproben heran gezwungen werden. Dadurch entsteht ein für alle Stichproben gemeinsames orthogonales Koordinatensystem.

**schiefwinklig** = 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 Variablenruppen, die wir in der Grafik als 2 „Punktwolken“ durch 2 Ellipsen dargestellt haben.

Die Achsen I und II bilden das rechtwinklige Ausgangs-Koordinatensystem. Die Achsen S<sub>1</sub> und S<sub>2</sub> sind die schiefwinkligen Achsen, die uns z.B. das Quartimin-Verfahren liefert. Sie laufen bei der Quartimin-Rotation ungefähr "mittig durch die beiden Punktwolken.

Die Achsen R<sub>1</sub> und R<sub>2</sub> sind die rechtwinkligen Achsen, die uns die rechtwinklige Varimax-Rotation liefert. Es ist deutlich zu erkennen, dass je dichter die beiden Punktwolken beieinander liegen, die Varimax-Rotation umso schlechter ist.

Beim Bootstrapping wird nun standardmäßig aus der quartimin-rotierten Faktorladungs-matrix der Originalstichprobe eine Zielmatrix abgeleitet. Zusätzlich wird die Matrix der Korrelationen zwischen den schiefwinkligen Achsen aus der Quartimin-Rotation der Faktorladungs-matrix der Originalstichprobe gebraucht. Also errechnet sie (ohne Zutun des Benutzers) aus den Winkeln zwischen den schiefwinkligen Achsen. Aus dieser Korrelationsmatrix und der Zielmatrix entsteht dann die "schiefwinklige Zielmatrix", an die die Ladungsmatrizen der Bootstrapsstichproben heran gezwungen werden. Dadurch entsteht ein für alle Stichproben gemeinsames schiefwinkliges Koordinatensystem. Wie noch gezeigt werden wird, bietet Almo dem Benutzer noch zwei weitere schiefwinklige Rotationsverfahren an, die Gruppenrotation und die benutzer-definierte Gruppenrotaton.

#### *Eingabefeld 2 der Sub-Box: Typ der schiefwinkligen Rotation*

Der Benutzer kann durch Klick auf den Zählknopf zwischen diesen 3 Verfahren wählen

0 = Quartimin-Rotation.

Das Verfahren entspricht ungefähr (nicht exakt) "Oblimin 0"

1 = Gruppen-Rotation.

Bei diesem Verfahren wird zuerst eine Quartimin-Rotation gerechnet. Dann wird jede Variable der Gruppe zugeordnet, auf deren Faktor ihre absolut größte Ladung liegt. Mit den so gewonnenen Gruppen wird dann der spezielle Kalkül der Gruppen-Rotation gerechnet. Als Ergebnis entstehen dann schiefwinklige Achsen, die "varianzmaximierend durch ihre Punktwolke verlaufen. Das Verfahren wird ausführlich im Almo-Dokument 15 "Faktorenanalys", Abschnitt P30.2.3.2.1 dargestellt.

2 = Benutzer-definierte Gruppen-Rotation.

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 (=Punktwolken) gezogen.

*Eingabefeld 3:* Dem Program muss mitgeteilt werden, in welcher Datei der Benutzer die Gruppenzugehörigkeit der zu faktorisierenden Variablen geschrieben hat. Wurde nicht das Rotationsverfahren 2 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 Prog30ml\_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. 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 in der Matrix bedeutet "Gruppenzugehörigkeit". Im Beispiel gehören die ersten 4 Variablen zur Gruppe 1 und die 5. und 6. Variablen zur Gruppe 2.

*Eingabefeld 4 der Sub-Box: Typ der Zielmatrix*

Die Zielmatrix wird aus der Faktorenanalyse der Originalstichprobe gewonnen. Drei verschiedenen Zielmatrizen stehen zur Auswahl

Zielmatrix bei **orthogonaler** Rotation

1 = eine -1,0,1 -Matrix,

die aus der Varimax-Faktorladungsmatrix der Originalstichprobe abgeleitet wird

- 2 = die gesamte Varimax-Faktorladungsmatrix aus der Originalstichprobe
- 3 = die gesamte nicht-rotierte orthogonale Faktorladungsmatrix aus der Originalstichprobe wird als Zielmatrix eingesetzt
- 4 = die selbst definierte Zielmatrix. Siehe Beispiel in Abschnitt P30.6.2.4, Punkt 4

Zielmatrix bei **schiefwinkliger** Rotation

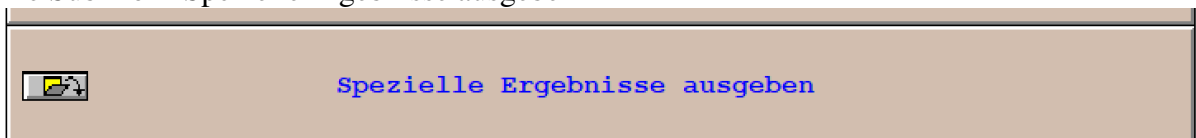
- 1 = eine -1,0,1-Matrix,  
die aus der Quartimin-rotierten, achsparallel projizierte Faktorladungsmatrix der Originalstichprobe abgeleitet wird. Zusätzlich wird den Bootstrapstichproben noch vorgegeben: Die Matrix der Winkel (bzw. Korrelationen) zwischen den schiefwinkligen Achsen aus der Quartimin-Rotation der Faktorladungsmatrix der Originalstichprobe
- 2 = die gesamte Quartimin-rotierte achsparallel projizierte Faktorladungsmatrix aus der Originalstichprobe.  
Zusätzlich wird den Bootstrapstichproben noch vorgegeben:  
Die Matrix der Winkel (bzw. Korrelationen) zwischen den schiefwinkligen Achsen aus der Quartimin-Rotation der Faktorladungsmatrix der Originalstichprobe
- 3 = nicht möglich bei schiefwinkliger Rotation. Also schaltet um auf 2
- 4 = die selbst definierte Zielmatrix.  
Muss ergänzt werden durch eine Matrix der Korrelationen zwischen den schiefwinkligen Achsen. Geschieht dies nicht, dann übernimmt Also die Faktor-Korrelations-Matrix aus der Quartimin-Rotation. Siehe Beispiel in Abschnitt P30.6.2.5, Punkt 3.

*Eingabefeld 5 der Sub-Box: Vom Benutzer definierte Zielmatrix aus Datei lesen*  
Der Benutzer teilt hier mit, wo er die von ihm geschriebene Datei gespeichert hat.

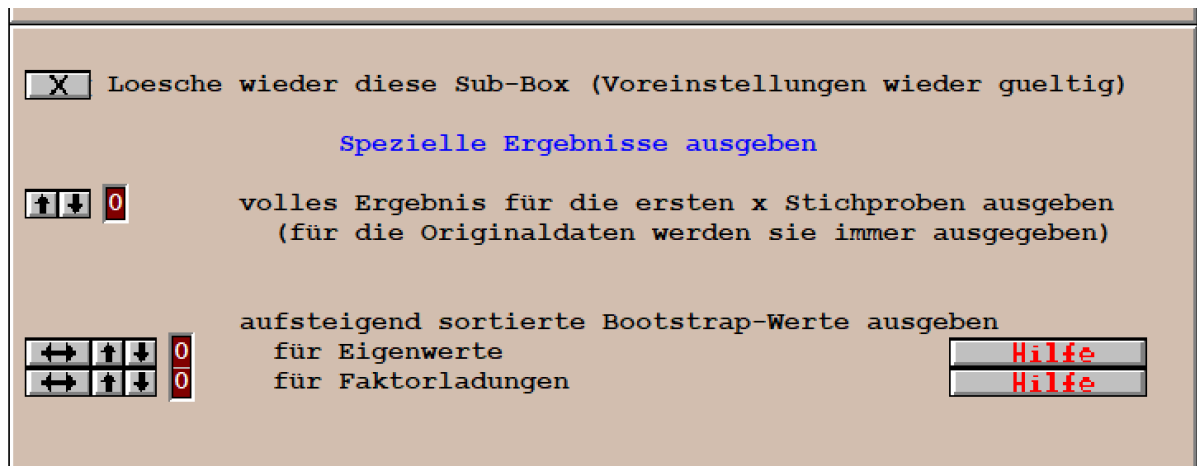
*Eingabefeld 6 der Sub-Box: Wenn schiefwinklige Achsen für Zielmatrix.  
Achsparallele oder rechtwinklige Projektion*

Wird im Eingabefeld 2 "1" eingesetzt (d.h. die Faktorladungen gewählt) und sollen die Faktorladungen schiefwinklig rotiert werden dann muss entschieden werden, ob  
1 =die Ladungsmatrix (mit achsparallel auf die Achsen projizierten Faktorladungen)  
oder  
2 =die Strukturmatrix (mit rechtwinklig projizierten Faktorladungen)  
dem Bootstrap-Verfahren unterworfen werden soll. Im Eingabefeld muss entsprechend 1 oder 2 eingesetzt werden. Werden die Faktorwert-Koeffizienten (als Bootstrap-Koeffizienten) gewählt, dann ist der Eintrag im 4. Eingabefeld irrelevant.

Die Sub-Box "Spezielle Ergebnisse ausgeben



Wird die Sub-Box geöffnet dann sieht man folgendes



*Eingabefeld 1 der Sub-Box: Volle Ergebnisse ausgeben.*

0 = die Ergebnisse für die Originalstichprobe werden ausgegeben. In einem deutlich getrennten 2. Ausgabeteil werden danach die kumulierten Bootstrap-Ergebnisse für die Gesamtzahl der Stichproben ausgegeben.

x = wird beispielsweise 3 eingesetzt, dann werden die Ergebnisse für die Originalstichprobe und zusätzlich für die Bootstrap-Stichproben 1, dann 2, dann 3 und erst danach die zusammengefassten Bootstrap-Ergebnisse für die Gesamtzahl der Stichproben ausgegeben

*Eingabefeld 2: Bootstrap-Werte für x-ten Eigenwert ausgeben*

0 = keine Ausgabe

x = Bootstrap-Werte für x-ten Eigenwert oder Faktorladung der x-ten Variablen

Ein Beispiel:

Der Benutzer möchte die aufsteigend sortierten Werte aus den 1000 Bootstrapstichproben für den 2. Eigenwert sehen. In das Eingabefeld ist dann 2 einzusetzen. Es darf nur eine Zahl eingetragen werden. Also liefert dann 1000 untereinander stehende, aufsteigend sortierte Werte. Sie werden allerdings erst sichtbar, wenn der Benutzer in der Ergebnisliste einen bestimmten Knopf anklickt. Siehe das Beispiel in P30.6.2.1.

*Eingabefeld 2 und 3: Bootstrap-Werte für Faktorladung einer Variablen ausgeben*

0 = keine Ausgabe

x = Bootstrap-Werte für Faktorladung der x-ten Variablen

Ein Beispiel:

Der Benutzer möchte die aufsteigend sortierten Faktorladungen aus den 1000 Bootstrapstichproben für die 4. Variable für die 3 extrahierten Faktoren sehen. In das Eingabefeld ist dann 4 einzusetzen. Es darf nur eine Zahl eingetragen werden. Einzutragen ist also die Stelle der Variablen in der Hintereinanderfolge der Variablen - nicht die Variablennummer.

