

# **Explorative Faktorenanalyse**

**Bachelorseminar:**

**Ausgewählte Aspekte der Wirtschafts- und**

**Sozialstatistik**

Autor: Simon Reitzner

Betreuerin: Eva Endres, M.Sc.

23. September 2015

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theorie</b>	<b>2</b>
2.1	Grundidee der Faktorenanalyse . . . . .	2
2.2	Das Modell . . . . .	3
2.3	Fundamentaltheorem der Faktorenanalyse . . . . .	4
2.4	Schritte der Faktorenanalyse . . . . .	5
2.5	Berechnung der Korrelationsmatrix und Standardisierung . . .	6
2.6	Schätzmethoden des Faktorenmodells . . . . .	6
2.7	Bestimmen der Faktorenanzahl . . . . .	9
2.8	Rotation der Faktoren . . . . .	12
2.9	Interpretation der Faktoren . . . . .	17
2.10	Faktorenwerte . . . . .	17
<b>3</b>	<b>Anwendungsbeispiel</b>	<b>18</b>
3.1	Erzeugen der Daten . . . . .	18
3.2	Berechnung der Korrelationsmatrix . . . . .	20
3.3	Bestimmung der Faktorenanzahl . . . . .	21
3.4	Schätzung der Kommunalitäten und Berechnung der Ladungs- matrix . . . . .	24
3.5	Faktorenanalyse zweiter Ordnung . . . . .	27
<b>4</b>	<b>Einbettung der Arbeit in den Seminarkontext</b>	<b>28</b>
	<b>Literaturverzeichnis</b>	<b>29</b>
<b>A</b>	<b>Anhang</b>	<b>31</b>

# 1 Einleitung

In der Statistik wird unterschieden zwischen manifesten Variablen, die direkt beobachtbar und messbar sind, und latenten Variablen, die nicht unmittelbar messbar sind. Das Erfassen dieser latenten Variablen stellt, aufgrund der fehlenden Messbarkeit, immer wieder ein Problem dar. Um dennoch mit latenten Variablen arbeiten zu können, werden aus messbaren Indikatoren, latente Konstrukte gebildet (Bühner; 2011, S.296).

Mithilfe der explorativen Faktorenanalyse, einem multivariaten Verfahren, kann von Indikatorvariablen auf latente Hintergrundvariablen geschlossen werden (Brosius; 2011, S.787). Ein weiteres Ziel der Faktorenanalyse ist, neben dem Entdecken von latenten Strukturen, die Dimensionsreduktion. Besteht ein hoher linearer Zusammenhang zwischen den Variablen, lassen sie diese mit einem möglichst geringem Informationsverlust zu homogenen latenten Gruppen zusammenfassen (Bühner; 2011, S.296). Dadurch kann die Komplexität der Daten verringert werden, weil der Datensatz auf eine kleinere Anzahl an Variablen reduziert wird (Brosius; 2011, S.787).

Im Folgenden werden die Gruppen bzw. die latenten Variablen „Faktoren“ genannt. Neben der explorativen, der strukturentdeckenden, gibt es die konfirmatorische Faktorenanalyse, welche bereits bestehende Konstrukte überprüft.

Die vorliegende Arbeit soll sich auf die explorative Faktorenanalyse beschränken und die gängigsten Methoden erklären bzw. veranschaulichen. Dabei werden Methoden zur Überprüfung, ob sich das Datenmaterial für eine Faktorenanalyse eignet, außer Acht gelassen und es wird angenommen, dass die Korrelation zwischen den Variablen geeignet hoch ist.

Nach der Erläuterung der einzelnen Schritte der explorativen Faktorenanalyse in Kapitel 2 wird anschließend in Kapitel 3 eine Faktorenanalyse an einem Datensatz mit der Statistiksoftware R durchgeführt.

## 2 Theorie

### 2.1 Grundidee der Faktorenanalyse

Das Modell der Faktorenanalyse nimmt an, dass der Wert einer Beobachtung  $i$  für eine Variable  $j$ , z.B. eine Antwort einer Person auf eine Frage in einem Test, sich durch eine Linearkombination aus gewichteten, latenten Faktoren und einem zufälligen Fehler darstellen lässt (Bühner; 2011, S.299):

$$z_{ij} = l_{j1}f_{i1} + \dots + l_{jl}f_{il} + \dots + l_{jk}f_{ik} + e_{ij}$$

$z_{ij}$  standardisierter Wert einer Beobachtung/Person  $i$  auf eine Variable  $j$

$l_{jl}$  Ladung der Variable  $j$  auf Faktor  $l$

$f_{il}$  Faktorwert von Beobachtung/Person  $i$  auf Faktor  $l$

$e_{ij}$  Fehlerkomponente/spezifische Faktoren

Die Grundlage der Faktorenanalyse ist ein lineares Modell mit den unbekannt-ten Größen  $l_{jl}$ ,  $f_{il}$  und  $e_{ij}$ . Ebenfalls unbekannt ist die Faktorenanzahl  $k$ . Mittels der Faktorenextraktion werden diese Größen geschätzt und anschließend zwecks Interpretierbarkeit der Faktoren diesen die Variablen zugeordnet. Folgende Grafik (in Anlehnung an (Bühl; 2008, S.527)) veranschaulicht die Zuordnung der Variablen zu den Faktoren:

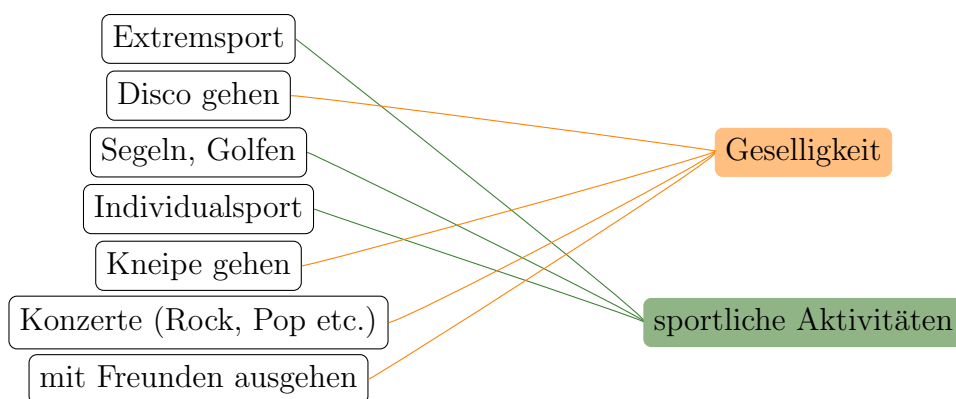


Abbildung 1: Zuordnung verschiedener Freizeitaktivitäten zu Faktoren

## 2.2 Das Modell

Das Modell der Faktorenanalyse basiert auf einem linearen Modell:

$$\mathbf{z} = \mathbf{L}\mathbf{f} + \mathbf{e}$$

$\mathbf{z}$  stellt einen  $p$ -dimensionalen Zufallsvektor dar.  $\mathbf{f}$  entspricht dem Vektor der gemeinsamen Faktoren. Da eine Reduktion der Dimension des Datensatz stattfindet, ist die Anzahl  $k$  an Faktoren kleiner als die Anzahl  $p$  an Variablen.  $\mathbf{e}$  steht für den Vektor der Restterme. Die Restterme sind für jede Variable unterschiedlich und werden deshalb auch spezifische Faktoren genannt. Diese entsprechen dem Anteil, der durch die Faktoren nicht erklärt werden kann.  $\mathbf{L}$  ist eine  $(p \times k)$ -Ladungsmatrix (Fahrmeir et al.; 1996, S.641ff.).

Über die spezifischen Faktoren werden folgende Annahmen zum Erwartungswert und zur Varianz getroffen:

$$\mathbb{E}(\mathbf{e}) = \mathbf{0} , \quad Cov(\mathbf{e}) = \mathbf{V}$$

wobei davon ausgegangen wird, dass  $\mathbf{e}$  nicht untereinander korreliert.

Folgendes gilt für den Vektor  $\mathbf{f}$ :

$$\mathbb{E}(\mathbf{f}) = \mathbf{0} , \quad Cov(\mathbf{f}) = \mathbf{\Phi}$$

Ist  $Cov(\mathbf{f}) = \mathbf{I}$ , liegt ein orthogonales Modell vor, d.h. die Faktoren sind voneinander unabhängig. Geht man von einer Abhängigkeit aus, spricht man von obliquen Faktoren (Fahrmeir et al.; 1996, S.642).

## 2.3 Fundamentaltheorem der Faktorenanalyse

Grundlage der Faktorenanalyse ist die Zerlegung der Varianz von  $\mathbf{z}$ . Diese lässt sich folgendermaßen formulieren (nach Fahrmeir et al.; 1996, S.643):

$$\begin{aligned}\Sigma &= Cov(\mathbf{z}) \\ &= \mathbb{E}(\mathbf{z}\mathbf{z}^T) \\ &= \mathbb{E}((\mathbf{L}\mathbf{f} + \mathbf{e})(\mathbf{L}\mathbf{f} + \mathbf{e})^T) \\ &= \mathbb{E}(\mathbf{L}\mathbf{f}\mathbf{f}^T\mathbf{L}^T) + \mathbb{E}(\mathbf{L}\mathbf{f}\mathbf{e}^T) + \mathbb{E}(\mathbf{e}\mathbf{f}^T\mathbf{L}^T) + \mathbb{E}(\mathbf{e}\mathbf{e}^T) \\ &= \mathbf{L} Cov(\mathbf{f}) \mathbf{L}^T + \mathbf{V}\end{aligned}$$

Im Fall eines orthogonalen Modells mit  $Cov(\mathbf{f}) = \mathbf{I}$ :

$$\begin{aligned}\Sigma &= \mathbf{L} Cov(\mathbf{f}) \mathbf{L}^T + \mathbf{V} \\ &= \mathbf{L}\mathbf{L}^T + \mathbf{V}\end{aligned}$$

Für die Kovarianz zwischen  $\mathbf{z}$  und  $\mathbf{f}$  gilt folgendes:

$$\begin{aligned}Cov(\mathbf{z}, \mathbf{f}) &= Cov(\mathbf{L}\mathbf{f} + \mathbf{e}, \mathbf{f}) \\ &= Cov(\mathbf{L}\mathbf{f}, \mathbf{f}) + Cov(\mathbf{e}, \mathbf{f}) \\ &= \mathbf{L} Cov(\mathbf{f})\end{aligned}$$

Im Fall eines orthogonalen Modells mit  $Cov(\mathbf{f}) = \mathbf{I}$ :

$$\begin{aligned}Cov(\mathbf{z}, \mathbf{f}) &= Cov(\mathbf{L}\mathbf{f} + \mathbf{e}, \mathbf{f}) \\ &= Cov(\mathbf{L}\mathbf{f}, \mathbf{f}) + Cov(\mathbf{e}, \mathbf{f}) \\ &= \mathbf{L}\end{aligned}$$

Im orthogonalen Modell entspricht die Ladung einer Variable auf einen Faktor der normierten Kovarianz zwischen dieser Variable und diesem Faktor. Dieser Zusammenhang ist wichtig für spätere Rotationsmethoden (siehe Kapitel 2.8) und Interpretationen. Die Summe der quadrierten Ladungen für einen Faktor entspricht seinem Eigenwert und dem Anteil an Varianz der Daten, die dieser

Faktor erklärt. Ein weiterer wichtiger Begriff ist die „Kommunalität“. Eine Kommunalität ist der Varianzanteil einer Variablen, der durch das gesamte Faktorenmodell erklärt wird.

