

# Skript Einführung in SPSS

## Faktorenanalyse

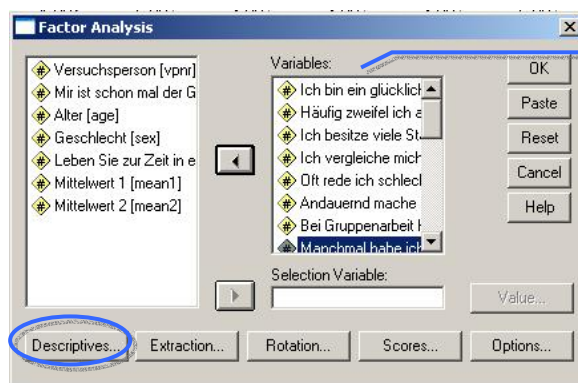
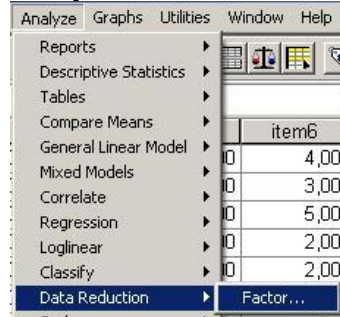
- Explorative Faktorenanalyse
  - Hauptkomponentenanalyse (PCA)
- Hinweise zum Schreiben des statistischen Reports

- 
1. Sämtliche Tabellen und Abbildungen im Report sollten nummeriert und beschriftet sein!
  2. Der Report sollte so gestaltet sein, dass jemand, der die Daten nicht kennt, alle Berechnungen nachvollziehen kann!
  3. Falls Ihr Euch bei der Bedeutung bestimmter Output-Elemente unsicher seid, könnt Ihr den *Result Coach* verwenden: Entsprechendes Element anklicken, dann rechte Maustaste ⇒ *Result Coach* (bzw. Ergebnis-Assistent in deutschen Versionen): Hier wird Euch Schritt für Schritt an einem Beispiel jede Kenngröße erklärt!

## Faktorenanalyse

### Hauptkomponentenanalyse (Principal Components Analysis)

Analysieren  $\Rightarrow$  Dimensionsreduktion (*data reduction*)  $\Rightarrow$  Faktorenanalyse (*factor analysis*)



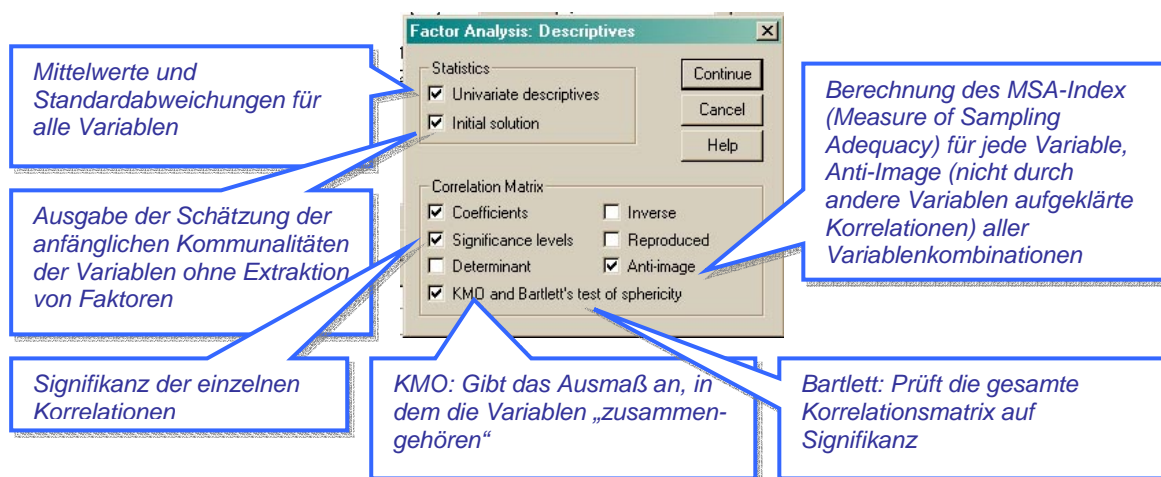
In dem Menü im linken Fenster alle Variablen, die in die Faktorenanalyse eingehen sollen, auswählen