Also zeigt dann (untereinander stehend) die 1000 Faktorladungen der 4. Variablen aus den 1000 Bootstrapstichproben - zuerst für Faktor 1, dann 2, dann 3. Die umfangreiche Ausgabe wird von Almo "verborgen". In der Ergebnisliste wird nur ein Knopf mit kurzem erläuternden

Text gezeigt. Nach Klick auf diesen Knopf werden die Daten dann geladen und in die Ergebnisliste eingefügt. Der Benutzer kann mit einfachem Mausklick den eingefügten Text wieder "verbergen"

### P30.6.4 Bootstrap-Ergebnisse aus Programm Prog30ml.Msk

Die Ergebnisse sind in 2 Abschnitte unterteilt

1. Ergebnisse für die Originalstichprobe. Sie stimmen weitgehend überein mit den Ergebnissen, die durch das Standard-Faktorenanalyse-Programm Prog30m2.Msk ausgegeben werden
2. Die Bootstrap-Ergebnisse. Diese bestehen aus 2 Tabellen
  - a. Bootstrap der Eigenwerte
  - b. Bootstrap der Ladungen für die x Faktoren

Da das Bootstrap-Verfahren mit recht- oder schiefwinkliger Zielmatrix gerechnet werden kann, werden wir die unter 1. und 2. genannten Ergebnisse zuerst für die rechtwinklige Zielmatrix und anschließend für die schiefwinkliger Zielmatrix zeigen.

#### P30.6.4.1 Ergebnisse aus Originalstichprobe bei rechtwinkliger Zielmatrix

Almo gibt zuerst die Ergebnisse für die Originalstichprobe aus. Sie sind identisch mit denen, die im Standardprogramm Prog30m2 zur Faktorenanalyse ausgegeben werden. Die wesentlichen Ergebnisse sind (1) die unrotierte Faktorladungsmatrix, die an die Zielmatrix angepasst wird, wodurch dann (2) die *orthogonale konfirmatorische Faktorladungsmatrix* für die Originalstichprobe entsteht - die dann noch graphisch dargestellt wird. Das wurde bereits oben in P30.6.2.1 gezeigt. Die Ergebnisse werden hier nochmals wiederholt.

Matrix der nicht-rotierten orthogonalen Faktorladungen der Originalstichprobe

		Faktor 1	Faktor 2
t6_parag	V14	0.8752	0.2121
t7_sente	V15	0.8513	0.3246
t9_word_	V17	0.8591	0.2444
t14_word	V22	0.4328	-0.6527
t15_num	V23	0.2310	-0.7418
t17_obje	V25	0.3487	-0.6257

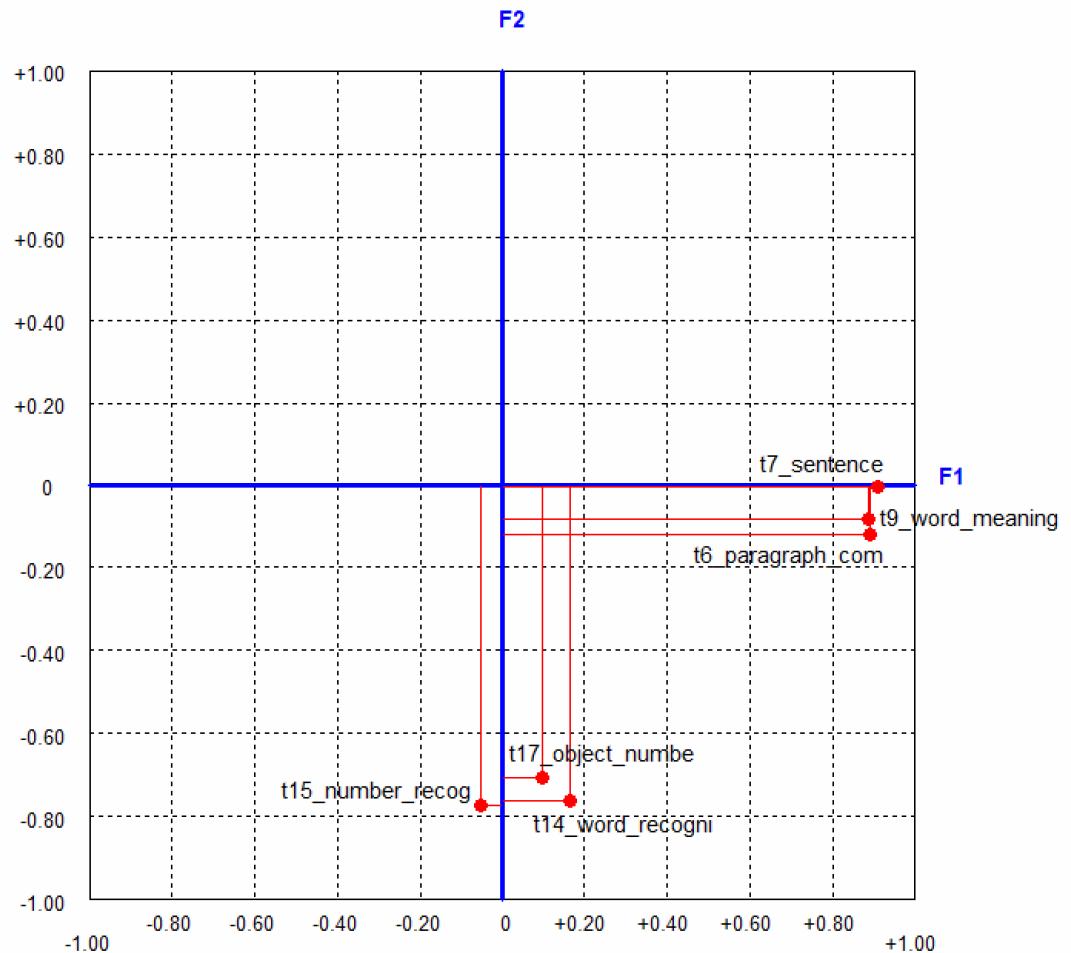
Matrix der orthogonalen konfirmatorischen Faktorladungen der Originalstichprobe

		Faktor 1	Faktor 2
t6_parag	V14	0.8925	-0.1199
t7_sente	V15	0.9111	-0.0065
t9_word_	V17	0.8892	-0.0839
t14_word	V22	0.1664	-0.7652
t15_num	V23	-0.0540	-0.7750
t17_obje	V25	0.0979	-0.7096

Graphisch dargestellt:

### Graphik 3

orthogonale konfirmatorische Faktorladungen



### P30.6.4.2 Ergebnisse aus Bootstrap bei rechtwinkliger Zielmatrix

Die Ergebnis-Ausgabe beginnt mit einem Protokoll der Einstellungen zum Bootstrap-Verfahren und zur Faktorenanalyse

#### Bootstrap-Einstellungen

```
-----  
Startzahl fuer Zufallsgenerator: 578125  
Konfidenzintervall u. Signifikanz p werden nach einfachen Perzentil-Verfahren  
berechnet  
Konfidenzniveau: 95%  
kleinst moeglicher berechenbarer p-Wert=0.000998
```

#### Einstellungen zur Faktorenanalyse

```
-----  
Faktorenanalys: "normale" Faktorenanalyse  
Kommunalitaet: Diagonale =1.0 mit 0 Iterationen  
Faktorenzahl: 2  
Eigenwert-Kalkuel: Tridiagonal-Qr-Verfahren  
Rotation: rechtwinkliges Varimax-Verfahren  
Zielmatrix: eine -1,0,1 -Matrix, gebildet aus der Varimax-Rotation
```

### P30.6.4.2.1 Bootstrap der Eigenwerte

Beachte: In Spalte \*g wird einseitig getestet, ob der Eigenwert  $\lambda \geq 1.0$  ist

und somit das Kaiser-Kriterium erfüllt  
 p ist die Wahrscheinlichkeit, dass die Annahme, der Eigenwert x sei  $\geq 1.0$  falsch ist  
 $(1-p) \cdot 100$  ist der Prozentsatz der Bootstrapstichproben (von 1000 Stichproben), deren Eigenwert  $x \geq 1.0$  ist  
 nur sinnvoll bei Kommunalitaetschätzung: 1 (Diagonale=1.0 und keine Iterationen)

	Original-Stichprobe	Ergebnisse aus 1000 Bootstrapstichproben der Eigenwerte					
		*a Eigenwert	*d	*e	*f	*g	*h
			Mitt.wert Eigenwert	Verzerr. Eigenwert	Standard fehler	Signifik. p	Konfidenzintervall Konf.niv=0.950 unten oben
Eigenwert 1	2.5909	2.6035	0.0126	0.0928	0.0010	2.4396	2.7970
Eigenwert 2	1.5778	1.5729	-0.0049	0.0899	0.0010	1.3852	1.7331
Eigenwert 3	0.6967	0.7052	0.0085	0.0489	0.9990	0.6159	0.7971
Eigenwert 4	0.5819	0.5735	-0.0084	0.0412	0.9990	0.4908	0.6536
Eigenwert 5	0.2950	0.2991	0.0041	0.0273	0.9990	0.2490	0.3569
Eigenwert 6	0.2577	0.2459	-0.0119	0.0219	0.9990	0.2044	0.2901

- \*a Eigenwerte aus Originalstichprobe
- \*d Mittelwert der Eigenwerte aus allen Bootstrap-Stichproben
- \*e mit "Verzerrung" wird die Differenz zwischen dem Mittelwert aus allen Bootstrap-Stichproben minus dem Wert aus der Originalstichprobe bezeichnet
- \*f Der Standardfehler ist gleich der Standardabweichung der Eigenwerte aus allen Bootstrap-Stichproben
- \*g Getestet wird, ob der Eigenwert groessergleich 1.0 ist  
 p ist dann die Wahrscheinlichkeit, dass diese Annahme falsch ist  
 1-p ist der Anteil der Eigenwerte groessergleich 1.0 an der Zahl von 1000 Bootstrapstichproben
- \*h Konfidenzintervall - nach dem vom Benutzer vorgegebenen Konfidenz-Niveau von 95.00%  
 Beim "einfachen" Perzentil-Verfahren bedeutet das:  
 Von den aufsteigend sortierten 1000 Werten aus den Bootstrap-Stichproben befinden sich 95.00% der Werte zwischen den Konfidenzgrenzen und je 2.50% oberhalb und unterhalb der Konfidenzgrenzen

Almo führt zuerst ein Bootstrap-Verfahren für die Eigenwert durch. Das Ergebnis ist eindeutig: 2 Faktoren sind mit  $p=0.001$  signifikant.

### P30.6.4.2 Bootsrap der Faktorladungen

Wir werden nur die Ergebnisse für den 1. Faktor ausführlich erläutern; die für den 2. Faktor werden kommentarlos abgebildet. Die von Almo gelieferte Tabelle ist sehr breit. Um sie hier abbilden zu können, müssen wir die ersten 3 Tabellenspalten als eigenständige Tabelle abtrennen. In ihr werden die Ergebnisse aus der Originalstichprobe in kompakter Form wiederholt.