## 2.4 Schritte der Faktorenanalyse

Die Faktorenanalyse kann man in folgende Schritte unterteilen:

1. Überprüfung der Bedingungen für eine Faktorenanalyse:

Notwendig für eine Durchführung der Faktorenanalyse ist ein ausreichend hoher Zusammenhang bzw. Korrelation zwischen den Variablen. Verfahren, um die Eignung der Daten für eine Faktorenanalyse zu untersuchen, werden in dieser Arbeit nicht behandelt.

2. Berechnung der Korrelationsmatrix der Daten und Standardisierung:

Ausgangspunkt dieses Modells ist die Korrelationsmatrix bzw. die standardisierte Varianz-Kovarianzmatrix, die aus dem Datenmaterial berechnet wird.

3. Schätzmethode des Faktorenmodells:

Je nachdem von welchen Annahmen ausgegangen wird, bieten sich verschiedene Verfahren an, die das Faktorenmodell aufstellen.

4. Bestimmung der Faktorenzahl:

Da die Faktorenzahl unbekannt ist, muss sie aus den Daten bestimmt werden. Bei den meisten Methoden werden die Eigenwerte der Korrelationsmatrix untersucht, um so wichtige Faktoren zu erkennen. In Kapitel 2.7 wird dieses Thema ausführlicher dargestellt.

5. Faktorenrotation:

Nach der Faktorenextraktion, also dem Schätzen der Größe des Faktorraum und dem Berechnen aller unbekannt Größen, werden die Faktoren für eine einfachere Interpretation und Zuordnung rotiert (siehe dazu Kapitel 2.8).

## 6. Interpretation der Faktoren:

Bei der Interpretation wird überlegt, für was die Faktoren inhaltlich stehen und welchen Sachverhalt sie beschreiben.

## 7. Berechnung der Faktorenwerte:

Oft ist es notwendig, Ausprägungen der Individuen auf die Faktoren vorliegen zu haben. So lassen sich weitere statistische Analysen anschließen.

## 2.5 Berechnung der Korrelationsmatrix und Standardisierung

Da die Varianz der Daten meistens unbekannt ist, wird diese durch die empirische Kovarianzmatrix bzw. der Korrelationsmatrix geschätzt. Für die Schätzung der Faktorenanalyse nach der Hauptkomponentenanalyse ist eine Standardisierung der Daten notwendig, da diese nicht invariant gegenüber Skalierungen ist (Fahrmeir et al.; 1996, S.662). Nach einer z-Standardisierung entspricht die Varianz-Kovarianzmatrix der Daten der Korrelationsmatrix. Dies erleichtert die Berechnung und die Interpretation der Ergebnisse.

## 2.6 Schätzmethoden des Faktorenmodells

Folgende Methode gehören zu den wichtigsten Verfahren, um das Faktorenmodell zu schätzen:

- Maximum-Likelihood-Methode
- Hauptkomponentenmethode
- Hauptfaktorenanalyse bzw. Hauptachsen-Faktorenanalyse

Mit diesen wird die unbekannt Ladungsmatrix  $\mathbf{L}$  bestimmt. Die drei genannten Methoden werden in den folgenden Unterkapitel kurz beschrieben.



### 2.6.1 Faktorenanalyse mit der Maximum-Likelihood-Methode

Grundlage für dieses Schätzverfahren ist das Maximum-Likelihood-Prinzip. Die unbekannte und zu schätzende Ladungsmatrix  $\mathbf{L}$  und die Varianz der spezifischen Faktoren  $\mathbf{V}$  werden mit der ML-Methode geschätzt. Zwingend notwendig ist dabei die Voraussetzung, dass die Daten multivariat-normalverteilt sind. Nach erster Schätzung wird das berechnete Modell mit einem Likelihood-Quotienten-Test auf Signifikanz geprüft. Bei fehlender Signifikanz lässt sich die Anzahl der Faktoren  $k$  ändern und das Modell wieder einem Test unterziehen.

### 2.6.2 Hauptkomponentenmethode

Das vollständige Modell der Hauptkomponentenmethode sieht folgendermaßen aus (vgl. Fahrmeir et al.; 1996, S.667f):

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^T$$

und der Zerlegung der Korrelationsmatrix:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T$$

Bei diesem Modell geht man davon aus, dass die Faktoren die gesamte Varianz erklären können. Es gibt keine spezifischen Faktoren. Die Ladungsmatrix wird aus der Spektralzerlegung erhalten. Hierbei wird die symmetrische Matrix  $\mathbf{R}$  in ihre Eigenvektoren und Eigenwerte zerlegt:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T = \mathbf{P} \mathbf{diag}(\lambda_1, \dots, \lambda_p) \mathbf{P}^T$$

Es liegen nun  $p$  Eigenvektoren und Eigenwerte vor. Für die Ladungsmatrix  $\mathbf{L}$  gilt:

$$\mathbf{L} = \mathbf{P} \mathbf{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$$

$\mathbf{P}$  ist die Matrix der Eigenvektoren von  $\mathbf{R}$  und  $\mathbf{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$  eine Diagonalmatrix mit den Eigenwerten auf der Diagonalen. Da eine Dimen-

sionsreduktion stattfinden soll, wird für der Berechnung von  $\mathbf{L}$  nur die  $k$  wichtigsten Eigenwerte und Eigenvektoren verwendet.

### 2.6.3 Hauptfaktorenanalyse

Anders als bei der Hauptkomponentenmethode geht die Hauptfaktorenanalyse (oder auch Hauptachsenanalyse) davon aus, dass es spezifische Faktoren gibt. Diese Methode entspricht dementsprechend der ursprünglichen Idee der Faktorenanalyse (vgl. Fahrmeir et al.; 1996, S.671f):

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^T + \mathbf{E}$$

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T + \mathbf{V} \iff \mathbf{R} - \mathbf{V} = \mathbf{L}\mathbf{L}^T \iff \mathbf{R}_h = \mathbf{L}\mathbf{L}^T$$

$\mathbf{R}_h$  wird reduzierte Korrelationsmatrix genannt. Sie unterscheidet sich von der Korrelationsmatrix insofern, dass auf ihrer Diagonale die Kommunalitäten  $h_i^2 = 1 - v_i^2$  stehen. Diese Kommunalitäten werden im Rahmen der Hauptfaktorenanalyse iterativ geschätzt. Bei der Hauptkomponentenmethode stellt sich diese Kommunalitäten-Problem nicht, da es keine Einzelrestvarianzen gibt und folglich alle Kommunalitäten gleich 1 sind. Als mögliche Startkommunalität bietet sich der quadrierte multiple Korrelationskoeffizient zwischen der Variable  $z_j$  und den übrigen Variablen oder der betragsmäßig größte Korrelationskoeffizient zweier Variablen an (Everitt und Hothorn; 2011, S.141). Die Spektralzerlegung von  $\mathbf{R}_h$  in ihre Eigenvektoren und Eigenwerte liefert:

$$\mathbf{R}_h = \mathbf{L}\mathbf{L}^T = \mathbf{P} \mathbf{diag}(\lambda_1, \dots, \lambda_p) \mathbf{P}^T$$

Analog wie bei der Hauptkomponentenmethode gilt für die Ladungsmatrix:

$$\mathbf{L} = \mathbf{P} \mathbf{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$$

wobei  $\mathbf{P}$  diesmal die Matrix der Eigenvektoren von  $\mathbf{R}_h$  und  $\mathbf{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p})$  eine Diagonalmatrix aus den Eigenwerten ist.

Wie auch bei der Hauptkomponentenmethode wird nun auf die  $k$  wichtigsten

Faktoren reduziert, indem nur die dazugehörigen Eigenwerte und Eigenvektoren in die Berechnung der Ladungsmatrix aufgenommen werden. Die von den Faktoren, die nicht in das Modell aufgenommen werden, erklärte Varianz wird zu  $\mathbf{V}$  hinzugefügt. Anders als bei der Hauptkomponentenmethode setzt nach diesem Punkt die iterative Schätzung der Kommunalitäten an. Die Kommunalitäten für die jeweils nächste Iteration berechnen sich nach jeder Spektralzerlegung durch  $\text{diag}(\mathbf{L}_k \mathbf{L}_k)$ , mit  $\mathbf{L}_k$  der auf  $k$  Faktoren reduzierten Ladungsmatrix. Diese werden wieder auf die Diagonale von  $\mathbf{R}$  gesetzt und die Korrelationsmatrix wird anschließend abermals zerlegt. Die beschriebenen Schritte werden wiederholt bis Konvergenz vorliegt (vgl. Fahrmeir et al.; 1996, S.672f).