### Einstellungen

#### Deskriptive Statistik, Eignung der Daten

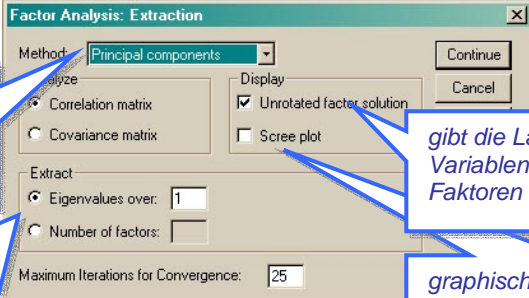
Unter dem Button **Deskriptive Statistik** (*Descriptives...*) kann man sich die deskriptiven Statistiken und Korrelationen aller Variablen ausgeben lassen. Darüber hinaus werden hier die Einstellungen vorgenommen, um die Eignung der Daten für eine Faktorenanalyse zu prüfen.

Es kann hilfreich sein, sich verschiedene Maße zur Eignung der Variablen anzusehen, um ein differenziertes Bild zu erhalten!



## Faktorextraktion

Im Menü Extraktion (Extraction) wird die Art der Faktorextraktion festgelegt.



*Als Methode solltet Ihr die Hauptkomponentenanalyse wählen (Principal Components)*

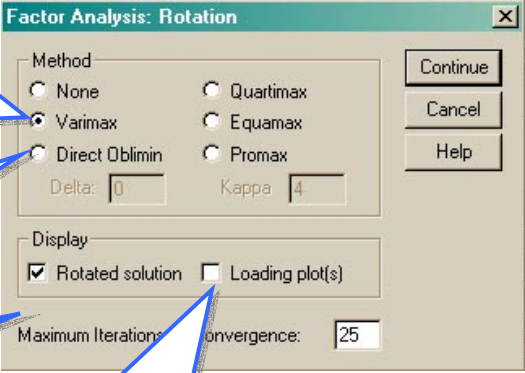
*gibt die Ladungen aller Variablen auf den unrotierten Faktoren an*

*graphische Entscheidungshilfe bei der Bestimmung der Anzahl der Faktoren*

*Standardeinstellung ist, dass alle Faktoren mit einem Eigenwert über 1 extrahiert werden. Diese Einstellung kann geändert werden, wenn man z.B. nur Faktoren mit sehr hohen Eigenwerten untersuchen will. Optional ist es auch möglich, die Anzahl der Faktoren von vornherein zu begrenzen (Number of factors).*

## Rotation der Faktoren

Im Menü Rotation kann festgelegt werden, ob und mit welcher Methode die Faktoren rotiert werden sollen.



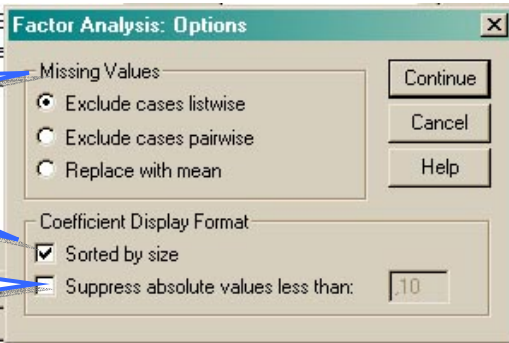
*Die einfachste Interpretation ermöglicht die Varimax-Rotation. Dabei werden die Faktoren so rotiert, dass hohe Faktorladungen noch stärker, niedrige noch niedriger werden (orthogonale) Faktoren. Somit wird die Varianz der Faktorladungen pro Faktor maximiert. Die Faktoren sind unabhängig (orthogonal).*

*Für eine schiefwinklige Rotation (die Faktoren korrelieren dann untereinander) ist die Oblimin-Methode gut geeignet.*