#### Ergebnisse fuer Faktor 1 der orthogonal konfirmatorischen Faktorladungen

Ergebnisse aus Originalstichprobe			
	*a	*b	*c
	unrotierte orthogonale Faktorladung	Varimax- Faktorladung	orthogonal konfirmator. Faktorladung
t6_paragraph_co V14	0.875175	0.891791	<u>0.892498</u>
t7_sentence V15	0.851345	0.911040	<u>0.911092</u>
t9_word_meaning V17	0.859052	0.888700	<u>0.889200</u>
t14_word_recogn V22	0.432765	0.162007	<u>0.166428</u>
t15_number_reco V23	0.230952	-0.058437	-0.053963
t17_object_num V25	0.348691	0.093774	0.097873

Ergebnisse aus Bootstrap der orthogonal konfirmatorischen Faktorladungen

Zielmatrix: -1,0,1 -Matrix aus Varimax *i)							
Mitt.wert *d	orthogonal	*e	*f	*g	*h		
konfirmator.	Verzerrung	Standard	Signifik.	Konfidenzintervall			
Faktorladung	Faktorladung	fehler	p	unten	oben	Breite	
t6	0.891229	-0.001269	0.012202	0.000999	0.865715	0.912623	0.046908
t7	0.910522	-0.000570	0.009245	0.000999	0.890842	0.927471	0.036630
t9	0.888337	-0.000863	0.011234	0.000999	0.864000	0.908476	0.044476
t14	0.168739	0.002311	0.048154	0.002000	0.074789	0.263801	0.189012
t15	-0.056682	-0.002719	0.059857	0.342000	-0.175103	0.057562	0.232666
t17	0.097781	-0.000092	0.060899	0.122000	-0.026843	0.218787	0.245630

### Erläuterung zu den Spalten \*a bis \*i in obiger Tabelle

- Spalte \*a Originalstichprobe  
unrotierte orthogonale Faktorladungen
- Spalte \*b Originalstichprobe  
schiefwinklig quartimin-rotierte Faktorladungen  
Ladungspunkte achsparallel auf Achsen projiziert
- Spalte \*c Originalstichprobe  
schiefwinklig konfirmatorische Faktorladungen  
Ladungspunkte achsparallel auf Achsen projiziert  
"konfirmatorisch" bedeutet: angepasst an schiefwinkl.  
Zielmatrix: -1,0,1-Matrix aus Quartimin-Rotat. \*i)
- Spalte \*d Mittelwert aus allen Bootstrapstichproben  
schiefwinklig konfirmatorische Faktorladungen  
Ladungspunkte achsparallel auf Achsen projiziert  
"konfirmatorisch" bedeutet: angepasst an schiefwinkl.  
Zielmatrix: -1,0,1-Matrix aus Quartimin-Rotat. \*i)  
BEACHTTE: Die Bootstrap-Ergebnisse Standardfehler, p-Wert und  
Konfidenzgrenzen beziehen sich auf \*c
- Spalte \*e mit "Verzerrung" wird die Differenz zwischen dem Mittelwert aus allen  
Bootstrap-Stichproben minus dem Wert aus der Originalstichprobe bezeichnet
- Spalte \*f Der Standardfehler ist gleich der Standardabweichung der  
konfirmatorischen Faktorladungen aus allen Bootstrap-Stichproben
- Spalte \*g Berechnet wird die zweiseitige Signifikanz p. \*\*Hilfe540\*\*
- Spalte \*h Konfidenzintervall - nach dem vom Benutzer vorgegebenen Konfidenz-Niveau  
von 95.00%. Beim "einfachen" Perzentil-Verfahren bedeutet das:  
Von den aufsteigend sortierten 1000 Werten aus den Bootstrap-Stichproben  
befinden sich 95.00% der Werte zwischen den Konfidenzgrenzen und je 2.50%  
oberhalb und unterhalb der Konfidenzgrenzen  
Befindet sich der Wert 0 nicht innerhalb der Konfidenzintervalls, dann kann  
interpretiert werden: Mit der Sicherheit des gewählten Konfidenzniveaus  
ist der Bootstrap-Koeffizient ungleich dem Wert 0. Oder umgekehrt die  
Irrtumswahrscheinlichkeit, dass der Wert 0 doch im Intervall liegt, kann  
nicht grösser (=schlechter) sein als  $1.0-95.00/100=0.05$
- Spalte \*i die Zielmatrix definiert das recht- oder schiefwinklige Koordinatensystem,  
an das die unrotierten orthogonalen Faktorladungsmatrizen der Original-  
und Bootstrapstichprobe "heranrotiert" werden müssen. Die Faktorladungs-  
matrizen aller Stichproben befinden sich dadurch im gleichen Koordinaten-  
Raum.  
Die Zielmatrix wird aus der Faktorenanalyse der Originalstichprobe gewonnen  
Möglich sind:  
die Zielmatrix 1 entsteht als -1,0,1 -Matrix aus der Varimax- bzw.  
Quartimin-Rotation  
die Zielmatrix 2 ist gleich der Varimax- bzw. Quartimin-Rotation  
die Zielmatrix 3 ist gleich der unrotierten Faktorladungsmatrix  
die Zielmatrix 4 ist eine vom Benutzer definierte Matrix

Das wesentliche Ergebnis des Bootstrap-Verfahrens ist nun folgendes:

Durch Bootstrap erhalten wir für die "orthogonal konfirmatorischen Faktorladungen" der  
*Originalstichprobe* (Spalte \*c in obiger Tabelle) einen verteilungsfreien Schätzer

1. des Standardfehlers (Spalte \*f)
2. des p-Wertes (Spalte \*g)
3. und des unteren und oberen Grenzwertes des Konfidenzintervalls (Spalte \*h)

### Der p-Wert

kann in folgender Weise interpretiert werden: Die Variable  $i$  besitzt eine Faktorladung auf dem Faktor  $j$  - wobei in unserem 2-faktoriellen Beispiel  $j=1$  und  $j=2$  sein kann. Der p-Wert sagt dann aus, ob die Faktorladung der Variablen  $i$  auf dem Faktor  $j$  den Wert  $=0$  besitzen könnte, oder ob sie auf dem Faktor  $j$  signifikant lädt. Üblicherweise wird ein p-Wert von  $\leq 0.05$  als signifikant und  $> 0.05$  als nicht signifikant gedeutet. Wenn mehrere Faktoren vorhanden sind, kann der Fall auftreten, dass die Variable nur auf einem Faktor signifikant lädt und auf den anderen nicht. Das ist häufig ein erwünschter Zustand: Jede Variable gehört signifikant nur zu einem Faktor.

Wichtig ist zu erkennen, dass damit die Schätzer für die "orthogonalen konfirmatorischen Faktorladungen" (Spalte \*c) gefunden sind - und nicht die Schätzer für die unrotierten orthogonalen Faktorladungen (Spalte \*a). In Abschnitt P30.6.2.2 haben wir bereits ausgeführt,

dass die drei Schätzwerte abhängig sind von der Lage der "Punktekonfiguration" der Faktorladungsmatrix im Koordinatensystem. Werden die Koordinatenachsen gedreht (rotiert), wobei die Punktekonfiguration stehen bleibt, dann entstehen andere Koordinatenwerte für die Punkte (die Faktorladungen) und damit andere Werte für die drei Schätzer. Es muss also entschieden werden, in welche Position das Koordinatensystem letztendlich rotiert werden soll. Dies sollte die Position sein, die der Forscher als diejenige interpretiert, die seinem Untersuchungsgegenstand am Besten entspricht. Das ist die Position der "konfirmatorischen" Faktorladungsmatrix. Beim Bootstrap entsteht sie durch das "Heranzwingen" an eine Zielmatrix.

Wird die Zielmatrix 2 verwendet (siehe P30.6.2.3) dann ist die "konfirmatorische" Faktorladungsmatrix der *Originalstichprobe* identisch mit der Varimax-Faktorladungsmatrix (bzw. im schiefwinkligen Fall der Quartimin-Faktorladungsmatrix). Die oben genannten 3 Schätzer, die das Bootstrapping erbringt, beziehen sich damit auf eine gewohnte Art von Faktorladungen - wo hingegen die "konfirmatorische" Faktorladung (nach dem "Heranzwingen" an eine -1,0,1 Zielmatrix) eher ungewohnt ist.

### Ergebnisse fuer Faktor 2 der orthogonal konfirmatorischen Faktorladungen

Ergebnisse aus Originalstichprobe				
		*a	*b	*c
		unrotierte	Varimax-	orthogonal
		orthogonale	Faktorladung	konfirmator.
		Faktorladung	Faktorladung	Faktorladung
t6_paragraph_co	V14	0.212128	-0.125057	-0.119906
t7_sentence	V15	0.324565	-0.011746	-0.006486
t9_word_meaning	V17	0.244441	-0.089078	-0.083945
t14_word_recogn	V22	0.652683	-0.766182	<u>-0.765234</u>
t15_number_reco	V23	0.741787	-0.774708	<u>-0.775033</u>
t17_object_num	V25	0.625707	-0.710142	<u>-0.709589</u>

Ergebnisse aus Bootstrap der orthogonal konfirmatorischen Faktorladungen				
Zielmatrix: -1,0,1 -Matrix aus Varimax *i)				
Mitt.wert *d	*e	*f	*g	*h
orthogonal				

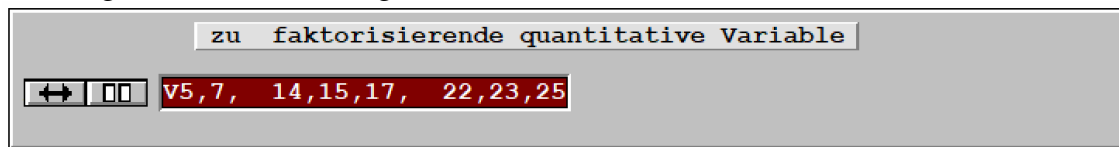
	konfirmator. Faktorladung	Verzerrung Faktorladung	Standard fehler	Signifik. p	Konfidenzintervall		
					unten	oben	Breite
t6	-0.120411	-0.000505	0.033689	0.000999	-0.187702	-0.056056	0.131647
t7	-0.005929	0.000557	0.040573	0.874000	-0.087395	0.072569	0.159964
t9	-0.083498	0.000447	0.038608	0.020000	-0.157489	-0.011414	0.146075
t14	-0.763238	0.001996	0.028808	0.000999	-0.815156	-0.703643	0.111513
t15	-0.774431	0.000602	<u>0.024261</u>	<u>0.000999</u>	-0.816633	-0.723414	0.093219
t17	-0.708194	0.001395	<u>0.041332</u>	<u>0.000999</u>	-0.777378	-0.620544	0.156835

### P30.6.4.2.3 Bootstrap der reproduzierten Kommunalitäten

Wir fügen zu den 6 Variablen unseres Beispiels noch 2 Variable hinzu.

V5    month\_since\_birthday    (Zahl der Monate bis zum Geburtstag)  
V7    age\_years    (Alter in Jahren)

Die Eingabe-Box ist dann folgende



Die erste hinzugefügte Variable hat keinen Bezug zu den 6 "Begabungs"-Variablen unseres Beispiels, die zweite nur einen geringen. Die Korrelationen in der Originalstichprobe sind

"Begabungsvariable" aus Holzinger/Swineford-Test

	t6	t7	t9	t14	t15	t17
month since birthday	-0.0170	-0.0372	0.0071	-0.0087	0.0590	0.0642
age in year	-0.2088	-0.2397	-0.1770	-0.0289	0.0831	0.0815

month\_since\_birthday korreliert mit Werten nahe 0 mit den 6 "Begabungs"-Variablen. age\_years korreliert schwach negativ, aber wohl signifikant, mit den ersten 3 "Begabungs"-Variablen, die (in der Terminologie Thurstones) auf dem Faktor "verbal reasoning" laden.