## 2.7 Bestimmen der Faktorenanzahl

Da bei den zuvor beschriebenen Schätzmethode nur die  $k$  wichtigsten Eigenwerte und Eigenvektoren relevant sind, muss zunächst die unbekannte Anzahl  $k$  an Faktoren ermittelt werden. Es soll  $k < p$  gelten, wobei  $k$  möglichst klein sein soll (möglichst wenige Faktoren, sollen möglichst viel Varianz erklären). Es sind die Faktoren zu extrahieren, die besonders viel Varianz von  $\mathbf{z}$  erklären. Das Problem bei der Bestimmung der Faktorenanzahl ist, dass es kein klares Kriterium gibt, wann ein Faktor viel Varianz erklärt. Diese Entscheidung sollte immer mit Betrachtung des vorliegenden Sachverhalt und mit mehr als nur einer Methode zur Anzahlbestimmung getroffen werden.

Es kann vorteilhaft sein, wenn bereits Vorwissen über die Dimension des Faktorraum vorhanden ist. Andernfalls lassen sich mit folgenden Methoden die Faktorenanzahl bestimmen:

- Kaiser-Guttman-Kriterium

Es wird davon ausgegangen, dass ein Faktor, der mehr als die Varianz einer standardisierten Variable erklärt, auffällig ist und extrahiert werden sollte. Zunächst werden die Eigenwerte von  $\mathbf{R}$  berechnet. Ist ein Eigenwert größer 1, wird dieser extrahiert. Die Faktorenanzahl entspricht der Anzahl an Eigenwerten größer 1 (Bortz und Schuster; 2010, S.415).

- Scree-Test

Bei diesem Verfahren nach Cattell trägt man die Eigenwerte in einem Plot abfallend nach ihrer Größe an (siehe Abbildung 2). Der Verlauf dieser fallenden Kurve enthält oft einen Knick oder Sprung zwischen zwei Eigenwerten. Die Anzahl an bedeutenden Faktoren entspricht der Anzahl an Eigenwerten vor diesem Sprung (Fahrmeir et al.; 1996, S.669).

- Parallelanalyse

Die Parallelanalyse nach Horn (1965) ähnelt der Analyse mit dem Scree-Test. Die empirischen Eigenwerte werden mit den Eigenwerten von simulierten Zufallsvariablen verglichen. Dazu werden die Eigenwerte von mehreren Bootstrap-Stichproben oder von normalverteilten Zufallsstichproben in den Scree-Plot gezeichnet (siehe Abbildung 2). Alle empirischen Eigenwerte, die größer als die simulierten Eigenwerte sind, gelten als auffällig und werden extrahiert.

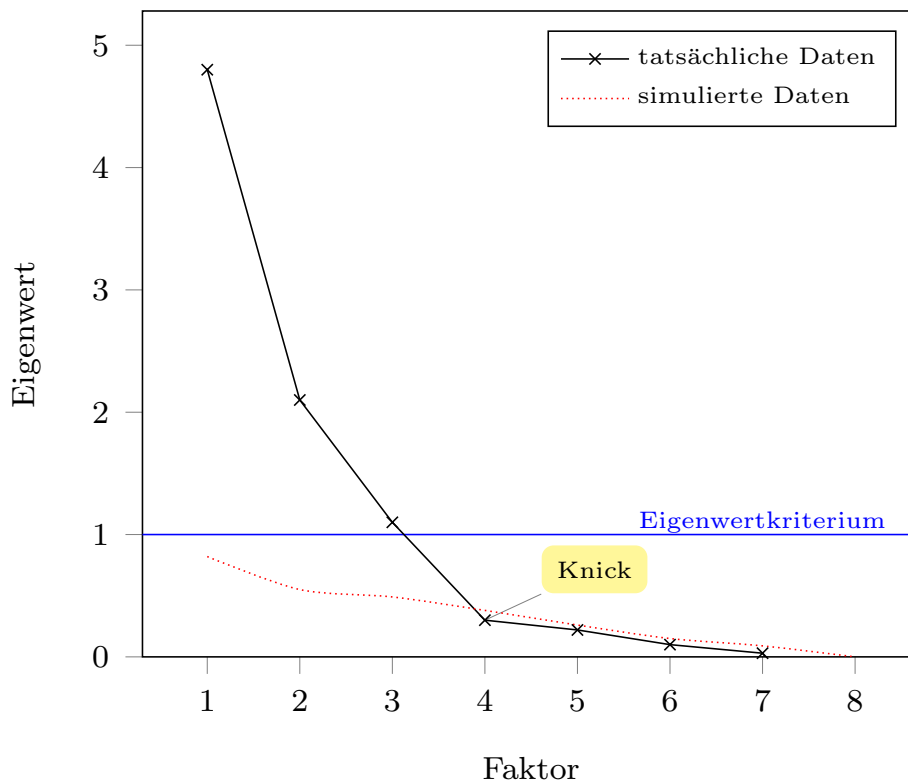


Abbildung 2: Scree-Plot mit verschiedenen Selektionskriterien

- Minimum-Average-Partial-Test (MAP-Test)

Für diesen Test nach Velicer wird zunächst eine Hauptkomponentenanalyse durchgeführt und der Faktor mit dem höchsten Eigenwert entfernt. Die Korrelation zwischen den Variablen wird erneut berechnet, allerdings ohne den Effekt des entfernten Faktors. Es liegt nun eine Matrix mit den partiellen Korrelationen der Variablen für den ersten Faktor vor. Für diese Matrix wird das arithmetische Mittel

$$MAP_l = \sum_{j=1}^p \sum_{\substack{j'=1 \\ j \neq j'}}^p \frac{r_{jj'}^{*2}}{p(p-1)}$$

berechnet.  $p$  steht für die Anzahl an möglichen Faktoren, die der Variablenanzahl entspricht.  $l$  ist der Index für den Faktor, dessen mittlere quadrierte Partialkorrelation berechnet wird. Die mittlere quadrierte Partialkorrelation ist das arithmetische Mittel an Restvarianz, welche die herauspartialisierten Faktoren nicht erklären. Ist diese sehr klein, erklärt das Modell viel der Korrelation der Daten. Dieses Verfahren wird mit allen Variablen durchgeführt, bis nur noch der Effekt eines Faktors in der Korrelationsmatrix enthalten ist. Für jeden theoretisch möglichen Faktor liegt die mittlere quadrierte Partialkorrelation vor. Diese werden untereinander verglichen. Angenommen der  $k$ -te Faktor hat die kleinste mittlere quadrierte Partialkorrelation, so werden  $k$  Faktoren extrahiert (vgl. Garrido E. et al.; 2011, S.553ff.).

- Likelihood-Quotienten-Test

Die Voraussetzung für diese Methode ist die Normalverteilung der Daten. Dieses Verfahren wird häufig verwendet, wenn eine ML-Faktorenanalyse verwendet wird. Es wird die Hypothese

$$H_0(l): \text{„}l \text{ Faktoren können die Daten erklären“}$$

getestet. Wird  $H_0$  abgelehnt, wird die Anzahl  $l$  an Faktoren erhöht,  $H_0$  erneut getestet und gegebenenfalls wiederholt bis Signifikanz erreicht ist (vgl. Fahrmeir et al.; 1996, S.654ff.).

## 2.8 Rotation der Faktoren

Für eine eindeutige Zuordnung der Variablen zu den Faktoren (wie in Grafik 1) sind entsprechende Ladungen erforderlich, weil sie das Kriterium für die Zuordnung sind. Die Variablen, die auf einen Faktor hochladen bzw. hoch mit diesem korrelieren, werden diesem Faktor zugeordnet. Für eine leichtere Interpretation und Zuordnung werden Ladungen angestrebt, sodass jeweils die Variable auf einen Faktor hoch- und auf die anderen Faktoren niedrig lädt (eine Ladung ab 0.5 gilt üblicherweise als „hoch“ (Fahrmeir et al.; 1996, S.682)). Oft ist eine eindeutige Zuordnung nicht möglich, weil Variablen auf mehrere Faktoren hochladen, weshalb Rotationsmethoden notwendig sind.

### 2.8.1 Geometrische Betrachtung der Faktoren

Werden die Faktoren als Vektoren angesehen und die Ladungen als Punkte in einem  $k$ -dimensionalen Vektorraum, so lassen sich die Rotationen gut veranschaulichen. Die Ladungen auf die Faktoren stellen die Koordinaten für eine Variable dar.

Die folgende Grafiken (angelehnt an Fahrmeir et al.; 1996, S.678f.) zeigen beispielhaft 2 Faktoren und ihre Ladungen:

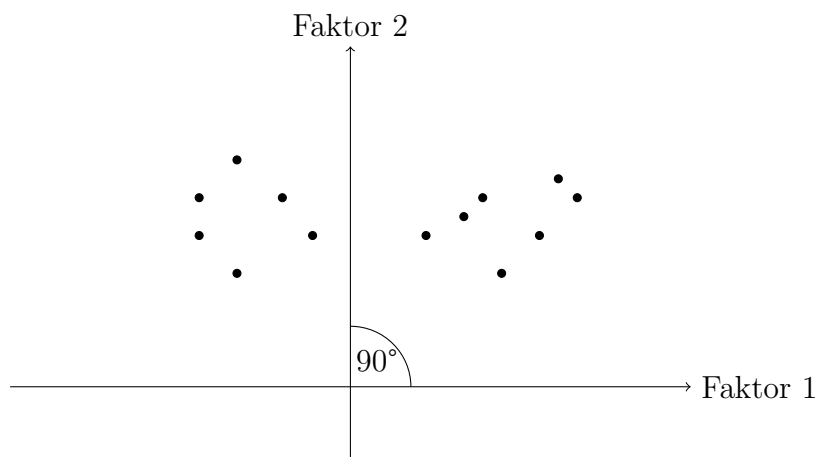


Abbildung 3: Faktoren mit Ladungen vor der Rotation

Vor der Rotation wird angenommen, dass die Faktoren unabhängig sind, d.h. sie stehen senkrecht aufeinander. Ziel aller Rotationen, sowohl für unabhängige, als auch für abhängige Faktoren ist es, die Faktoren im Faktorraum rotieren zu lassen, sodass sie möglichst gut zu den Ladungen passen.