*gibt die Faktorladungen aus, die sich durch die Rotation ergeben*

*stellt die Ladungen der Variablen auf den ersten drei Faktoren graphisch dar*

## Optionen



*Legt fest, wie mit Missings verfahren werden soll*

*Die Variablen werden absteigend nach Höhe ihrer Faktorladungen auf den einzelnen Faktoren sortiert*

*Die Ausgabe geringer Faktorladungen kann der Übersichtlichkeit wegen auch unterdrückt werden. Festzulegen ist dann ein Mindestwert (z.B. 0.3)*

## Werte

Sollen die Ergebnisse (Faktoren) der Analyse später weiterverwendet werden, bietet SPSS die Möglichkeit, die Faktoren zu speichern. Dazu wählt das Dialogfeld **Werte** aus und aktiviert das Kästchen **Als Variablen speichern**.

Durch die Aktivierung dieses Feldes werden in der Datendatei so viele neue Variablen eingefügt, wie Faktoren extrahiert wurden.

faktorenanalyse - SPSS Daten-Editor

	mean1	mean2	fac1_1	fac2_1	fac3_1
1	4,33	2,15	1,15279	-,23082	-,14242
2	3,33	1,08	,52609	-,77352	-,77369
3	3,78	,62	,76630	-,85731	-1,00923
4	3,56	2,00	,64810	-,16800	-,73533
5	1,78	1,92	-1,39493	-,32505	,36218
6	3,83	3,00	,78114	2,12608	-,05445
7	3,06	,54	-,54592	-,49461	-,67873
8	2,11	2,23	-1,55258	,31805	,51609
9	2,83	1,31	-,38070	-1,06737	-,22459
10	4,28	,85	1,63428	-,69795	-,62119
11	2,28	2,62	-1,34740	0,5327	-,40551

**Faktorenanalyse: Faktorwerte**

☒ Als Variablen speichern

Methode

☒ Regression

☐ Bartlett

☐ Anderson-Rubin

☐ Koeffizientenmatrix der Faktorwerte anzeigen

Weiter

Abbrechen

Hilfe

Die den neuen Faktorvariablen zugeordneten Werte werden geschätzt. Dazu stehen in SPSS diese drei Methoden zur Verfügung. (Regression ist die Standardeinstellung)

- ☞ Alle drei Methoden führen zu einem Variablenmittelwert von Null.
- ☞ Unter der Hauptkomponentenanalyse (dem in den Sozialwissenschaften gebräuchlichsten Verfahren zur Durchführung einer Faktorenanalyse) ergeben alle drei Methoden die gleichen Faktorwerte.

## Output

### Eignung der Daten für die Faktorenanalyse

Nach den deskriptiven Statistiken wird die Korrelationsmatrix aller Variablen ausgegeben. Je höher der Anteil hoher Korrelationen, umso besser sind die Daten für eine Faktorenanalyse geeignet.

Im unteren Teil dieser Tabelle befindet sich eine zweite Matrix, diese enthält die p-Werte der Signifikanztests der Korrelationen. Analog gilt, je mehr Korrelationen signifikant sind, umso besser.

Korrelationsmatrix							
		Ich bin ein glücklicher Mensch.	Häufig zweifel ich an mir selbst	Ich besitze viele Stärken.	Ich vergleiche mich oft mit anderen Personen.	Oft rede ich schlecht über mich selbst.	Andauernd mache ich mir Gedanken, was an mir besser sein könnte.
Korrelation	Ich bin ein glücklicher Mensch.	1,000	,395	,186	,104	,184	,107
	Häufig zweifel ich an mir selbst	,395	1,000	,107	,441	,488	,466
	Ich besitze viele Stärken.	,186	,107	1,000	,191	,068	
	Ich vergleiche mich oft mit anderen Personen.	,104	,441	,191	1,000	,392	
	Oft rede ich schlecht über mich selbst.	,184	,488	,068	,392	1,000	
	Andauernd mache ich mir Gedanken, was an mir besser sein könnte.	,107	,466	,163	,606	,555	
Signifikanz (1-seitig)	Ich bin ein glücklicher Mensch.		,000	,066	,201	,068	
	Häufig zweifel ich an mir selbst		,000	,194	,000	,000	
	Ich besitze viele Stärken.		,066	,194	,061	,294	
	Ich vergleiche mich oft mit anderen Personen.		,201	,000	,061	,001	,000
	Oft rede ich schlecht über mich selbst.		,068	,000	,294	,001	,000
	Andauernd mache ich mir Gedanken, was an mir besser sein könnte.		,195	,000	,094	,000	,000