Diese 8 Variablen werden faktorisiert. Aus der *orthogonalen unrotierten* Faktorladungsmatrix werden die "reproduzierten" Kommunalitäten gewonnen. Beim Bootstrapping werden die Ladungsmatrizen aus der Originalstichprobe und aus allen Bootstrapstichproben an eine Zielmatrix angepasst. Es entstehen die *konfirmatorischen* Faktorladungsmatrizen. Das wurde in Abschnitt P30.6.2.3 ausführlich beschrieben.

Es gilt nun:

Die reproduzierten Kommunalitäten, ermittelt aus der *orthogonalen unrotierten* Faktorladungsmatrix einer Stichprobe und die reproduzierten Kommunalitäten, ermittelt aus der *konfirmatorischen* Faktorladungsmatrix derselben Stichprobe sind gleich.

Betrachten wir als Beispiel die Originalstichprobe. Aus der nicht-rotierten Ausgangsfaktorladungsmatrix ergibt sich durch variablenweise Summierung der quadrierten Faktorladungen folgende Matrix der reproduzierten Kommunalitäten

reproduzierte Kommunalitäten je Variable  
=Varianzbeiträge der Variablen

month_si	V5	0.0249
age_year	V7	0.2161
t6_parag	V14	0.7941
t7_sente	V15	0.8135
t9_word_	V17	0.7667
t14_word	V22	0.5751
t15_num	V23	0.5848
t17_obje	V25	0.5224
Summe		4.2976

Exakt dieselben Werte erhält man auch aus der konfirmatorischen Faktorladungsmatrix.

Die Werte bleiben auch gleich, wenn eine schiefwinklige Zielmatrix vorgegeben wird. Bei schiefwinkliger Zielmatrix werden die reproduzierten Kommunalitäten in folgender Weise berechnet: Die Matrix der reproduzierten Korrelationen ergibt sich aus

$$\text{Repro} = \mathbf{B} \cdot \mathbf{R} \cdot \mathbf{B}'$$

**B** =schiefwinklige Ladungsmatrix (mit achsparalleler Projektion)

**B'** =deren Transponierte

**R** =Korrelationsmatrix der schiefwinkligen Achsen

Die reproduzierten Kommunalitäten sind in der Diagonale von Repro enthalten. Der Benutzer kann sich die reproduzierte Korrelationsmatrix ausgeben lassen, wenn er die Zwischenergebnisse in der Optionbox "weitere Optionen" anfordert.

Beim *Bootstrap der Faktorladungen* der 8 Variablen erhalten wir folgendes Ergebnis für die p-Werte (das wir aus der großen Ergebnis-Tabelle heraus geschnitten haben

		Faktor 1	Faktor 2
		Signifikanz	Signifikanz
		P	P
V5	month_since_bir	0.4840	0.2940
V7	age_years	0.0010	0.0480
V14	t6_paragraph_co	0.0010	0.0010
V15	t7_sentence_com	0.0010	0.2420
V17	t9_word_meaning	0.0010	0.0040
V22	t14_word_recogn	0.0020	0.0010
V23	t15_number_reco	0.4420	0.0010
V25	t17_object_num	0.1480	0.0010

Faktor 1 ist der bereits erwähnte Faktor des "verbal reasoning" und Faktor 2 des "space reasoning".

month\_since\_birthday ist mit  $p=0.4840$  und  $p=0.2940$  auf beiden Faktoren nicht signifikant. age\_years hat hingegen einen hoch signifikanten p-Wert von  $0.001$  auf Faktor 1 und einen gerade noch signifikanten p-Wert von  $p=0.048$  auf Faktor 2.

Aus dem *Bootstrap der reproduzierten Kommunalitäten* bzw. Varianzbeiträge der Variablen erhalten wir folgende Ergebnisse:

**Bootstrap der reproduzierten Kommunalitäten**  
(=der Varianzbeiträge der Variablen)

	Ergebnisse aus 1000 Bootstraptstichproben der reproduzierten Kommunalitaeten						
	*a	*d	*e	*f	*g	*h	
	reproduz. Kommunal.	Mitt.wert reproduz. Kommunal.	Verzerr. reproduz. Kommunal.	Standard fehler	Pseudo-t M/S	Konfidenzintervall Konf.niv=0.950 unten oben	
month_since_bir	0.0249	0.0494	0.0245	0.0449	1.1008	0.0013	0.1682
age_years	0.2161	0.2248	0.0087	0.0665	3.3803	0.0993	0.3486
t6_paragraph_co	0.7941	0.7925	-0.0016	0.0220	35.9875	0.7486	0.8321
t7_sentence_com	0.8135	0.8142	0.0007	0.0186	43.7374	0.7773	0.8463
t9_word_meaning	0.7667	0.7663	-0.0005	0.0248	30.9222	0.7152	0.8111
t14_word_recogn	0.5751	0.5735	-0.0016	0.0496	11.5685	0.4732	0.6636
t15_number_reco	0.5848	0.5837	-0.0012	0.0411	14.1940	0.5004	0.6571
t17_object_num	0.5224	0.5217	-0.0008	0.0531	9.8203	0.4112	0.6146

- \*a reproduzierte Kommunalitaeten der Variablen in Originalstichprobe = Varianzbeitraege der Variablen zu der durch die extrahierten Faktoren zusammen erklarte Varianz
- \*d Mittelwert der reproduzierte Kommunalitaeten aus allen Bootstrap-Stichproben
- \*e mit "Verzerrung" wird die Differenz zwischen dem Mittelwert aus allen Bootstrap-Stichproben minus dem Wert aus der Originalstichprobe bezeichnet
- \*f Der Standardfehler ist gleich der Standardabweichung der Werte der reproduzierte Kommunalitaeten aus allen Bootstrap-Stichproben
- \*g Es wird ein dem t-Wert analoger Signifikanz-Koeffizient gerechnet nach der Formel M/S. Dabei ist M= Mittelwert aus Spalte \*d und S= Standardfehler aus Spalte \*f. Ist M/S groesser ca. 2.0 dann darf (analog zum t-Test) angenommen werden, dass der Varianzbeitrag der Variablen *signifikant* groesser 0 ist
- \*h siehe oben bei Tabelle der Faktorladungen

Der Pseudo-t-Wert ist für die 6 "Begabungs"-Variablen hoch, woraus wir schließen können, dass sie zu der durch die 2 Faktoren erklärten Varianz signifikant beitragen.

### Pseudo-t

Die reproduzierte Kommunalität, bzw. der Beitrag der Variablen zur erklärten Varianz kann nur 0 oder größer sein. Negative Werte sind nicht möglich. Daraus folgt, dass sich aus dem Perzentil-Verfahren kein p-Wert ermitteln lässt. Als Ersatz-Koeffizient verwendet Almo einen von Zientek/Thompson vorgeschlagenen Pseudo-t -Wert.

Es wird ein dem t-Wert analoger Koeffizient gerechnet nach der Formel M/S. Dabei ist M= Mittelwert aus Spalte \*d in obiger Tabelle und S= Standardfehler aus Spalte \*f. Ist M/S größer ca. 2.0 dann darf (analog zum t-Test) angenommen werden, dass der Varianzbeitrag der Variablen *signifikant* größer 0 ist. Eine Irrtumswahrscheinlichkeit von  $p=0.05$  daraus abzuleiten, wie beim parametrischen t-Test, ist unzulässig.

In obiger Tabelle erhält erwartungsgemäß die Variable **month\_since\_birthday** einen sehr kleinen M/S-Wert. Sie besitzt keinerlei Gemeinsamkeit mit den 6 Begabungs-Variablen. Ihr Beitrag zu der durch die beiden Faktoren "verbal reasoning" und "space reasoning" erklärten Varianz ist *nicht signifikant* größer 0.

**age\_years** erklärt 0.2161 Varianzeinheiten von **4.2976** insgesamt. Dieser Beitrag ist mit  $M/S=3.3803$  signifikant größer 0 - wenn man, wie von Zientek/Thompson vorgeschlagen, **2.0** als Grenzwert festlegt - jedoch deutlich geringer als die M/S-Werte der 6 "Begabungsvariablen".

Wir fassen nochmals die relevanten Ergebnisse zusammen:

Zu erklarende Gesamtvarianz	8.0000		
Eigenwerte (erklärte Varianz) je Faktor	2.6519,	1.6457	Summe 4.2976
Durch 2 Faktoren erklarte Varianz	4.2976		

A	B	M/S
---	---	-----

month_si	V5	1.0000	0.0249	1.1008
age_year	V7	1.0000	0.2161	3.3803
t6_parag	V14	1.0000	0.7941	35.9875
t7_sente	V15	1.0000	0.8135	43.7374
t9_word_	V17	1.0000	0.7667	30.9222
t14_word	V22	1.0000	0.5751	11.5685
t15_num	V23	1.0000	0.5848	14.1940
t17_obje	V25	1.0000	0.5224	9.8203
Summe		8.0000	4.2976	

A = in Originalstichprobe eingesetzte Kommunalitäten  
 B = aus Originalstichprobe reproduzierte Kommunalitäten  
 bzw. Varianzbeiträge der Variablen zu der durch 2 Faktoren  
 erklärten Varianz  
 M/S = Pseudo-t -Wert aus Bootstrap

### P30.6.4.3 Ergebnisse aus Bootstrap bei schiefwinkliger Zielmatrix

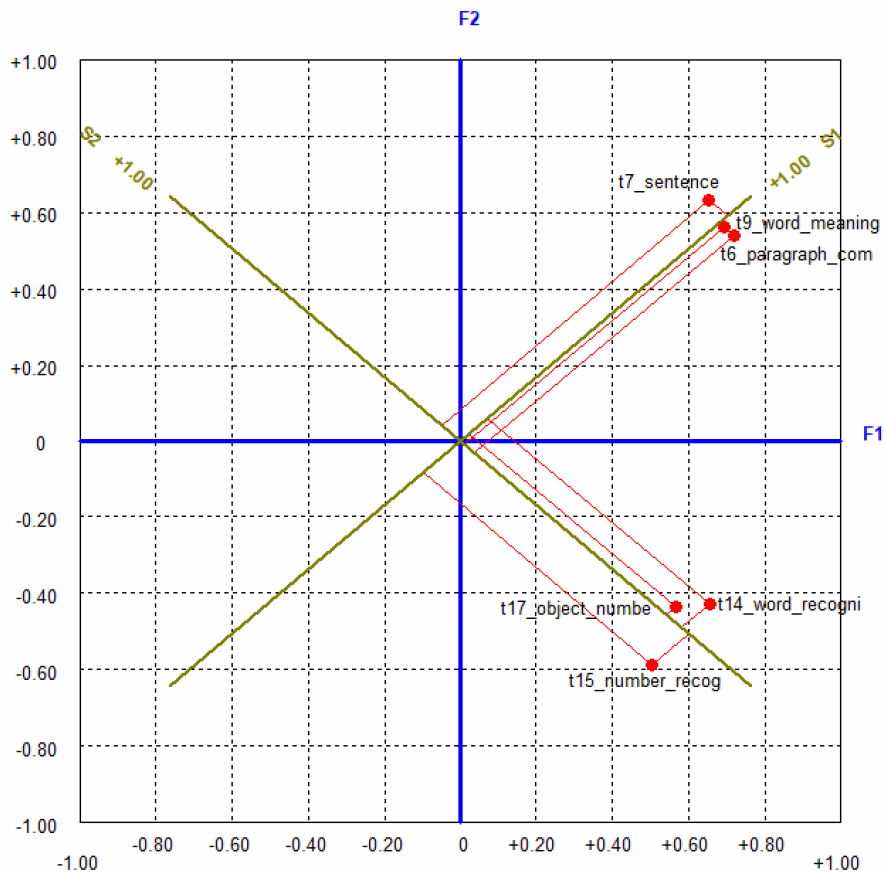
#### Die Ergebnisse aus der Originalstichprobe