Sind Faktoren unabhängig voneinander, so sind sie auch nach der Rotation noch orthogonal:

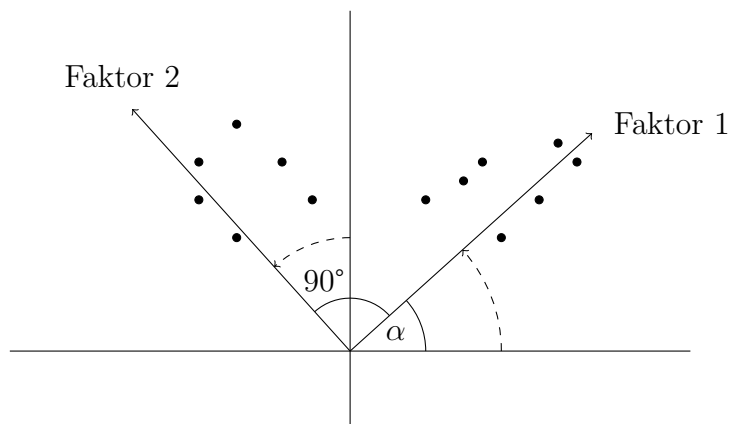


Abbildung 4: Faktoren nach orthogonaler Rotation

Bei abhängigen Faktoren weicht der Winkel zwischen den Faktoren von  $90^\circ$  ab:

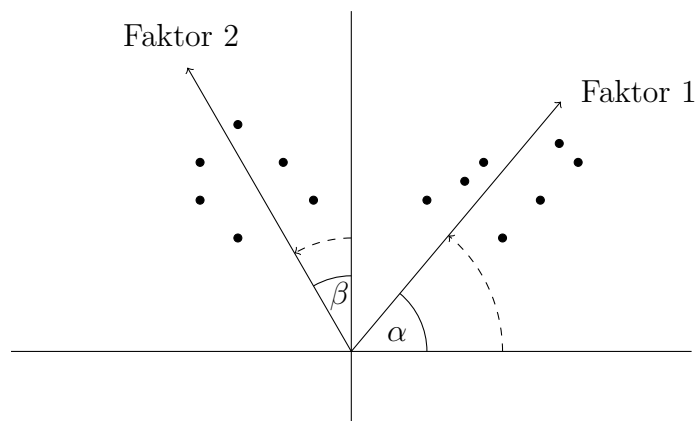


Abbildung 5: Faktoren nach obliquen Rotation

Nach der Extraktion der Faktoren, liegen orthogonale Faktoren vor, d.h.

die Faktoren stehen senkrecht zueinander. Während bei einer orthogonalen Rotation der 90° Winkel zwischen den Faktoren erhalten bleibt, lässt eine oblique Rotation die Faktoren in beliebigen Winkeln drehen.

Wird angenommen, dass die Faktoren miteinander korrelieren, findet erst mit einer obliquen Rotation eine Beachtung dieser Abhängigkeit statt.

### 2.8.2 Faktorentransformation

Bei einer Transformation darf sich nicht der Erklärungsanteil an der Varianz der Variablen ändern. Warum eine Faktorentransformation möglich ist, zeigt folgende Berechnung (vgl. Fahrmeir et al.; 1996, S.678):

$$\begin{aligned} \mathbf{z} &= \mathbf{L}\mathbf{f} + \mathbf{e} = \mathbf{L}\mathbf{M}\mathbf{M}^{-1}\mathbf{f} + \mathbf{e} \\ &= \mathbf{L}\mathbf{M} \mathbf{M}^{-1}\mathbf{f} + \mathbf{e} = \tilde{\mathbf{L}}\tilde{\mathbf{f}} + \mathbf{e} \end{aligned}$$

$\mathbf{M}$  sei eine Rotationsmatrix aus  $\mathbb{R}^{k,k}$  mit der Inversen  $\mathbf{M}^{-1}$ .

Es ist zu erkennen, dass die Ladungsmatrix und die Matrix der Faktorenwerte  $\mathbf{F}$  (siehe Kapitel 2.10) nicht eindeutig sind. Sowohl  $\mathbf{L}$  und  $\mathbf{F}$ , als auch  $\tilde{\mathbf{L}}$  und  $\tilde{\mathbf{F}}$  stellen Lösungen der Zerlegung von  $\mathbf{z}$  dar. Die verschiedenen Rotationsmethoden versuchen alle durch Varianzmaximierung eine Ladungsmatrix aus der Menge der existierenden Ladungsmatrizen zu finden, sodass die Faktoren möglichst leicht interpretiert und Variablen möglichst eindeutig zugeordnet werden können.

### 2.8.3 Iterativer Rotationsprozess

Bei vielen Rotationsmethoden wird die Rotationsmatrix über einen iterativen Prozess bestimmt. Folgende Schritte werden durchgeführt:

1. Aufstellen der Initial-Rotations- und Ladungsmatrix:

$$\text{Initial-Rotationsmatrix:} \quad \mathbf{M}_1 = \mathbf{I}_k \quad (\hat{=} \text{Einheitsmatrix})$$

$$\text{Initial-Ladungsmatrix:} \quad \mathbf{L}_0 = \mathbf{L}$$

2. Maximieren der Varianzen durch Anwendung des Rotationskriterium



auf  $\tilde{\mathbf{L}}_1 = \mathbf{L}_0 \mathbf{M}_1 = (\tilde{l}_{jl})$  mit  $\tilde{l}_{jl} = l_j^T m_l$  (Ladung nach Rotation):  
 Bsp.: Varimaxkriterium (siehe Varimaxkriterium in 2.8.4)

$$Q_{Varimax} = \sum_{l=1}^k \underbrace{\sum_{j=1}^p (\tilde{l}_{jl}^2 - \frac{\sum_{l=1}^p \tilde{l}_{jl}^2}{p})^2}_{p \cdot s_l^2} \rightarrow \max$$

mit der Bedingung, dass  $\mathbf{M}$  orthogonal ist. (Lagrange Methode!)  
 $\rightarrow$  durch Ableiten nach  $\mathbf{M}$  und Nullsetzen, liegt eine quadratische Matrix  $\mathbf{B}_1$  vor.

### 3. Singulärwertzerlegung von $\mathbf{B}$

$$\mathbf{B}_1 = \mathbf{U} \mathbf{diag}(\boldsymbol{\lambda}) \mathbf{V}^T = \underbrace{\mathbf{U} \mathbf{V}^T}_{\mathbf{M}_2} \underbrace{\mathbf{V} \mathbf{diag}(\boldsymbol{\lambda}) \mathbf{V}^T}_{\mathbf{A}_1} = \mathbf{M}_2 \mathbf{A}_1$$

### 4. $\mathbf{M}_2$ ist die Rotationsmatrix für die 2. Iteration

Die Schritte werden wiederholt bis Konvergenz eintritt. Ausführliche Darstellung dieser Iteration ist in Knüsel (2008) zu finden.

## 2.8.4 Orthogonale Rotationsmethoden

Bei einer orthogonalen Rotation ist die Unabhängigkeit der Faktoren nach der Rotation noch vorhanden. Das ist in Abbildung 4 zu erkennen, weil beide Faktoren um den gleichen Winkel  $\alpha$  rotiert werden. Zu den orthogonalen Verfahren zählen diese zwei Rotationen:

- Varimax

Ziel dieser Methode nach Kaiser (1958) ist es, eine hohe und ansonsten niedrige Ladungen für eine Variable vorliegen zu haben. Dies wird bei dieser Methode durch die Maximierung der Varianz der Ladungsqua-

drate für jeden Faktor erreicht. Folgender Ausdruck wird maximiert:

$$Q_{Varimax} = \sum_{l=1}^k \underbrace{\sum_{j=1}^p (l_{jl}^2 - \frac{\sum_{j=1}^p l_{jl}^2}{p})^2}_{p \cdot s_l^2} \rightarrow max$$

$s_l^2$  ist die Varianz der quadrierten Ladungen für den Faktor l

- Quartimax

Die Quartimax ist eine Rotation, ähnlich wie die Varimax-Methode. Allerdings werden die Varianzen der Variablen (Zeilen) maximiert, anstatt der Faktoren. Folgender Ausdruck hier soll maximiert werden (vgl. Fahrmeir et al.; 1996, S.680):

$$Q_{Quartimax} = \sum_{j=1}^p \underbrace{\sum_{l=1}^k (l_{jl}^2 - \frac{\sum_{l=1}^k l_{jl}^2}{k})^2}_{k \cdot s_j^2} \rightarrow max$$

$s_j^2$  ist die Varianz der quadrierten Ladungen für die Variable j

### 2.8.5 Oblique Rotationsmethoden

Wird Oblique rotiert, gibt man die Unabhängigkeit der Faktoren auf. In Abbildung 5 wird Faktor 1 um den Winkel  $\alpha$  und Faktor 2 um den Winkel  $\beta$  gedreht. Diese zwei Beispiel werden den schwiefwinkligen/obliquen Verfahren zugeordnet:

- Promax

Der Ausgangspunkt der Promax-Rotation ist das Ergebnis einer Varimax-Rotation. Zunächst wird eine Zielladungsmatrix  $\mathbf{T}$  konstruiert, indem die Elemente der varimax-rotierten Ladungsmatrix  $\tilde{\mathbf{L}}$  potentiert werden. Vorzeichen der Ladungen bleiben dabei erhalten:

$$\mathbf{T} = (\tilde{l}_{jl} \cdot |\tilde{l}_{jl}^{m-1}|) \quad \text{mit } \tilde{\mathbf{L}} = (\tilde{l}_{jl})$$

Dieser Zielmatrix wird die Ladungsmatrix aus der Varimax-Rotation durch Regression angenähert (Revelle (2015, siehe Funktion: Promax)).

- Oblimax

Diese Rotation entspricht der orthogonalen Quartimax-Methode. Allerdings wird die Bedingung der orthogonalen Faktoren fallen gelassen.

## 2.9 Interpretation der Faktoren

Nach der Zuordnung der Variablen zu den Faktoren, werden die Faktoren für gewöhnlich interpretiert. Anhand der Variablen wird versucht eine Bezeichnung zu finden. In Abbildung 1 liegen Variablen vor, die bereits zugeordnet sind. Die Bezeichnungen „Geselligkeit“ und „Sportliche Aktivitäten“ sind das Ergebnis der Interpretation.