☞ **Aber Vorsicht:** Die Signifikanz hängt auch von der Anzahl der Versuchspersonen ab! Auch geringe Korrelationen können bei genügend hoher Anzahl signifikant werden, sind damit aber trotzdem nicht besser für eine Faktorenanalyse geeignet!

Schneller überschaubar ist die nächste Ausgabe: **KMO-Index** und **Bartlett-Test**. Damit wird überprüft, ob ein nennenswerter Zusammenhang zwischen allen Variablen besteht. Ist dies nicht der Fall, macht die Faktorenanalyse keinen Sinn.

☞ Der KMO-Index gilt als bestes zur Verfügung stehendes Verfahren zur Prüfung der Eignung der Daten!

KMO- und Bartlett-Test		
Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,725
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	93,576
	df	15
	Signifikanz nach Bartlett	,000

☞ Ein KMO-Index von .73 entspricht einem „ziemlich guten“ Ausmaß an Interkorrelation zwischen allen Variablen (zur Bewertung der Höhe vgl. Folien!)  
 ☞ Der Bartlett-Test prüft die  $H_0$ , dass in der Population kein Zusammenhang zwischen den Variablen besteht. Wird der Test signifikant, ist diese Hypothese mit einer Irrtumswahrscheinlichkeit von höchstens 5% widerlegt.

☞ Es existieren weitere Tests, die dem Bartlett-Test überlegen sind (z.B. Tests von Brien, Steiger), diese sind jedoch in SPSS nicht realisiert.

☞ Deshalb sollte der Bartlett-Test nicht in Eurem Bericht erscheinen, greift lieber auf KMO und MSA-Index zurück!

Die folgende Tabelle ist eine **Anti-Image-Matrix**. Anti-Image ist der Teil einer Korrelation, der nicht durch andere Variablen erklärt werden kann. Es werden also alle anderen Variablen aus einer Korrelation auspartialisiert (somit also Partialkorrelationen höherer Ordnung  $r_{12.34...}$  berechnet). Das Anti-Image berechnet sich als  $1 - r_{12.34...}$ .

☞ Die Begriffe *Image* und *Anti-Image* stammen von Guttman. *Image* bedeutet soviel wie „Abbildung einer Variablen durch andere Variablen“, also der gemeinsame Varianzanteil zweier Variablen, der durch alle anderen Variablen erklärbar ist. Entsprechend wird der Teil, der nicht durch die anderen Variablen aufgeklärt wird, als *Anti-Image* bezeichnet. Dieser ist bis zu einem bestimmten Grad vergleichbar mit den Residuen in der Regression, nur dass es sich dabei um den nicht erklärten Varianzanteil **einer** Variablen (des Kriteriums) handelt, während es hier um den nicht erklärten **gemeinsamen** Varianzanteil **zweier** Variablen, also um eine *Korrelation*, geht.

Da es für die Faktorenanalyse wichtig ist, dass die Variablen hoch korrelieren und Ähnliches messen, sollte der nicht erklärte gemeinsame Varianzanteil zweier Variablen, das Anti-Image gering sein. Je mehr Werte in der Anti-Image-Matrix nahe 0 sind, umso besser sind die Daten für eine Faktorenanalyse geeignet.

**Achtung:** In der **Diagonale** stehen **keine Anti-Image**werte, sondern der **MSA-Index** für jede Variable! Dieser entspricht dem KMO-Index, nur wird dieser in dem Fall nicht für die gesamte Matrix, sondern für einzelne Variablen auf Grundlage der Anti-Image-Matrix berechnet. Somit gibt er an, inwiefern eine Variable „zu den restlichen passt“. Je höher der Wert, desto stärker die Zusammengehörigkeit (vgl. Folien).