Almo rechnet zuerst eine Faktorenanalyse für die Originalstichprobe. Die Faktorladungsmatrix wird mit dem Quartimin-Verfahren (=Oblimin 0) schiefwinklig rotiert. Aus der Quartimin-Matrix wird die schiefwinkliger Zielmatrix abgeleitet und die Matrix der Winkel und Korrelationen zwischen den schiefwinkligen Achsen gewonnen, die für die "Prokrustes-Anpassung" der Ladungsmatrizen aus Original- und aller Bootstrap-Stichproben gebraucht werden. Diese Matrizen sind oben in Abschnitt P30.6.2.4 bereits abgebildet worden. Wir zeigen hier das Ergebnis der Prokrustes-Anpassung, die "konfirmatorische Faktorladungsmatrix" der Originalstichprobe. Sie ist identisch mit der Quartimin-Faktorladungsmatrix, wenn Zielmatrix 2 verwendet wird.

konfirmatorische Ladungsmatrix aus der Originalstichprobe  
 Matrix der auf die schiefwinkligen Achsen  
 achsparallel projizierten Faktorladungen

		Faktor 1	Faktor 2
t6_parag	V14	0.8904	-0.0514
t7_sente	V15	0.9198	0.0648
t9_word_	V17	0.8905	-0.0152
t14_word	V22	0.0965	-0.7610
t15_num	V23	-0.1271	-0.7882
t17_obje	V25	0.0324	-0.7101

Graphisch dargestellt



Die beiden schiefwinkligen Achsen S1 und S2 stehen mit ihren positiven Schenkeln S1+ und S2+ mit einem offenen Winkel von  $180 - 80.2735 = 99,7265$  Grad aufeinander. Dem entspricht eine Korrelation zwischen den Achsen von  $-0.1689$ . Der Winkel zwischen S1+ und S2- beträgt  $-80.2735$  Grad.

### Die Bootstrap-Ergebnisse

Almo gibt zuerst die Tabelle des Bootstraps der Eigenwerte aus. Sie ist identisch mit der aus dem prthogonalen Bootstrapping.

Danach wird die Tabelle des Bootstraps der Faktorladungen ausgegeben. Wir zeigen nur die Ergebnisse für den 1. Faktor, wobei wir hier wieder die ersten 3 Tabellenspalten der überbreiten Ergebnistabelle als eigenständige Tabelle abtrennen

Bootstrap-Ergebnisse fuer Faktor 1  
 der schiefwinkligen konfirmatorischen Faktorladungen aus Originalstichprobe  
 Die Ladungspunkte wurden achsparallel auf die schiefwinkligen Achsen projiziert

Ergebnisse aus Originalstichprobe				
		*a	*b	*c
		unrotierte	Quartimin-	schiefwinkl.
		orthogonale	Faktorladung	konfirmator.
		Faktorladung	Faktorladung	Faktorladung
t6_paragraph_co	V14	0.875175	0.890783	0.890416
t7_sentence	V15	0.851345	0.919989	0.919825
t9_word_meaning	V17	0.859052	0.890753	0.890452
t14_word_recogni	V22	0.432765	0.097875	0.096459
t15_number_reco	V23	0.230952	-0.125712	-0.127109
t17_object_numbe	V25	0.348691	0.033717	0.032413

Ergebnisse aus Bootstrap der schiefwinkl. konfirmatorischen Faktorladungen Zielmatrix: -1,0,1-Matrix aus Quartimin *i)							
Mitt.wert *d	*e	*f	*g	*h			
schiefwinkl. konfirmator. Faktorladung	Verzerrung Faktorladung	Standard fehler	Signifik. p	Konfidenzintervall		Breite	
				unten	oben		
t6	0.889080	-0.001336	0.013553	0.000999	0.861709	0.914602	0.052892
t7	0.919295	-0.000530	0.010374	0.000999	0.898185	0.938155	0.039970
t9	0.889617	-0.000836	0.012544	0.000999	0.863473	0.911674	0.048200
t14	0.098959	0.002500	0.049953	0.048000	0.000834	0.199017	0.198183
t15	-0.129820	-0.002711	0.060180	0.028000	-0.248372	-0.015342	0.233030
t17	0.032445	0.000032	0.062629	0.580000	-0.092035	0.160561	0.252596

Da die beiden schiefwinkligen Achsen beinahe rechtwinklig aufeinander stehen, ist das Ergebnis gegenüber dem mit der rechtwinkligen Zielmatrix nur gering verschieden.

#### Rechtwinklig oder schiefwinklig

Natürlich stellt sich die Frage, was ist vorzuziehen, "die aus der Anpassung an die rechtwinklige oder an die schiefwinklige Zielmatrix hervorgegangene konfirmatorische Faktorladungsmatrix?". Die Frage muss unbeantwortet bleiben. Beide sind zulässig. Der Forscher muss entscheiden, welche der beiden, seinem Forschungsgegenstand und seinen Daten am ehesten entspricht.

#### P30.6.4.4 Ergebnisse aus Bootstrap bei Faktorenanalyse mit 1 Faktor

Wir zeigen hier das Ergebnis aus dem Beispiel-Programm "Prog30ml\_1\_Faktor.Alm", das oben in Abschnitt P30.6.2.0 vorgestellt wurde. Die Ausgabe wurde etwas verkürzt.

Bootstrap-Ergebnisse fuer Faktor 1  
der orthogonalen Faktorladungen aus 1-faktorieller Analyse

	Ergebnisse Original-Stichprobe	Ergebnisse aus 1000 Bootstrapstichproben der Faktorladungen					
		*a	*d	*e	*f	*g	*h
		Faktorladung	Mitt.wert Fakt.lad.	Verzerr. Fakt.lad.	Standard fehler	Signifik. p	Konf.niv=0.950 unten oben
t5_general V13	0.8239	0.8241	0.0001	0.0184	0.0010	0.7829	0.8561
t6_paragra V14	0.8405	0.8403	-0.0002	0.0163	0.0010	0.8069	0.8698
t7_sentenc V15	0.8404	0.8411	0.0007	0.0138	0.0010	0.8127	0.8665
t9_word_me V17	0.8697	0.8695	-0.0002	0.0133	0.0010	0.8428	0.8933
t22_proble V30	0.6874	0.6862	-0.0011	0.0338	0.0010	0.6169	0.7433
t23_series V31	0.6642	0.6615	-0.0028	0.0384	0.0010	0.5791	0.7311
t24_woody_ V32	0.6261	0.6240	-0.0021	0.0394	0.0010	0.5382	0.6938

- \*a unrotierte Faktorladungen aus der Originalstichprobe
- \*d Mittelwert der unrotierten Faktorladungen aus allen Bootstrap-Stichproben
- \*e mit "Verzerrung" wird die Differenz zwischen dem Mittelwert aus allen Bootstrap-Stichproben minus dem Wert aus der Originalstichprobe bezeichnet
- \*f Der Standardfehler ist gleich der Standardabweichung der unrotierten Faktorladungen aus allen Bootstrap-Stichproben
- \*g Berechnet wird die zweiseitige Signifikanz p.  
Sie wird beim einfachen Perzentil-Verfahren in folgender Weise gewonnen: Also ermittelt zuerst das Konfidenzniveau, bei dem sich der Wert 0 des Bootstrap-Koeffizient gerade noch ausserhalb des Konfidenzintervalls befindet, d.h. bei dem 0 direkt unterhalb der unteren Konfidenzgrenze oder direkt oberhalb der oberen Konfidenzgrenze plaziert ist. Dieses Konfidenzintervall wird in Also auch als "optimales Konfidenzintervall" bezeichnet. Es kann so interpretiert werden: "Gerade noch" mit der Wahrscheinlichkeit (Sicherheit) von k Prozent, die für das optimale Konfidenzintervall gefunden wurde, ist der Bootstrap-Koeffizient von 0

verschieden. Die Gegenwahrscheinlichkeit, die Irrtumswahrscheinlichkeit, der p-Wert ist dann  $p = 1-k/100$

- \*h Konfidenzintervall nach dem vom Benutzer vorgegebenen Konfidenz-Niveau von 95.00%. Beim "einfachen" Perzentil-Verfahren bedeutet das:  
Von den aufsteigend sortierten 1000 Werten aus den Bootstrap-Stichproben befinden sich 95.00% der Werte zwischen den Konfidenzgrenzen und je 2.50% oberhalb und unterhalb der Konfidenzgrenzen

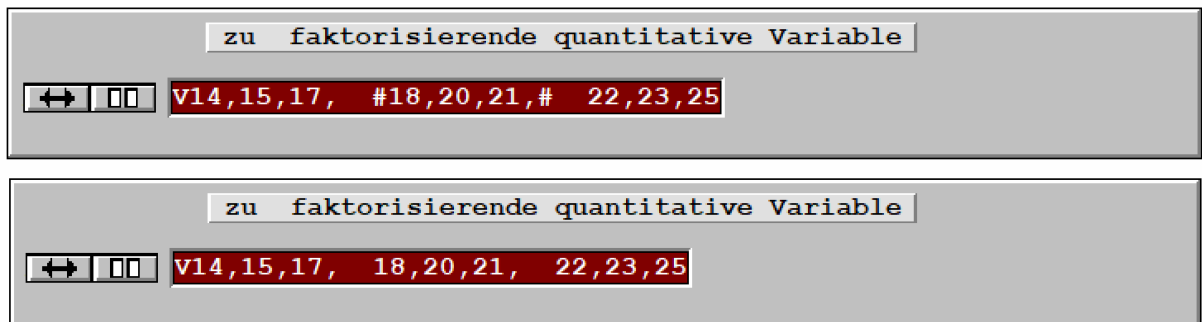
## Anhang 1: Normierung der Faktorladungen: Vergleich der Studie von Zientek & Thompson mit Almo

Dieser Abschnitt ist nur für den Leser interessant, der sich mit der Arbeit von Zientek/Thompson beschäftigt hat und der wissen möchte, worin das durch die Normierung der Faktorladungen verursachte Problem besteht

Die 3-faktorielle Bootstrap-Faktorenanalyse der 9 Items aus den Holzinger/Swineford-Daten führt bei den Programmen von Almo (in der Version ohne Normierung) und von Zientek/Thompson zu deutlich verschiedenen "prokrustes-rotierten" Faktorladungsmatrizen für die Originalstichprobe und für die zusammengefassten Bootstrapstichproben. Almo (mit Normierung) und das Zientek/Thompson-Programm erbringen die gleichen Ergebnisse.

In Almo wird es vorgezogen von "orthogonal *confirmatorischer* Faktorladungsmatrix" zu sprechen. Wir werden hier jedoch mit gleicher Bedeutung den bei Zientek/Thompson verwendeten Begriff der "orthogonal *prokrustes-rotierten* Faktorladungsmatrix" übernehmen.