## 2.10 Faktorenwerte

Zu Beginn der Faktorenanalyse war die Datenmatrix mit den Ausprägungen der Objekte/Personen auf den Variablen vorhanden. Der letzte Schritt der Faktorenanalyse ist es, falls notwendig, die Ausprägungen der Objekte/Personen auf die Faktoren zu bestimmen. Mit den Faktorenwerten ist es möglich, weitere Analysen, wie eine Faktorenanalyse zweiter Ordnung, durchzuführen (Noack; 2007, S.55). Wurde eine Hauptkomponentenmethode durchgeführt, können die Faktorenwerte direkt berechnet werden, weil keine spezifischen Faktoren geschätzt werden müssen. Nach einer ML-Faktorenanalyse oder der Hauptfaktorenanalyse sind je nach Ausgangslage des Datensatzes und Eigenschaften der Faktoren unter anderem folgende Schätzverfahren möglich: (gewichtete) Summenscores, Multiple Regression, Bartlett-Methode, Anderson-Rubin-Verfahren (DiStefano et al.; 2009) oder die Schätzung nach Ten Berge (Grice; 2001, S.433).

### 3 Anwendungsbeispiel

Folgendes Anwendungsbeispiel veranschaulicht die Durchführung einer explorativen Faktorenanalyse mit dem Statistikprogramm R. Die notwendigen Funktionen bieten die Pakete „psych“ (Revelle (2015)) und „GPArotation“ (Bernards und I.Jennrich (2005)). Für Grafiken zur Veranschaulichung wird zusätzlich das Paket „corrplot“ (Wei (2013)) benutzt.

#### 3.1 Erzeugen der Daten

Zu Beginn werden erst Daten für die Anwendung gezogen. Dies kann man mit der Funktion `sim.hierarchical` machen. Mit dieser Funktion können Korrelationsmatrizen in gewünschter hierarchischer Faktorenstruktur simuliert werden. Um eine anschauliche Faktorenstruktur darstellen zu können, werden hier 20 Variablen erzeugt, die sich in 5 Faktoren zusammenfassen lassen. Diese korrelierten Faktoren lassen sich anschließend in einer weiteren Faktorenanalyse in Faktoren 2.Ordnung anordnen.

Folgende Grafik veranschaulicht die Faktorenstruktur:

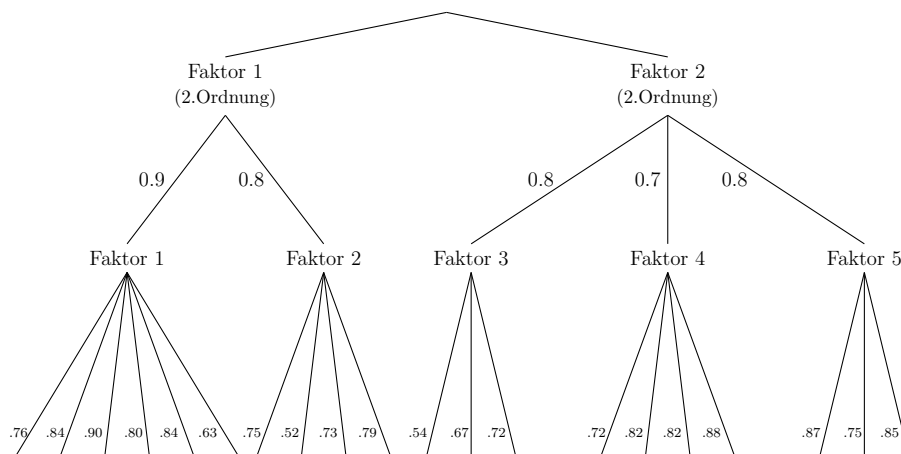


Abbildung 6: Faktorenstruktur des erzeugten Datensatzes

Für die Erzeugung der Daten müssen die Ladungen ( $\mathbf{L}$ ) der Variablen auf die Faktoren und die Ladungen ( $\mathbf{L}_f$ ) der Faktoren auf die Faktoren der 2.Ordnung angegeben werden. Diese sind in Abbildung 6 zu sehen. Die angegebenen Ladungen sind für eine eindeutige Zuordnung entsprechend hoch gewählt. Es

werden die Werte von 1000 Individuen für 20 Variablen ausgegeben.

Folgende Theorie steckt hinter der Datengenerierung:

Um den Zufallsvektor  $\mathbf{z}$  zu erhalten, mit dem die Faktorenanalyse durchgeführt wird, wird von folgender Beziehung ausgegangen:

$$\mathbf{z} = \mathbf{G}\mathbf{x}$$

$\mathbf{z}$  ist der Vektor der korrelierten Zufallsvariablen/Daten mit gewünschter Kovarianzstruktur  $\Sigma$ .

$\mathbf{x}$  ist ein Vektor aus unkorrelierten Zufallsvariablen.

Durch Multiplikation von  $\mathbf{x}$  mit der Matrix  $\mathbf{G}$  erhält man den Vektor  $\mathbf{z}$ .

Für die Kovarianz von  $\mathbf{z}$  gilt:

$$\Sigma = Cov(\mathbf{z}) = Cov(\mathbf{G}\mathbf{x}) = \mathbf{G} Cov(\mathbf{x}) \mathbf{G}^T = \mathbf{G}\mathbf{G}^T$$

und aus der Spektralzerlegung von  $\Sigma$  folgt:

$$\mathbf{G} = \mathbf{P} \text{diag}(\sqrt{\lambda})$$

wobei  $\mathbf{P}$  die Matrix der Eigenvektoren und  $\lambda$  der Vektor der Eigenwerte von  $\Sigma$  ist.

Wie in 2.3 gezeigt, gilt für die Kovarianz von  $\mathbf{z}$  mit der Modellgleichung  $\mathbf{z} = \mathbf{L}\mathbf{f}$ :

$$\Sigma = \mathbf{L} Cov(\mathbf{f}) \mathbf{L}^T$$

mit  $Cov(\mathbf{f}) = \mathbf{L}_f \cdot \mathbf{L}_f^T$ ,

wobei  $\mathbf{L}_f$  die Matrix der Ladungen der Faktoren auf die Faktoren der 2. Ordnung ist.

Mit den als Argumente der Funktion `sim.hierarchical` anzugebenden Ladungsmatrizen wird  $\Sigma$  berechnet. Um die korrelierten Daten  $\mathbf{z}$  zu generieren, werden unkorrelierte, normalverteilte Zufallszahlen  $\mathbf{x}$  gezogen und mit der Matrix  $\mathbf{G}$  multipliziert.

### 3.2 Berechnung der Korrelationsmatrix

Folgende Grafik zeigt die Korrelationen zwischen den 20 simulierten Variablen als Heatplot/Corrplot dargestellt:

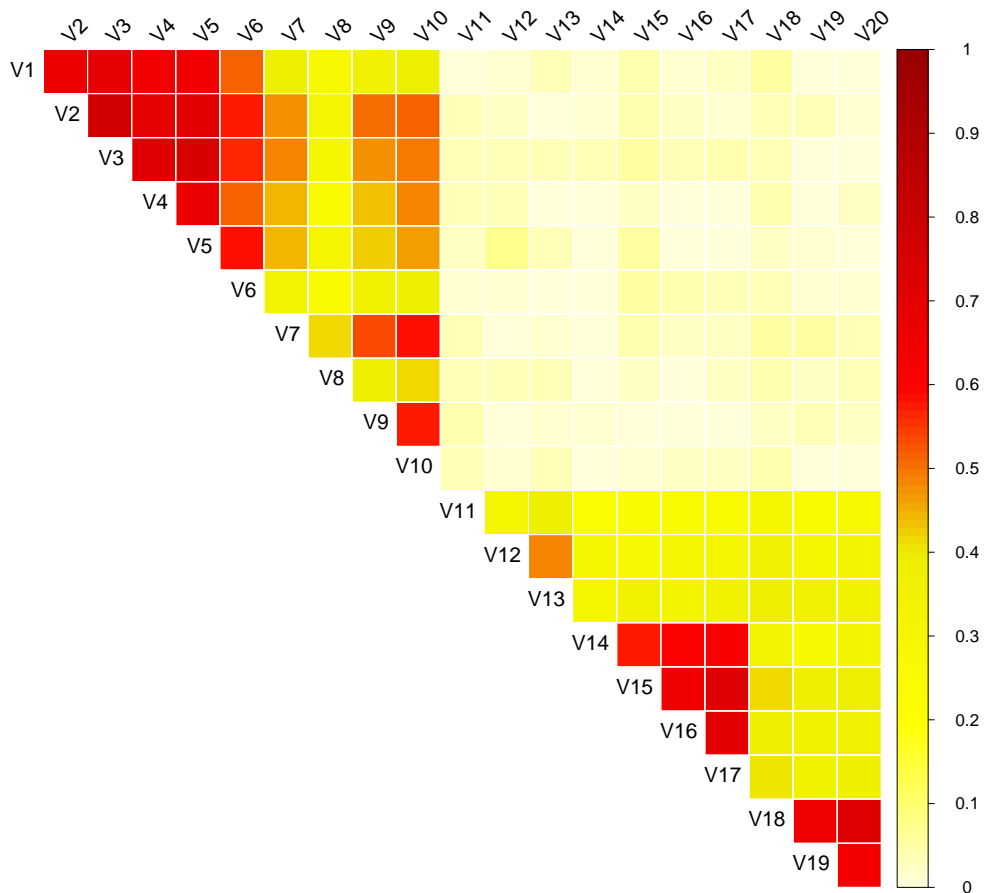


Abbildung 7: Korrelationsmatrix der 20 Variablen

Dieser Plot veranschaulicht gut die Korrelationsstruktur, wobei Rot für eine hohe und Weiß und Hellgelb für keine bzw. niedrige Korrelation steht: Es ist zu erkennen, dass die Variablen 1 bis 10 und die Variablen 11 bis 20 untereinander korrelieren. Die Variablen 1 bis 10 korrelieren aber nur sehr gering mit den Variablen 11 bis 20. Die Variablen 1 bis 10 werden den Faktoren 1 und 2 zugeordnet, welche den übergeordneten Faktor 1 der 2.Ordnung bilden. Die Variablen 11 bis 20 bilden die Faktoren 3, 4 und 5,

welche in einer Faktorenanalyse 2. Ordnung dem Faktor 2 der 2. Ordnung zugeordnet werden (hierzu nochmal die Veranschaulichung durch Grafik 6). Da die Faktoren der 2. Ordnung, im Gegensatz zu den Faktoren 1 bis 5 (1. Ordnung) orthogonal, d.h. unabhängig sind, besteht auch zwischen den Variablen von verschiedenen Faktoren der 2. Ordnung keine bzw. eine sehr geringe Korrelation. Dies erklärt die auffällige Korrelationsmatrix.