⇒ Beispiel: Ein MSA-Index von .53 bedeutet „kläglich“!

Anti-Image-Matrizen				
		Ich bin ein glücklicher Mensch.	Häufig zweifel ich an mir selbst	Ich besitze viele Stärken.
Anti-Image-Kovarianz	Ich bin ein glücklicher Mensch.	,823	-,320	-,140
	Häufig zweifel ich an mir selbst	-,320	,843	-,034
	Ich besitze viele Stärken.	-,140	-,034	,964
Anti-Image-Korrelation	Ich bin ein glücklicher Mensch.	,525	-,384	-,157
	Häufig zweifel ich an mir selbst	-,384	,530	-,038
	Ich besitze viele Stärken.	-,157	-,038	,638

a. Maß der Stichprobeneignung

Die Anti-Image-Kovarianz-Matrix ist nicht weiter interessant für die Frage der Eignung der Variablen für die Faktorenanalyse!

MSA-Indizes

Anti-Images

☞ Die Berechnung des MSA-Index liefert Anhaltspunkte, falls es erforderlich wird, Variablen zu entfernen.

☞ Die Anti-Image-Matrix gehört nicht in Euren Report! Falls Ihr die MSA-Indizes erwähnen wollt, ist es sinnvoller, diese in einer eigenen Tabelle darzustellen!

## Kommunalitäten

Anteil der Gesamtvarianz einer Variablen, der durch die Faktoren erklärt wird ( $h^2$ )

Kommunalitäten		
	Anfänglich	Extraktion
Ich bin ein glücklicher Mensch.	1,000	,730
Häufig zweifel ich an mir selbst	1,000	,620
Ich besitze viele Stärken.	1,000	,399
Ich vergleiche mich oft mit anderen Personen.	1,000	,619
Oft rede ich schlecht über mich selbst.	1,000	,600
Andauernd mache ich mir Gedanken, was an mir besser sein könnte.	1,000	,746

Extraktionsmethode: Hauptkomponentenanalyse.

Anteil der Varianz, der durch **alle** Faktoren (nicht nur die extrahierten) erklärt wird

Anteil der Varianz, der durch **alle** Faktoren, die nach der Extraktion übrig blieben, erklärt wird (war z.B. das Extraktionskriterium ein Eigenwert über 1, so gibt der Wert die Varianzaufklärung durch alle Faktoren mit einem Eigenwert größer/gleich 1 an)

☞ je höher die Werte (Wertebereich 0-1), umso besser die Varianzaufklärung und umso stärker die Faktoren!



## Eigenwerte, Gesamtvarianzaufklärung

Diese Tabelle liefert die Eigenwerte der Faktoren, woraus sich wiederum die Varianzaufklärung berechnen lässt. Ein Eigenwert ist der Anteil der Gesamtvarianz, der durch einen Faktor erklärt wird. Da mit z-standardisierten Variablen gerechnet wird, ist die Varianz jeder einzelnen Variable 1.

- ⇒ Insgesamt sind im Beispiel sechs Variablen verwendet worden. Diese haben somit eine Gesamtvarianz von 6. Der Eigenwert gibt den Varianzanteil an, den der jeweilige Faktor aufklärt. Für die erste Komponente liegt er bei 2.631. Daraus lässt sich der relative erklärte Varianzanteil errechnen:  $2.631/6 = .4385$ . Dies entspricht 43.85% der Gesamtvarianz. Dieser Wert ist ebenfalls in der Tabelle verzeichnet.
- ⇒ Eigenwerte lassen sich als Summe der quadrierten Ladungen aller Variablen auf einem Faktor berechnen. Faktor 1:  $.81^2 + .776^2 + \dots + .297^2 = 2.631$  (Faktorladungen siehe Komponentenmatrix!)