Im Almo-Programm Prog30ml müssen in der Variableneingabe die Kommentarzeichen # herausgenommen werden



Die unterschiedliche Vorgehensweise lässt sich an den Ergebnissen für die Originalstichprobe zeigen. Die Varimax-rotierte Faktorladungsmatrix für die Originalstichprobe ist in beiden Programmen gleich. Das bedeutet, dass auch die in beiden Programmen ermittelten Korrelationsmatrizen und orthogonalen nicht-rotierten Faktorladungsmatrizen (die bei Zientek/Thompson nicht ausgegeben werden) identisch sind. Die beiden schlussendlich "prokrustes-rotierten" Faktorladungsmatrizen der Originalstichprobe aus Almo (in der Version ohne Normierung) und Zientek/Thompson und auch der Bootstrapstichproben sind erheblich verschieden, wenn auch "strukturell" gleich.

Betrachten wir zuerst die Bootstrap-Ergebnisse für die Eigenwerte

Ergebnisse aus Almo

	Original-Stichprobe	Ergebnisse aus 1000 Bootstrapstichproben der Eigenwerte					
		*a	*d	*e	*f	*g	*h
		Eigenwert	Mitt.wert Eigenwert	Verzerr. Eigenwert	Standard fehler	Signifik. p	Konfidenzintervall Konf.niv=0.950 unten oben
Eigenwert 1	2.9454	2.9639	0.0185	0.1757	0.0010	2.6451	3.3296
Eigenwert 2	1.7603	1.7795	0.0193	0.1233	0.0010	1.5525	2.0383
Eigenwert 3	1.3963	1.3904	-0.0060	0.1000	0.0010	1.1817	1.5849
Eigenwert 4	0.7166	0.7366	0.0200	0.0496	0.9990	0.6447	0.8400
Eigenwert 5	0.6292	0.6274	-0.0017	0.0405	0.9990	0.5531	0.7130
Eigenwert 6	0.5351	0.5303	-0.0049	0.0348	0.9990	0.4652	0.6053
Eigenwert 7	0.4780	0.4518	-0.0261	0.0345	0.9990	0.3870	0.5188
Eigenwert 8	0.2885	0.2852	-0.0033	0.0251	0.9990	0.2420	0.3409
Eigenwert 9	0.2506	0.2350	-0.0157	0.0201	0.9990	0.1967	0.2758

Siehe die Erläuterungen \*a bis \*h in Abschnitt P30.6.2.1. Durch den p-Wert in Spalte \*g wird getestet, ob ein Eigenwert größergleich 1.0 ist, also das Kaiser-Kriterium erfüllt. Das Ergebnis ist eindeutig: 3 Eigenwert sind signifikant.

#### Ergebnisse Zientek/Thompson

	Original-Stichprobe	Ergebnisse aus 1000 Bootstrapstichproben der Eigenwerte		
		*d	*f	
		Mitt.wert Eigenwert M	Standard fehler S	M/S
Eigenwert 1	2.9454	2.9586	0.1808	16.3656
Eigenwert 2	1.7603	1.7782	0.1256	14.1545
Eigenwert 3	1.3963	1.3914	0.0975	14.2729
Eigenwert 4	0.7166	0.7395	0.0518	14.2669
Eigenwert 5	0.6292	0.6265	0.0399	15.7205
Eigenwert 6	0.5351	0.5311	0.0340	15.6128
Eigenwert 7	0.4780	0.4531	0.0343	13.1961
Eigenwert 8	0.2885	0.2866	0.0259	11.0782
Eigenwert 9	0.2506	0.2351	0.0200	11.7313

Die Eigenwerte aus der Originalstichprobe bei Almo und Zientek/Thompson stimmen exakt überein. Die Mittelwerte der Bootstrap-Eigenwerte differieren nur an der 3. Kommastellen. Auch die Standardfehler differieren nur minimal.

Zientek/Thompson berechnen analog zum t-Wert einen Koeffizienten M/S aus "Mittelwert dividiert durch Standardfehler", den wir in obiger Tabelle in der letzten Spalte ausgegeben haben. Wir werden am Ende des Anhangs diesen Koeffizienten diskutieren. Hier ist er eher von geringer Bedeutung.

#### Vergleich der Faktorladungen aus der Originalstichprobe

Die Varimax-rotierte Faktorladungsmatrix von Almo und Zientek/Thompson stimmen exakt überein, ebenso die daraus abgeleitete Zielmatrix für die Prokrustes-Rotation.. Bei Zientek/Thompson wird erstere allerdings nur mit 2 Kommastellen ausgegeben. Man würde nun erwarten, dass auch die an die Zielmatrix heran rotierte Faktorladungsmatrizen der Originalstichprobe die gleichen sind. Sie sind jedoch deutlich verschieden.

	Prokrustes-rotierte Faktorladungsmatrix der Originalstichprobe Almo (nicht normiert)			Zientek/Thompson		
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3
t6_parag	0.8868	-0.0880	0.1203	0.9845	0.1077	0.1381
t7_sente	0.9061	-0.0961	-0.0010	0.9932	0.1160	0.0032
t9_word_	0.8821	-0.1013	0.0755	0.9883	0.1238	0.0890

t10_addi	0.0375	-0.7676	0.1694	0.0365	0.9749	0.2198
t12_coun	0.0482	-0.8348	0.0355	0.0468	0.9978	0.0470
t13_stra	0.1996	-0.6937	0.0776	0.2643	0.9579	0.1122
t14_word	0.1872	0.0471	0.7857	0.2283	-0.0602	0.9717
t15_numb	-0.0429	0.0097	0.7883	-0.0580	-0.0174	0.9982
t17_obje	0.0504	-0.3393	0.6532	0.0600	0.4566	0.8876

Die Faktorladungen aus Almo (nicht normiert) und Zientek/Thompson gehen zwar in die gleiche Richtung sind in ihren Werten jedoch erheblich verschieden. Alle Ladungswerte sind bei Zientek/Thompson größer.

In Almo kann die Normierung der Faktorladungen durch eine Option ebenfalls gerechnet werden. Das Almo-Ergebnis für die Originalstichprobe stimmt dann exakt mit dem bei Zientek/Thompson überein. Die Bootstrap-Ergebnisse sind minimal (zufällig) verschieden.

In Almo wird in folgender Weise normiert:

In der orthogonalen, nicht-rotierten Faktorladungsmatrix der Original-Stichprobe und auch der Bootstrap-Stichproben wird für jede Matrixzeile  $i$  die Summe  $s(i)$  der quadrierten Ladungen ermittelt. Die Summe  $s(i)$  ist die *reproduzierte* Kommunalität der Variablen  $i$ , die sich von der durch den Benutzer vorgegebenen unterscheidet. Jede Ladung der Zeile  $i$  wird dann durch die Wurzel aus  $s(i)$  dividiert. Das Ergebnis ist dann die zeilenweise normierte, orthogonale, nicht-rotierte Faktorladungsmatrix. Diese wird dann der "Prokrustes-Anpassung" unterzogen. Die Summe der quadrierten normierten Faktorladung in einer Matrixzeile ist 1.0.

Das Zientek/Thompson-Programm ist in der Matrix-Syntaxsprache von SPSS geschrieben. Das Programm kann im Internet herunter geladen werden, Es ist unter <https://www.shsu.edu/lrz002/bfa/> zu finden. Der zentrale Programmteil, die "Prokrustes-Rotation", wurde von Thompson programmiert. Eine Beschreibung des von ihm verwendeten Algorithmus konnten wir nicht finden.

Um den Algorithmus aus dem Programm zu rekonstruieren, muss man in der komplexen SPSS-Matrix-Syntaxsprache firm sein. Das sind wir nicht genügend. Der Programm-Code liegt im doc-Format in 2 Teile aufgetrennt vor und ist sehr übersichtlich gestaltet. Im 2. Teil ist durch die Programmzeile **TITLE='Normalization Factor for Rows'** die Stelle markiert, an der die Normierung der Faktorladungen vorgenommen wird.

Im Artikel von Zientek/Thompson (2007) ist kein Hinweis darauf zu finden, dass normiert wird und warum. Der Grund warum normiert wird ist allerdings offensichtlich: Es wird angenommen, dass die Faktorladungsmatrizen der Originalstichprobe und aller Bootstrapstichproben einheitlicher werden und dass es damit gelingt, sie besser in einen "gemeinsamen Faktorenraum" zu stellen. Das mag zwar gelingen, ist aber an unserem Beispiel nicht erkennbar. Die Norm mit der die Ladungen für die Faktormatrixzeile bzw. Variable  $i$  gewichtet werden, ist die Wurzel aus der reproduzierten Kommunalität der Variablen  $i$ . Die  $x$  Faktorladungen für die Variable  $i$  werden also durch ein Konstante gewichtet. Die Konstante hat jedoch für jede Zeile/Variable wieder einen anderen Wert. Graphisch betrachtet können wir dann interpretieren: Der Ladungspunkt  $i$  wird in der Punktekonfiguration im Koordinatensystem um einen anderen Betrag verschoben als der Ladungspunkt  $j$ . Daraus folgt, dass die Normierung die "Konfiguration der Ladungspunkte" verändert. Die Distanzen zwischen den Ladungspunkten und die reproduzierte Korrelationsmatrix stimmen nicht mehr mit der der Ausgangs-Faktorladungsmatrix überein.

Die Normierung verletzt, wie wir meinen, die zentrale Bedingung, dass die aus der "prokrustes-rotierte" Faktorladungsmatrix reproduzierte Korrelationsmatrix gleich sein muss mit der, die aus der ursprünglichen nicht-rotierten orthogonalen Faktorladungsmatrix hervorgegangen ist. Geometrisch formuliert: Die Punktwolke der Faktorladungen darf durch die "Prokrustes-Rotation" nicht verändert werden.

Wir vergleichen die "reproduzierten Korrelationsmatrizen" die aus folgenden Faktorladungsmatrizen entstanden

1. aus der unrotierten Faktorladungsmatrix der Originalstichprobe
2. aus der Prokrustes-rotierten Ladungsmatrix von Almo (nicht normiert)
3. aus der Prokrustes-rotierten Ladungsmatrix von Zientek/Thompson

Die Tabellen aus 1 und 2 sind identisch.