### 3.3 Bestimmung der Faktorenanzahl

Für die Bestimmung der Faktorenanzahl wird zuerst ein Scree-Plot betrachtet:

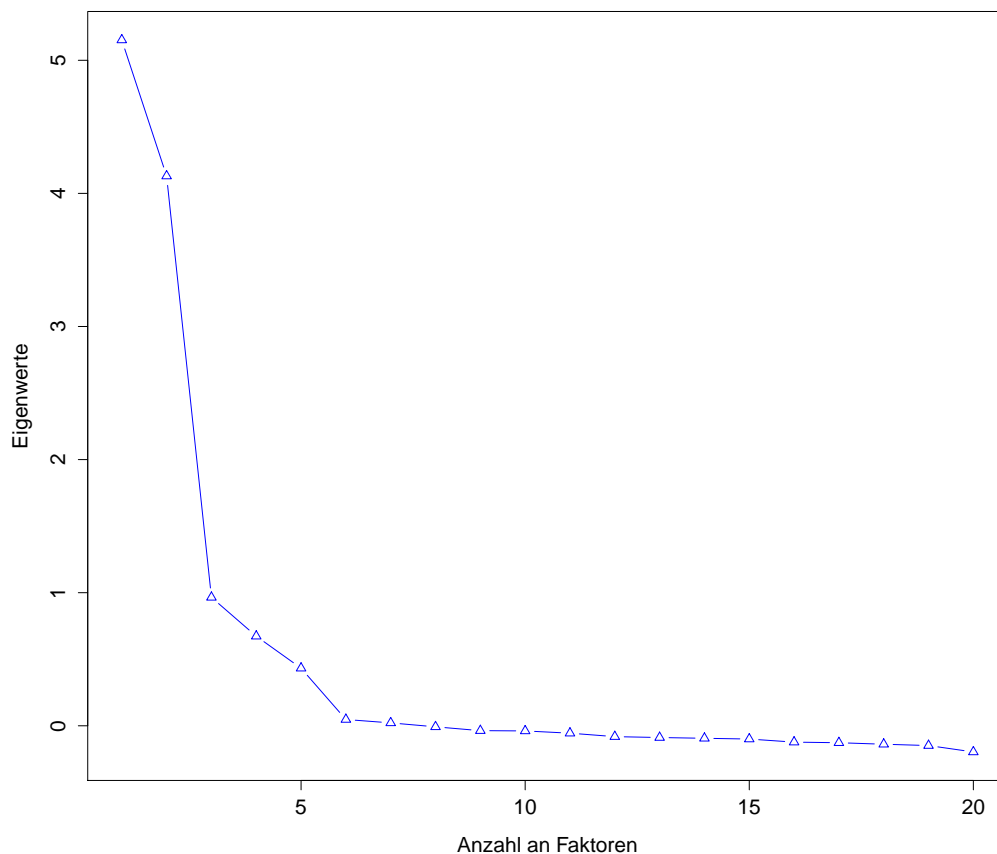


Abbildung 8: Scree-Plot

Die „Knickstellen“ im Kurvenverlauf der Eigenwerte trennen die bedeutsamen von den weniger bedeutsamen Faktoren, d.h. alle Faktoren mit Eigenwerten

vor der Knickstelle werden extrahiert. In Abbildung 8 sind zwei Knicke zu erkennen. Der erste liegt bei Eigenwert Nummer 3 und der zweite bei Eigenwert Nummer 6. Dementsprechend ist es nach der grafischen Scree-Plot-Methode sinnvoll 2 bzw. 5 Faktoren zu extrahieren.

Hier zeigt sich auch ein Problem der Methode: Oft enthält der Kurvenverlauf der Eigenwerte mehrere Knickstellen bzw. keine klar zu erkennende Knickstelle, sodass eine eindeutige Entscheidung, wie viele Faktoren extrahiert werden sollen, nicht möglich ist. Sowohl die Entscheidung für die Extraktion von 2 als auch von 5 Faktoren wäre, in Anbetracht der generierten Faktorenstruktur (siehe Abbildung 6), zulässig.

Für eine eindeutigere Anzahlbestimmung werden dennoch zusätzlich das Eigenwertkriterium und die Parallelanalyse angewendet:

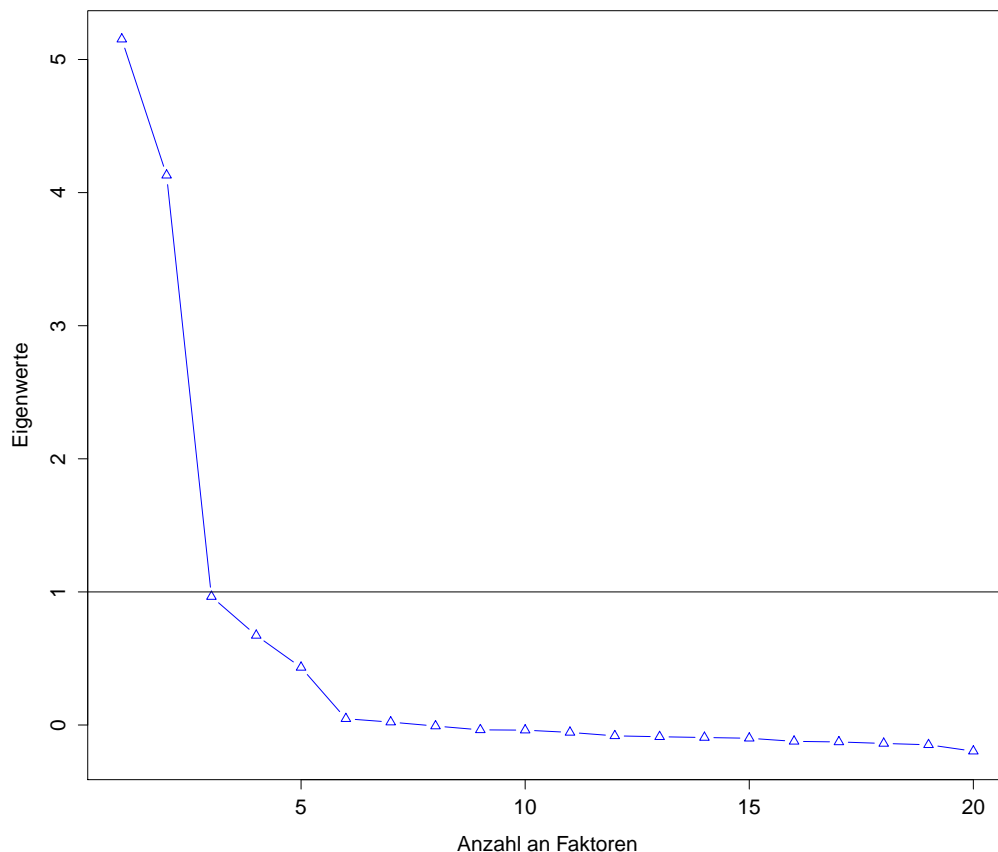


Abbildung 9: Eigenwertkriterium



Nach dem Eigenwertkriterium entspricht die Anzahl an Faktoren 2, weil zwei Eigenwerte größer 1 sind ( $\approx 5.15$ ,  $\approx 4.13$ ).

Wird eine Parallelanalyse durchgeführt, sind 5 Eigenwerte auffällig, da diese größer als die Eigenwerte der Bootstrap-Stichproben sind. Die rote, gestrichelte Linie kennzeichnet den Eigenwertverlauf der Bootstrap-Ziehungen.

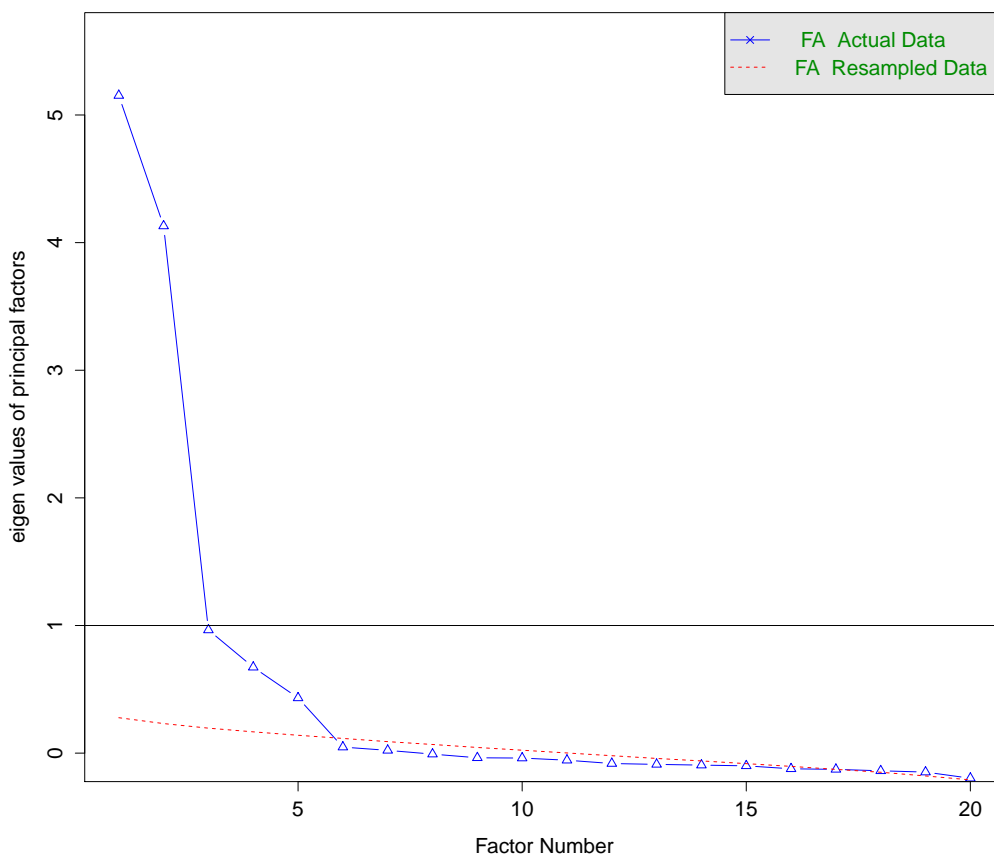


Abbildung 10: Parallelanalyse

Für die Analyse wurden 100 Bootstrapstichproben aus dem ursprünglichen Datensatz gezogen mit gleichem Stichprobenumfang und gleicher Anzahl an Variablen. Eine Änderung der Anzahl an Bootstrapstichproben beeinflusst die Faktorenanzahl nicht.