Komponente	Erklärte Gesamtvarianz								
	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,631	43,855	43,855	2,631	43,855	43,855	2,411	40,190	40,190
2	1,082	18,039	61,894	1,082	18,039	61,894	1,302	21,704	61,894
3	,939	15,645	77,539						
4	,570	9,507	87,046						
5	,433	7,214	94,260						
6	,344	5,740	100,000						

Extraktionsmethode: Hauptkomponentenanalyse

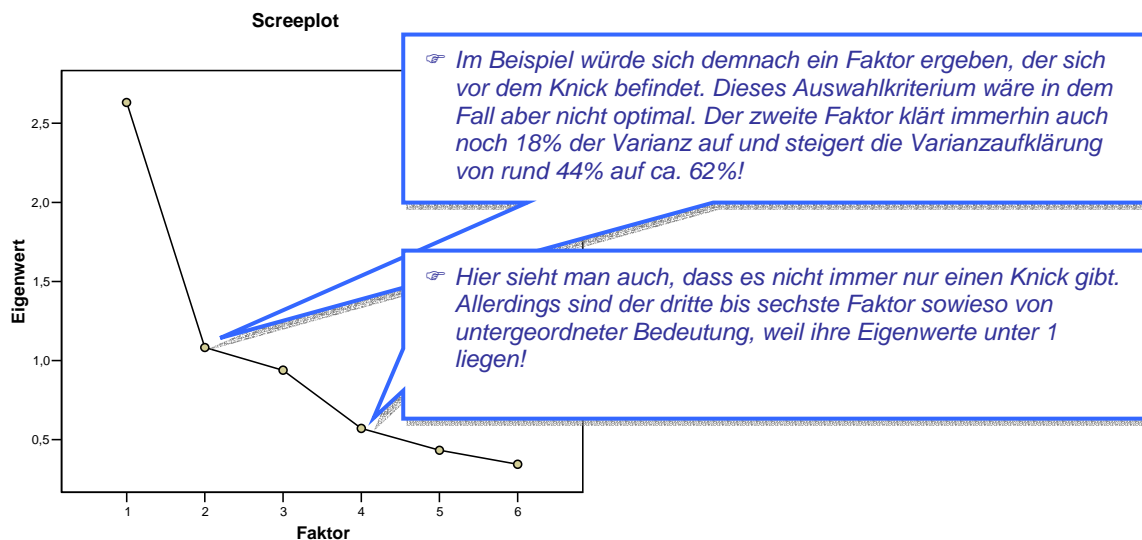
Eigenwerte und Varianzaufklärung aller Faktoren ohne Extraktion

Eigenwerte und Varianzaufklärung aller Faktoren mit Eigenwerten über 1

Eigenwerte und Varianzaufklärung aller Faktoren mit Eigenwerten über 1 nach Rotation: wie man sieht, hat sich die Gesamtvarianzaufklärung (61.89%) nicht geändert, aber die Eigenwerte und somit die Varianzaufklärungen für die einzelnen Faktoren. Dies ist auf die varianzmaximierende Rotation zurückzuführen.

## Screeplot

Der Screeplot stellt eine Unterstützung bei der Auswahl der Faktoren dar. Dabei wählt man die Variablen aus, die den Hang darstellen, das „Geröll“ hingegen, also Variablen mit eher geringen Eigenwerten, lässt man weg. Man orientiert sich dabei am „Knick“.



## Komponentenmatrix

Die Komponentenmatrix stellt die Ladungen der Variablen mit den Faktoren (Komponenten) dar. Je höher eine Variable auf einem Faktor lädt, umso stärker repräsentiert er sie. Die Variablen sind absteigend nach Größe ihrer Faktorladungen sortiert!