Tabelle für 1 und 2

durch 3 Faktoren reproduzierte Korrelationsmatrix der Originalstichprobe  
entstanden aus der unrotierten orthogonalen Faktorladungsmatrix  
und  
aus der Prokrustes-rotierten Ladungsmatrix von Almo (nicht normiert)

	t6_para	t7_sente	t9_word	t10_add	t12_cou	t13_str	t14_wor	t15_num	t17_obj
t6_parag	0.8087	0.8118	0.8003	0.1212	0.1205	0.2474	0.2564	0.0560	0.1532
t7_sente	0.8118	0.8302	0.8089	0.1076	0.1238	0.2474	0.1643	-0.0406	0.0776
t9_word	0.8003	0.8089	0.7941	0.1236	0.1297	0.2522	0.2197	0.0207	0.1281
t10_addi	0.1212	0.1076	0.1236	0.6193	0.6486	0.5531	0.1040	0.1245	0.3730
t12_coun	0.1205	0.1238	0.1297	0.6486	0.7004	0.5915	-0.0025	0.0178	0.3089
t13_stra	0.2474	0.2474	0.2522	0.5531	0.5915	0.5271	0.0656	0.0459	0.2961
t14_word	0.2564	0.1643	0.2197	0.1040	-0.0025	0.0656	0.6545	0.6118	0.5067
t15_num	0.0560	-0.0406	0.0207	0.1245	0.0178	0.0459	0.6118	0.6234	0.5095
t17_obje	0.1532	0.0776	0.1281	0.3730	0.3089	0.2961	0.5067	0.5095	0.5444

Aus der "prokrustes-rotierten" Faktorladungsmatrix entsteht bei Zientek/Thompson. die folgende reproduzierte Korrelationsmatrix, die erheblich andere Werte beinhaltet, als die der unrotierten Faktorladungsmatrix

Tabelle für 3

reproduzierte Korrelationsmatrix der Originalstichprobe  
(entstanden aus der Prokrustes-rotierten Ladungsmatrix von Zientek/Thompson)

	t6_para	t7_sente	t9_word	t10_add	t12_cou	t13_str	t14_wor	t15_num	t17_obj
t6_parag	0.9999	0.9907	0.9986	0.1713	0.1600	0.3789	0.3525	0.0789	0.2308
t7_sente	0.9907	0.9999	0.9962	0.1500	0.1624	0.3740	0.2229	-0.0564	0.1154
t9_word	0.9986	0.9962	1.0000	0.1763	0.1740	0.3898	0.3047	0.0294	0.1948
t10_addi	0.1713	0.1500	0.1763	1.0001	0.9848	0.9682	0.1632	0.2003	0.6424
t12_coun	0.1600	0.1624	0.1740	0.9848	1.0000	0.9734	-0.0037	0.0268	0.5001
t13_stra	0.3789	0.3740	0.3898	0.9682	0.9734	1.0000	0.1117	0.0800	0.5528
t14_word	0.3525	0.2229	0.3047	0.1632	0.0037	0.1117	0.9999	0.9578	0.8487
t15_num	0.0789	0.0564	0.0294	0.2003	0.0268	0.0800	0.9578	1.0001	0.8746
t17_obje	0.2308	0.1154	0.1948	0.6424	0.5001	0.5528	0.8487	0.8746	0.9999

Der Wert 0.9999 bzw. 1.0001 in der Diagonale bedeutet 1.0. Wir konnten nur mit 4 Kommastellen die reproduzierte Korrelationsmatrix aus der normierten Faktorladungsmatrix berechnen.

Bei der Rotation wird (graphisch betrachtet) die Punktekonfiguration "in geschlossener Formation" im Koordinatenraum gedreht. Dabei müssen alle euklidischen Distanzen zwischen allen Punkten gleich bleiben. Wir haben die Distanzmatrix für die nicht-rotierte, orthogonale Faktorladungsmatrix ermittelt (mit Almo-Prog30mf.Msk) und mit der Distanzmatrix aus der oben abgebildeten Prokrustes-rotierten Faktorladungsmatrix aus Almo (nicht normiert) und mit der normierten von Zientek/Thompson verglichen. Bei

Almo (nicht normiert) stimmt sie überein, bei Zientek/Thompson nicht. Das bedeutet, dass bei Zientek/Thompson, die Punktekonfiguration durch die Prokrustes-Rotation verändert wurde

### Vergleich der Faktorladungen aus den Bootstrapstichproben

Die Faktorladungsmatrizen der Bootstrapstichproben werden in Richtung auf die gleiche Zielmatrix (wie bei der Originalstichprobe) mit demselben Algorithmus "prokrustes-rotiert". Wie nicht anders zu erwarten war stimmen die Bootstrap-Ergebnisse von Almo (nicht-normiert) und normiert bei Zientek/Thompson in der Tendenz, aber nicht in den Werten überein. Bei Almo mit normierten Faktorladungen und Zientek/Thompson stimmen die Werte mit nur minimalen (zufälligen) Differenzen überein

**Bootstrap-Ergebnisse aus Almo - Faktorladungen nicht normiert**  
Tabelle gekürzt

Var.	Faktor 1				Faktor 2				Faktor 3			
	Sample	Boot-strap	SE	signif. p	Sample	Boot-strap	SE	signif. p	Sample	Boot-strap	SE	signif. p
T6	.8868	.8848	.0133	.0010	-.0880	-.0895	.0383	.0180	.1203	.1206	.0345	.0020
T7	.9061	.9048	.0099	.0010	-.0961	-.0962	.0385	.0200	-.001	-.0014	.0412	.9520
T9	.8821	.8802	.0124	.0010	-.1013	-.1027	.0432	.0200	.0755	.0747	.0389	.0480
T10	.0375	.0377	.0533	.4900	-.7676	-.7671	.0306	.0010	.1694	.1707	.0598	.0060
T12	.0482	.0511	.0506	.3180	-.8348	-.8329	.0204	.0010	.0355	.0318	.0596	.5980
T13	.1996	.1996	.0626	.0020	-.6937	-.6913	.0406	.0010	.0776	.0792	.0642	.2100
T14	.1872	.1896	.0479	.0010	.0471	.0499	.0555	.3460	.7857	.7828	.0276	.0010
T15	-.043	-.046	.0594	.4400	.0097	.0084	.0604	.8940	.7883	.7850	.0253	.0010
T17	.0504	.0502	.0607	.3780	-.3393	-.3399	.0650	.0010	.6532	.6512	.0516	.0010

**Bootstrap-Ergebnisse aus Almo - Faktorladungen normiert**  
Tabelle gekürzt

Var.	Faktor 1				Faktor 2				Faktor 3			
	Sample	Boot-strap	SE	signif. p	Sample	Boot-strap	SE	signif. p	Sample	Boot-strap	SE	signif. p
T6	.9845	.9825	.0078	.0010	-.1077	-.1093	.0442	.0140	.1381	.1384	.0397	.0010
T7	.9932	.9912	.0052	.0010	-.1160	-.1161	.0435	.0080	.0032	.0027	.0471	.9660
T9	.9883	.9859	.0078	.0010	-.1238	-.1253	.0500	.0120	.0890	.0880	.0453	.0460
T10	.0365	.0369	.0676	.5880	-.9749	-.9691	.0187	.0010	.2198	.2205	.0768	.0060
T12	.0468	.0505	.0602	.4320	-.9978	-.9934	.0062	.0010	.0470	.0423	.0725	.5680
T13	.2643	.2633	.0841	.0020	-.9579	-.9494	.0287	.0010	.1122	.1150	.0901	.1920
T14	.2283	.2314	.0598	.0010	.0602	.0628	.0682	.3200	.9717	.9665	.0158	.0010
T15	.0580	-.062	.0751	.4140	.0174	.0149	.0767	.8440	.9982	.9922	.0079	.0010
T17	.0600	.0596	.0814	.4260	-.4566	-.4555	.0879	.0010	.8876	.8789	.0465	.0010

**Bootstrap-Ergebnisse aus Zientek/Thompson**

Var.	Faktor 1				Faktor 2				Faktor 3			
	Sample	Boot-strap	SE	MBR/SE	Sample	Boot-strap	SE	MBR/SE	Sample	Boot-strap	SE	MBR/SE
T6	.9845	.9824	.0073	<u>133.96</u>	.1077	.1089	.0454	2.40	.1381	.1391	.0388	3.59
T7	.9932	.9913	.0055	<u>181.54</u>	.1160	.1149	.0451	2.55	.0032	.0055	.0456	0.12
T9	.9883	.9863	.0078	<u>126.39</u>	.1238	.1239	.0490	2.53	.0890	.0867	.0443	1.96
T10	.0365	.0369	.0687	0.54	<u>.9749</u>	.9689	.0189	<u>51.15</u>	.2198	.2210	.0768	2.88
T12	.0468	.0486	.0614	0.80	<u>.9978</u>	.9935	.0061	<u>162.58</u>	.0470	.0433	.0700	0.62
T13	.2643	.2622	.0811	3.23	<u>.9579</u>	.9502	.0271	<u>35.09</u>	.1122	.1134	.0908	1.25
T14	.2283	.2296	.0647	3.55	.0602	.0633	.0719	0.88	<u>.9717</u>	.9662	.0174	<u>55.46</u>
T15	.0580	.0583	.0776	0.75	.0174	.0107	.0718	0.15	<u>.9982</u>	.9926	.0074	<u>133.77</u>
T17	.0600	.0600	.0829	0.72	.4566	.4517	.0882	5.12	<u>.8876</u>	.8806	.0469	<u>18.76</u>

Note: MBR stands for mean bootstrap results, SE for the standard error

## Pseudo-t -Wert

Zientek/Thompson berechnen analog zum t-Wert einen Koeffizienten MBR/SE aus "Mittelwert dividiert durch Standardfehler", der in obiger Tabelle in der jeweils letzten Spalte ausgegeben wird.

Er wird bei Zientek/Thompson dazu verwendet um *annähernd* beurteilen zu können, ob Faktorladungen signifikant ungleich 0 sind. Er darf aber nicht streng als t-Wert verstanden werden. Zientek/Thompson meinen aber dann doch, dass ein Quotient MBR/SE größergleich 2.0 darauf hinweist, dass der betreffende Koeffizient ungleich 0 sein könnte (Seite 323)

Dieser Pseudo-t-Wert ist eine Art Stabilitätsindex bzw. informeller Signifikanzwert, der sich wie man an obiger Tabelle erkennt, als brauchbar erweist. Wie der p-Wert von Almo gibt er an, ob die Faktorladung der Variablen i auf dem Faktor j gleich 0 sein könnte oder ob die Variable i signifikant auf dem Faktor j lädt. Werte >2 werden gedeutet als "ist signifikant", und <2 als "ist nicht-signifikant".

Wir vergleichen die p-Werte aus Almo (nicht-normiert) mit den **MBR/SE**-Werten aus Zientek/Thompson. **MBR/SE**-Werte größer 2.0 und der jeweils entsprechende p-Wert sind unterstrichen

	Faktor 1		Faktor 2		Faktor 3	
	p	MBR/SE	p	MBR/SE	p	MBR/SE
T6	<u>0.0010</u>	<u>133.96</u>	0.0180	2.40	<u>0.0020</u>	<u>3.59</u>
T7	<u>0.0010</u>	<u>181.54</u>	<u>0.0200</u>	<u>2.55</u>	0.9520	0.12
T9	<u>0.0010</u>	<u>126.39</u>	<u>0.0200</u>	<u>2.53</u>	0.0480	1.96
T10	0.4900	0.54	<u>0.0010</u>	<u>51.15</u>	<u>0.0060</u>	<u>2.88</u>
T12	0.3180	0.80	<u>0.0010</u>	<u>162.58</u>	0.5980	0.62
T13	<u>0.0020</u>	<u>3.23</u>	<u>0.0010</u>	<u>35.09</u>	0.2100	1.25
T14	<u>0.0010</u>	<u>3.55</u>	0.3460	0.88	<u>0.0010</u>	<u>55.46</u>
T15	<u>0.4400</u>	<u>0.75</u>	0.8940	0.15	<u>0.0010</u>	<u>133.77</u>
T17	0.3780	0.72	0.0010	5.12	<u>0.0010</u>	<u>18.76</u>

p-Wert und MBR/SE entsprechen sich recht gut.

## Anhang 2: Vertauschte Faktoren bei den Bootstrapschproben

Die Ladungsmatrizen der Bootstrapschproben können nicht nur auf Grund der oben in Abschnitt P30.6.2.3.1 erwähnten Spiegelung anders im Koordinaten-Raum liegen, sondern auch weil die Faktoren im Vergleich zur Originalstichprobe vertauscht sind. Es ist durchaus möglich, dass unter den vielen Bootstrapschproben einige wenige dabei sind, bei denen beispielsweise die 2. Variablengruppe (t14, t15, t17) dem 1. Faktor zugeordnet werden kann und nicht dem 2. Faktor, wie bei der Originalstichprobe, und entsprechend die erste Variablengruppe dem 2. Faktor. Bei nur 2 Faktoren dürften allerdings Vertauschungen sehr selten ein, jedoch nicht bei 3 und mehr Faktoren, wie wir noch zeigen werden.