Mit dem Wissen der wahren Faktorenstruktur und um sich die Möglichkeit offenzuhalten, bei ausreichend hoher Korrelation den zwischen Faktoren eine Faktorenanalyse 2.Ordnung durchzuführen, wird eine Faktorenzahl von 5 angenommen.

### **3.4 Schätzung der Kommunalitäten und Berechnung der Ladungsmatrix**

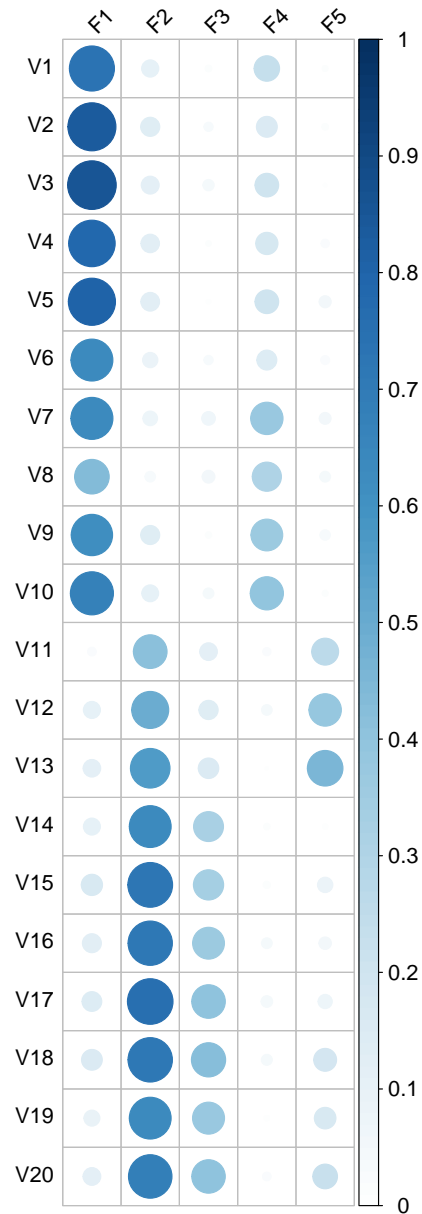
Nun wird eine Hauptfaktorenanalyse angewendet, weil davon ausgegangen wird, dass spezifische Faktoren vorliegen.

Mithilfe der Spektralzerlegung wird die reduzierte Korrelationsmatrix in ihre Eigenvektoren und Eigenwerte zerlegt. Für die anschließende iterative Berechnung der Ladungsmatrix sind Startkommunalitäten zu bestimmen, die auf die Diagonale der Korrelationsmatrix gesetzt werden. Die Funktion `fa` aus dem Paket „psych“ bietet die Möglichkeit die quadrierten multiplen Korrelationskoeffizienten zwischen der Variable  $z_j$  und den übrigen Variablen als Startwerte zu wählen. Mit diesen Werten auf der Hauptdiagonale der Korrelationsmatrix beginnt die Iteration, die endet wenn die Kommunalitäten nicht nach 100 Wiederholungen konvergieren oder, falls Konvergenz eintritt, die Differenz der Kommunalitäten zu dem vorangegangenen Iterationsschritt nicht mehr als 0.001 beträgt.

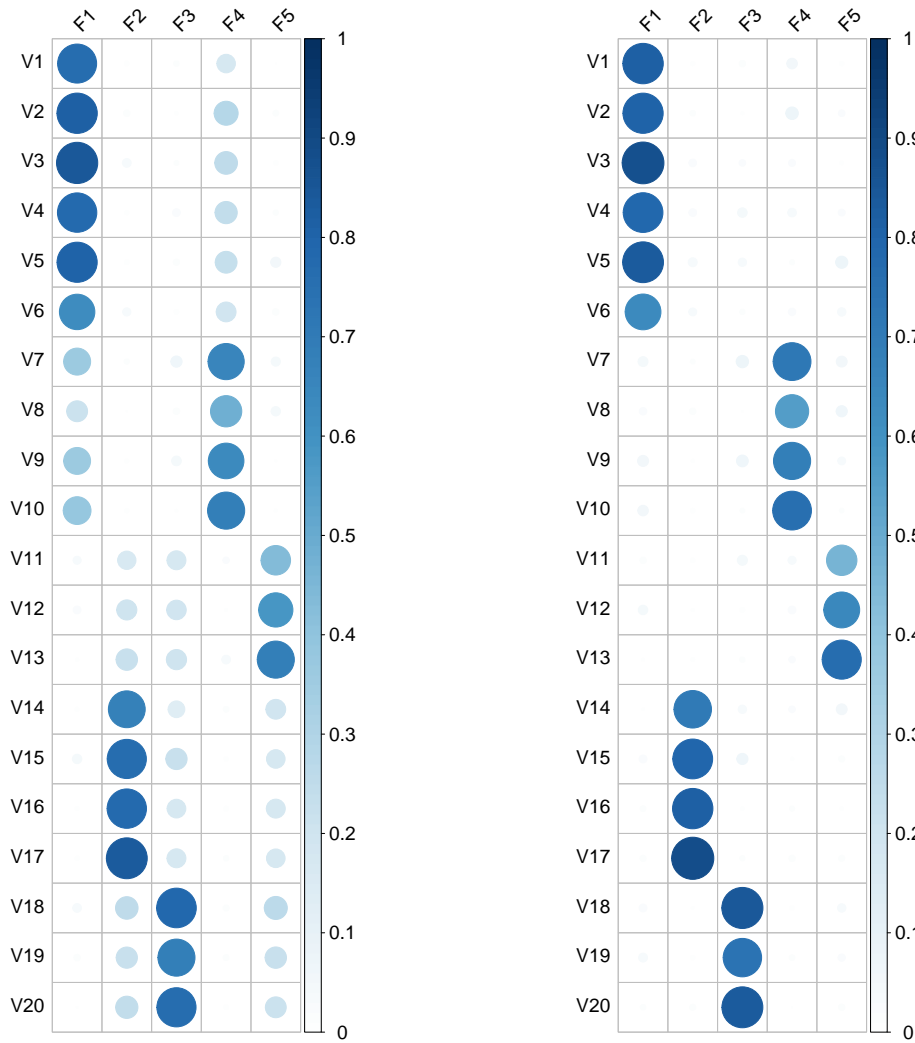
In diesem Anwendungsbeispiel konvergieren die Kommunalitäten nach 10 Schritten.

Abbildung 11: Ladungsmatrix vor der Rotation

Mit dem Vorliegen der Kommunalitäten und der daraus ermöglichten Berechnung der Ladungsmatrix ist eine Zuordnung der Variablen zu den Faktoren erstmals durchführbar. Für eine eindeutigere Zuordnung der Variablen, ist eine Rotation sinnvoll. Nebenstehende Grafik macht dies ersichtlich: In Abbildung 11 sind die Ladungen der 20 Variablen auf die Faktoren als Kreise dargestellt. Je größer und blauer ein Kreis ist, desto betragsmäßig größer ist die Ladung. Eine Variable wird dem Faktor zugeordnet, auf den sie am höchsten lädt. Es ist zu erkennen, dass manche Variablen auf mehrere Faktoren hochladen. Dann ist eine Zuordnung nicht möglich. Außerdem laden die Variablen hauptsächlich auf den ersten und zweiten Faktor, was eine Interpretation dieser Faktoren erschwert. Durch Rotationsverfahren kann dieses Problem behoben werden, indem große Ladungen maximiert und kleine Ladungen minimiert werden.



Vergleicht man eine orthogonale und eine oblique Rotation miteinander, fällt auf, dass eine oblique Rotation eine eindeutigerere Lösung liefert wie in den Grafiken 12 offensichtlich. Die Varimax-Rotation scheint, verglichen mit der



(a) Varimax-Rotation (orthogonal)

(b) Promax-Rotation (oblique)

Abbildung 12: Vergleich von orthogonaler und obliquen Rotationslösung

obliquen Promax-Rotation, nicht optimal zu sein, da es nach ihrer Anwendung dennoch viele verhältnismäßig große Nebenladungen gibt. Da eine oblique Rotation eine eindeutigerere Zuordnung ermöglicht, handelt es sich um Fakto-

ren, die miteinander korrelieren. Andere Rotationen für korrelierte Faktoren erweisen sich genauso zielführend und bestätigen das Ergebnis nach der Promax-Methode. Die Zuordnung der Faktorenanalyse, d.h. die Gruppierung der einzelnen Variablen zu Faktoren, entspricht der Faktorenstruktur, wie in Abbildung 6 dargestellt.