Komponentenmatrix <sup>a</sup>			Rotierte Komponentenmatrix <sup>a</sup>			Rotierte Komponentenmatrix <sup>a</sup>		
	Komponente			Komponente			Komponente	
	1	2		1	2		1	2
Andauernd mache ich mir Gedanken, was an mir besser sein könnte.	<b>.810</b>	<b>-.300</b>	Andauernd mache ich mir Gedanken, was an mir besser sein könnte.	<b>.863</b>	<b>.027</b>	Andauernd mache ich mir Gedanken, was an mir besser sein könnte.	.863	
Häufig zweifel ich an mir selbst	.776	.135	Ich vergleiche mich oft mit anderen Personen.	.784	.064	Ich vergleiche mich oft mit anderen Personen.	.784	
Ich vergleiche mich oft mit anderen Personen.	.751	-.236	Oft rede ich schlecht über mich selbst.	.770	.089	Oft rede ich schlecht über mich selbst.	.770	
Oft rede ich schlecht über mich selbst.	.746	-.208	Häufig zweifel ich an mir selbst	.668	.417	Häufig zweifel ich an mir selbst	.668	
Ich bin ein glücklicher Mensch.	.407	.751	Ich bin ein glücklicher Mensch.	.094	.849	Ich bin ein glücklicher Mensch.		.849
Ich besitze viele Stärken.	.297	.558	Ich besitze viele Stärken.	.065	.628	Ich besitze viele Stärken.		.628
Extraktionsmethode: Hauptkomponentenanalyse. a. 2 Komponenten extrahiert			Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. a. Die Rotation ist in 3 Iterationen konvergiert.			Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. a. Die Rotation ist in 3 Iterationen konvergiert.		

Dies sind die Faktorladungen ohne Rotation!

Faktorladungen nach Varimax-Rotation

Faktorladungen nach Varimax-Rotation, wobei Faktorladungen unter .5 unterdrückt wurden, es ist auf einen Blick ersichtlich, welche Variablen auf welchen Faktoren hoch laden!

⇒ Beispiel: Die erste Variable lädt ursprünglich mit .81 auf dem ersten Faktor. Nach Rotation ist diese Korrelation auf .86 gestiegen. Auf dem zweiten Faktor hat sie ursprünglich eine Ladung von -.3 (korreliert also negativ mit ihm). Nach Rotation liegt diese Faktorladung beinahe bei Null.

☞ Dies verdeutlicht das Prinzip der Varianzmaximierung durch Rotation: Hohe Ladungen werden zusätzlich verstärkt, niedrige Ladungen weiter abgeschwächt. Somit wird die Varianz der Variablen auf den Faktoren maximiert. Nach der Rotation gibt es kaum noch mittlere Ladungen! Somit kann zwischen den Variablen besser differenziert werden.

## Komponenten-Transformationsmatrix

Diese Matrix macht Aussagen über die Stärke der Rotation:

Komponententransformationsmatrix		
Komponente	1	2
1	.926	.377
2	-.377	.926

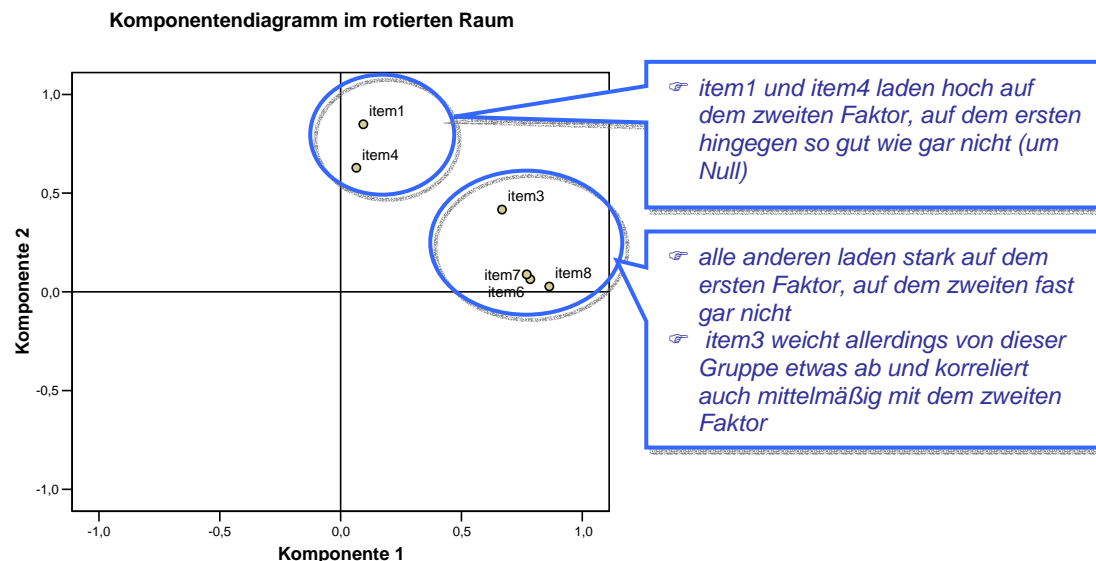
Extraktionsmethode: Hauptkomponentenanalyse.  
Rotationsmethode: Varimax mit Kaiser-Normalisierung.