Wir wollen gleich vorweg nehmen, dass Faktoren-Vertauschungen kein Problem darstellen. Durch die Werte -1,0,1 in der Zielmatrix wird den Variablen "gezeigt wo sie hin gehören". Allgemein gilt: Durch den Kalkül der konfirmatorischen Faktorenanalyse werden Faktoren-Vertauschung automatisch wieder rückgängig gemacht. Sie bilden kein Problem für das Bootstrapping der Faktorenanalyse. Wir werden das im Folgenden anhand

eines Beispiels mit rechtwinkliger Zielmatrix zeigen. Bei schiefwinkliger Zielmatrix wäre der Ablauf derselbe.

Zientek/Thompson berichten, dass sie in ihrer Analyse der Holzinger-Swineford-Daten mit 9 ausgewählten Variablen und 3 Faktoren in der 334. Bootstrapsstichprobe (von 1000) eine Vertauschung von Faktor 2 und 3 entdeckt haben. Sie werden dabei vermutlich nicht systematisch gesucht haben. Tatsächlich treten Faktoren-Vertauschungen sehr häufig auf

Wir rechnen ebenfalls mit diesen 9 Variablen und 3 Faktoren und 1000 Bootstrapsstichproben. Die Referenz ist prinzipiell die Originalstichprobe. Die 3 Variablengruppen sind den Faktoren 1,2,3 zugeordnet. Wir haben (indem wir in den Quellcode unseres Programms eingegriffen haben) in jeder Bootstrapsstichprobe untersucht, ob die absolut maximalen varimax-rotierten Ladungen

**der 2. Variablengruppe auf Faktor 3 liegen  
und die  
der 3. Gruppe auf Faktor 2.**

War dies der Fall, dann galt dies als Vertauschung. Dieses Verfahren ist ausschließlich auf unsere Beispieldaten mit 3 deutlich voneinander getrennten Variablengruppen ("Punktwolken") und 3 Faktoren ausgerichtet. Den Versuch einen allgemein verwendbaren Algorithmus zu entwickeln, mit dem es gelingt alle Vertauschungen zwischen allen Faktoren zu entdecken, unternehmen Zhang, Preacher, Luo (2010). Es erweist sich, dass es außerordentlich aufwendig ist, alle Faktor-Vertauschungen, auch solche zwischen nicht-benachbarten Faktoren, zu entdecken.

Da sich aber beim Kalkül der konfirmatorischen Faktorenanalyse das Problem der Faktor-Vertauschung und Vorzeichen-Umkehrung nicht stellt, wenn beim Bootstrapping der Kalkül der konfirmatorischen Faktorenanalyse eingesetzt wird, wollen wir dieses Thema nicht weiter behandeln.

Mit unserem ad-hoc-Verfahren haben wir 201 Vertauschungen zwischen Faktor 2 und 3 gefunden, dabei auch gleich eine Vertauschung in den Bootstrapsstichprobe Nr. 1 bis 4. Deren Ergebnisse können wir uns ausgeben lassen. Das kann in der Bootstrap-Optionsbox (in der Sub-Box "Spezielle Ergebnisse") angefordert werden.

Bei schiefwinkliger Zieldimension wurden 199 Vertauschungen festgestellt, wobei die Bootstrapsstichproben mit vertauschten Faktoren fast zu 100 Prozent dieselben sind.

**Nummern der Bootstrapsstichproben mit vertauschten Faktoren 2 und 3  
bei Zielmatrix**

<b>rechtwinklig</b>	<b>schiefwinklig</b>
1	1
2	2
3	3
4	4
-	11
13	13
17	17
24	24
25	25
.	.
.	.
.	.
970	970
975	975
979	979

980            980  
 -                994

Wir wollen nun untersuchen, wie sich die Vertauschung in Bootstrapstichprobe 1 auf den Bootstrap-Prozeß auswirkt.

----- unrotierte orthogonale Faktorladungen -----

	Originalstichprobe			Bootstrapstichprobe 1 Faktoren 2 und 3 vertauscht		
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3
t6_parag	0.7764	0.4485	0.0684	0.8139	0.1432	-0.3910
t7_sente	<u>0.7487</u>	0.5184	-0.0291	<u>0.7829</u>	0.2194	-0.4317
t9_word_	<u>0.7622</u>	0.4609	0.0254	<u>0.8257</u>	0.2257	-0.2580
t10_spee	0.4811	<u>-0.5076</u>	-0.3608	0.5613	-0.0591	0.5756
t12_coun	0.4718	<u>-0.4708</u>	-0.5062	0.5191	0.2454	0.6221
t13_stra	0.5339	-0.3150	-0.3779	0.6027	0.1568	0.3858
t14_word	0.4231	-0.2583	<u>0.6394</u>	0.3598	-0.6922	-0.1664
t15_num	0.2650	-0.4269	<u>0.6090</u>	0.1716	-0.7413	0.0302
t17_obje	0.4612	-0.5004	0.2851	0.4540	-0.6322	0.0997

Wir haben oben (zu Beginn von Abschnitt P30.6.2.3) bereits ausgeführt, dass bei grafischer Interpretation der Faktorladungsmatrix "eigenständige Punktwolken" erst nach dem Rotieren erkannt werden können. In oben abgebildeter Originalstichprobe besitzen die Variablen der Gruppe 1 (t6, t7, t9) ihre absolute maximale Ladung auf Faktor 1. Bei der 2. Variablen-Gruppe liegt nur eine Variable (t10) mit absolut maximaler Ladung auf Faktor 2. In Bootstrapstichprobe 1 besitzen zwei Variablen der Gruppe 2 (t10 und t12) absolute maximale Ladungen auf Faktor 3.

Eindeutige Verhältnisse existieren jedoch bei den varimax-rotierten Faktorladungen. Die Vertauschung der Faktoren 2 und 3 ist in den beiden nachfolgenden Tabellen klar erkennbar. Bei der Originalstichprobe liegen die absolut maximalen Ladungen der Variablen der Gruppe 1 auf Faktor 1, die von Gruppe 2 auf Faktor 2 und die von Gruppe 3 auf Faktor 3. In der Ladungsmatrix der Bootstrapstichprobe 1 sind Faktor 2 und 3 klar vertauscht.

Derartige Eindeutigkeit ist jedoch nicht immer gegeben. Wenn Punktwolken in der Originalstichprobe nicht so eindeutig voneinander getrennt sind wie in unserem Beispiel, dann kann es schwierig bis unmöglich werden, Faktoren-Vertauschungen in Bootstrapstichproben zu erkennen.

Prinzipiell gilt: Faktoren-Vertauschungen werden erst nach der Varimax-Rotation (im schiefwinkligen Fall nach der Quartimin- oder Gruppen-Rotation) eindeutig erkannt.

----- Varimax-rotierte Faktorladungen -----

	Originalstichprobe			Bootstrapstichprobe 1 Faktoren 2 und 3 vertauscht		
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3
t6_parag	<u>0.8863</u>	-0.0944	0.1191	<u>0.8928</u>	-0.1299	0.1478
t7_sente	<u>0.9058</u>	-0.0987	-0.0024	<u>0.9126</u>	-0.0487	0.1098
t9_word_	<u>0.8817</u>	-0.1063	0.0739	<u>0.8485</u>	-0.0523	0.2769
t10_spee	0.0350	<u>-0.7727</u>	0.1449	0.0907	-0.2199	<u>0.7702</u>

t12_coun	0.0457	<u>-0.8356</u>	0.0089	0.1140	0.0834	<u>0.8347</u>
t13_stra	0.1974	<u>-0.6964</u>	0.0558	0.2896	-0.0325	<u>0.6722</u>
t14_word	0.1858	0.0215	<u>0.7871</u>	0.1858	<u>-0.7740</u>	-0.0518
t15_num	-0.0444	-0.0153	<u>0.7881</u>	-0.0855	<u>-0.7567</u>	-0.0014
t17_obje	0.0481	-0.3601	<u>0.6422</u>	0.1234	<u>-0.7407</u>	0.2279

Almo errechnet die varimax-rotierten Faktorladungen nur für die Originalstichprobe. Und dies auch nur, um die Zielmatrix zu gewinnen, an die dann alle Stichproben "herangezungen" werden. Um zu prüfen, ob Faktoren vertauscht sind, muss man somit mühsam die Bootstrapstichproben varimax-rotieren. Dazu kann Programm-Maske Prog30m4.Msk verwendet werden.

Aus der varimax-rotierten Faktorladungsmatrix der Originalstichprobe wird die folgende Zielmatrix gewonnen.

Vorgegebene Zielmatrix  
aus Varimax-rotierter  
Originalstichprobe

	Faktor 1	Faktor 2	Faktor 3
t6_parag	1.0000	0	0
t7_sente	1.0000	0	0
t9_word_	1.0000	0	0
t10_spee	0	-1.0000	0
t12_coun	0	-1.0000	0
t13_stra	0	-1.0000	0
t14_word	0	0	1.0000
t15_num	0	0	1.0000
t17_obje	0	0	1.0000

Wir können etwas zugespitzt formulieren: Durch die Zielmatrix werden die Variablen-gruppen "angewiesen" ihre absolut maximalen Ladungen bei Faktor 2 zu plazieren.

#### Bei schiefwinkliger Zielmatrix

Die unrotierten orthogonalen Ladungsmatrizen der Originalstichprobe und der Bootstrapstichprobe 1 werden an die Zielmatrix "herangezungen". Dazu muss noch eine Transformationsmatrix T errechnet werden. Im Kalkül für T wird berücksichtigt ob die unrotierte Faktorladungsmatrix oder die varimax-rotierte Faktorladungsmatrix an die Zielmatrix angepasst werden soll. In Almo werden unrotierten Faktorladungsmatrizen eingesetzt; die varimax-rotierten sind eher ein zeitaufwendiger Umweg. Siehe die Formeln zu T im Almo-Dokument "Konfirmatorische Faktorenanalyse".

----- Transformationsmatrix -----

	Originalstichprobe			Bootstrapstichprobe 1		
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3
Faktor 1	0.7799	-0.5409	0.3150	0.7799	-0.5409	0.3150
Faktor 2	0.2716	-0.1609	-0.9489	0.2716	-0.1610	-0.9489
Faktor 3	-0.5639	-0.8256	-0.0215	-0.5640	-0.8255	-0.0214

Aus der Multiplikation der unrotierten orthogonalen Faktorladungen der beiden Stichproben entstehen dann die konfirmatorischen orthogonalen Faktorladungen nach der Formel

$$B = A * T$$



Zhang, Preacher, Luo: Bootstrap Confidence Intervals for Ordinary Least Squares Factor and Correlations in Exploratory Factor Analysis, *Multivariate Behavioral Research*, 2010,45:1,S.104-134  
(<https://doi.org/10.1080/00273170903504836>)

*Siehe auch Literatur-Angabe in Almo-Dokument Nr 15, "Faktorenanalyse" und  
Nr.16 "Konfirmatorische Faktorenanalyse"*