☞ Je höher die Werte über bzw. unter der Diagonalen, umso stärker war die Drehung (>0.5: starke Rotation, nahe Null: schwache Rotation). Über die Cosinus-Funktion könnte der Rotationswinkel berechnet werden.



## Rotiertes Komponentendiagramm

In diesem Diagramm sind die Ladungen der Variablen auf den Faktoren graphisch dargestellt.



☞ Achtung: Die Kennzeichnung der Items stimmt nicht mit der Reihenfolge in den Komponentenmatrizen überein, weil sie dort nach Größe der Ladungen sortiert sind!

## Interpretation der Faktoren

Dies ist der spannendste, aber auch kniffligste Teil der Faktorenanalyse: Die Faktoren müssen interpretiert und benannt werden. Eine gute Orientierung erhält man, wenn man sich an den ein oder zwei Variablen orientiert, die am höchsten auf dem jeweiligen Faktor laden, denn sie steuern schließlich auch am meisten bei.

Auf Faktor 1 laden die folgenden Items:

- ⇒ Andauernd mache ich mir Gedanken, was an mir besser sein könnte.
- ⇒ Ich vergleiche mich oft mit anderen Personen.
- ⇒ Oft rede ich schlecht über mich selbst.
- ⇒ Häufig zweifle ich an mir selbst.

Dieser Faktor wäre möglicherweise als „Selbstbewertung“ o. ä. interpretierbar

Faktor 2:

- ⇒ Ich bin ein glücklicher Mensch.
- ⇒ Ich besitze viele Stärken.

Eventuell „Selbstbeschreibung“?

Hier wird deutlich, dass die Ergebnisse der Faktorenanalyse stark von der Interpretation des Forschers abhängen. Für dieses kleine Beispiel allein wären schon etliche Varianten denkbar. Hinzu kommt, dass die Ergebnisse immer davon abhängen, welche Variablen überhaupt eingehen. Hätte man weitere Variablen hinzugenommen, wären vermutlich völlig andere Faktoren herausgekommen.

Das ist sowohl eine Chance, Systematiken und Modelle zu entdecken, gleichzeitig aber auch ein Risiko, weil man die Variablen solange hin- und herschieben kann, bis die gewünschten Ergebnisse zustande kommen. Deshalb sollte bei der Durchführung einer Faktorenanalyse kritisch mit den Daten umgegangen werden.

## Hinweise zum Schreiben eines statistischen Reports

- **Fragestellung konkret formulieren**
- **Deskriptive Statistik**
- **Beschreibung der verwendeten Methoden**
- **Ergebnisdarstellung der durchgeführten Tests**
  - Prüfen der Voraussetzungen für Korrelationsberechnung:
    - siehe Handout Korrelation und Regression, wichtig insbesondere:
      - die Intervallskalierung der Daten,
      - Normalverteilung und
      - keine Ausreißer
  - Prüfung der Daten auf ihre Eignung für die Faktorenanalyse (vgl. S. 4/5)
  - Korrelationsmatrix (nur die Hälfte unter der Diagonalen)
  - Faktoren: Anzahl, Eigenwerte, Varianzaufklärung
  - Komponentenmatrix: nichtrotierte, rotierte Lösung (dabei können niedrige Ladungen unterdrückt werden, wobei jedoch die Grenze kenntlich gemacht werden sollte!)
  - Interpretation der Faktoren