### **3.5 Faktorenanalyse zweiter Ordnung**

Um die eben extrahierten fünf Faktoren weiter zu gruppieren, werden Faktorenwerte benötigt. Die Matrix der Faktorenwerte stellt die neue Datenmatrix dar, mit der die eben durchgeführten Schritte der Faktorenanalyse wiederholt werden. Die Faktorenwerte werden nach der Methode von „Ten Berge“ berechnet, die die Korrelation zwischen den Faktoren bei der Schätzung berücksichtigt. Nach der Berechnung der Korrelationsmatrix wird das Eigenwertkriterium angewendet und eine Parallelanalyse hinzugezogen. Bei beiden Verfahren scheinen 2 Faktoren eine sinnvolle Wahl für die Faktorenzahl zu sein. Analog zur Faktorenanalyse 1.Ordnung wird die Ladungsmatrix nach der Hauptfaktorenanalyse berechnet, unter der Annahme, dass spezifische Faktoren vorhanden sind. Allerdings wird im nächsten Schritt diesmal mit einer Varimax-Rotation rotiert, weil davon ausgegangen wird, dass die Faktoren 2.Ordnung orthogonal zueinander sind. Im Anhang befinden sich sowohl die rotierte und unrotierte Ladungsmatrix als auch die Korrelationsmatrix der Faktoren als 'corrplot', zudem die Grafik einer Parallelanalyse. Das Ergebnis der 2.Analyse bestätigt die simulierte Faktorenstruktur: Zwei Faktoren mit den Variablen 1 bis 10 bilden einen übergeordneten Faktor und drei Faktoren mit den Variablen 11 bis 20 bilden einen weiteren übergeordneten Faktor.

## 4 Einbettung der Arbeit in den Seminar-kontext

In der Wirtschafts- und Sozialstatistik finden viele Methoden Verwendung, um latente Variablen zu modellieren. Umfangreiche Studien, bauen maßgeblich auf solchen Methode auf. Die explorative Faktorenanalyse, als eine davon, kann man, wie schon erwähnt, große Datenmengen für weitere Verfahren vereinfachen, indem einzelne Items zu übergeordneten Variablen zusammengefasst werden. Außerdem ist es möglich, mit der explorativen Faktorenanalyse Dimensionen zu untersuchen. So lässt sich z.B. Eindimensionalität überprüfen, eine Annahme des Rasch-Modells, zudem wichtig für die Test- und Fragebogenkonstruktion. Wird beispielsweise ein Test konstruiert, mithilfe dessen bestimmte Kompetenzen ermittelt werden sollen, so ist es notwendig diejenigen Items zu ermitteln, die die gleiche Kompetenz erfassen. Die Items sind mit Schwierigkeiten verbunden, die nicht eindeutig einer Kompetenz zu zuordnen sind, da bei der späteren Analyse nicht sicher ist, auf was sich die Items beziehen, d.h. was die Items messen sollen. Ob eine Item in einen Fragebogen aufgenommen wird, kann maßgeblich davon abhängen, ob die Annahme der Eindimensionalität verletzt ist oder nicht.

Mit der explorativen Faktorenanalyse stößt man allerdings dann an seine Grenzen, sobald Daten vorliegen, die nicht metrisch bzw. ordinalskaliert sind. Bei vielen Fragebögen werden Antworten gegeben, die kategorial sind, aber nicht ordinales Skalenniveau besitzen. Items, wie Geschlecht, Beruf, Religion oder Staatsangehörigkeit sind grundsätzlich problematisch, wenn mit der explorativen Faktorenanalyse gearbeitet wird.

## Literatur

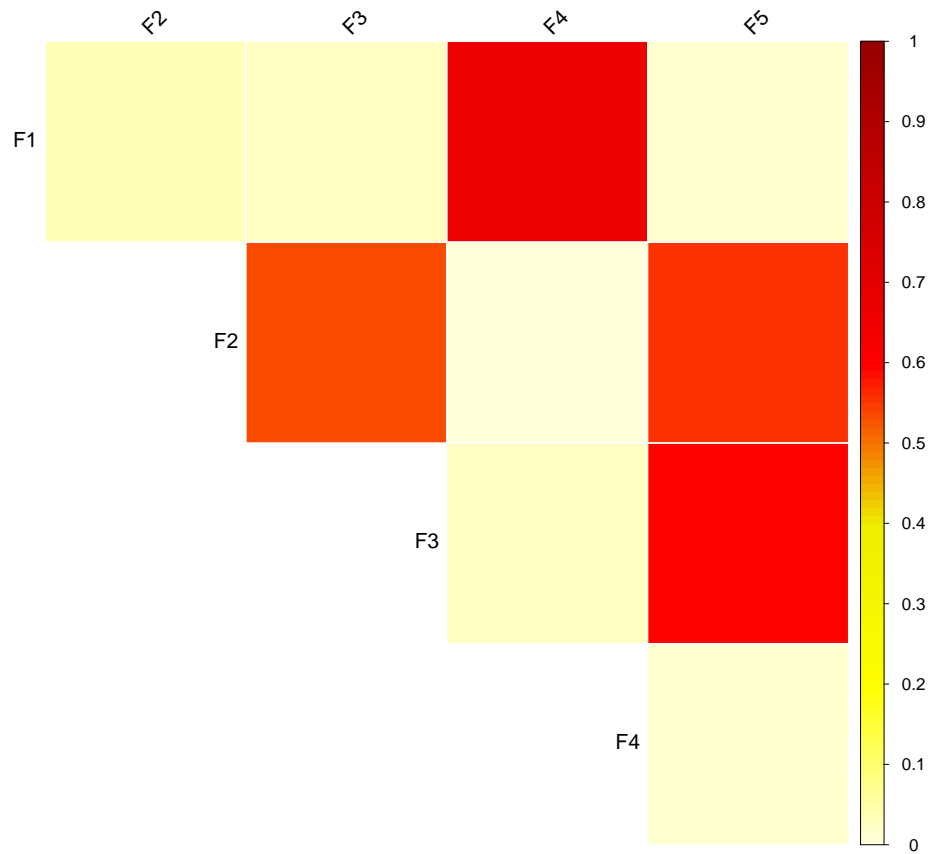
- Bernaards, C. A. und I.Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis, *Educational and Psychological Measurement* **65**: 676–696.
- Bortz, J. und Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler. Lehrbuch mit Online-Materialien*, Springer.
- Brosius, F. (2011). *SPSS 19*, mitp.
- Bühl, A. (2008). *SPSS 16: Einführung in die moderne Datenanalyse*, Pearson Studium.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*, Pearson Studium.
- DiStefano, C., Zhu, M. und Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher, *Practical Assessment, Research & Evaluation* **14**(20): S. 1–11.
- Everitt, B. und Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*, Springer New York.
- Fahrmeir, L., Brachinger, W., Hamerle, A. und Tutz, G. (1996). *Multivariate statistische Verfahren*, de Gruyter.
- Garrido E., L., Abad J., F. und Ponsoda, V. (2011). Performance of velicer’s minimum average partial factor retention method with categorical variables, *Educational and Psychological Measurement* **71**: S.551–570.
- Grice, J. W. (2001). Computing and evaluating factor scores, *Psychological Methods* **6**(4): S.430–450.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis, *Psychometrika* **30**(2): S.179–185.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis, *Psychometrika* **23**(3): S.287–200.

- Knüsel, L. (2008). Factor analysis: Chisquare as rotation criterion, Website. aufgerufen am: 25.05.2015.  
**URL:** <http://epub.ub.uni-muenchen.de/6350/1/tr040.pdf>
- Noack, M. (2007). Faktorenanalyse, Website. aufgerufen am: 20.05.2015.  
**URL:** <https://www.uni-due.de/imperia/md/content/soziologie/stein/faktorenanalyse.pdf>
- Revelle, W. (2015). *psych: Procedures for Psychological, Psychometric, and Personality Research*, Northwestern University, Evanston, Illinois. R package version 1.5.4.  
**URL:** <http://CRAN.R-project.org/package=psych>
- Wei, T. (2013). *corrplot: Visualization of a correlation matrix*. R package version 0.73.  
**URL:** <http://CRAN.R-project.org/package=corrplot>

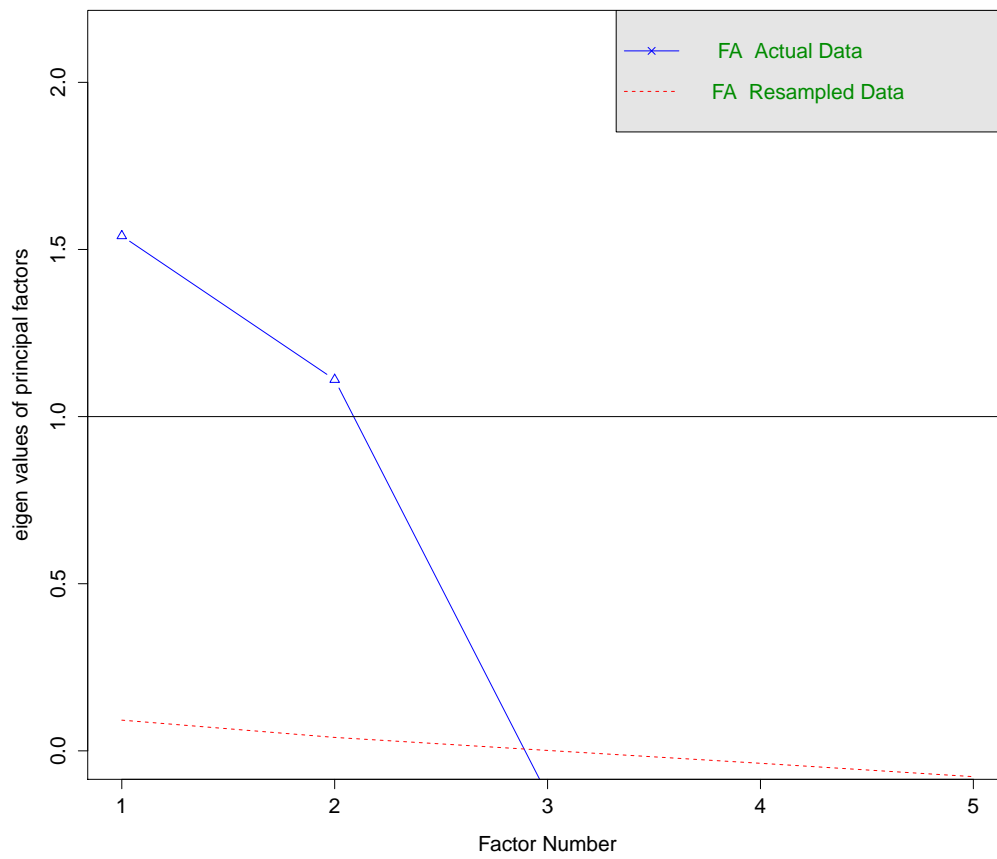


# A Anhang

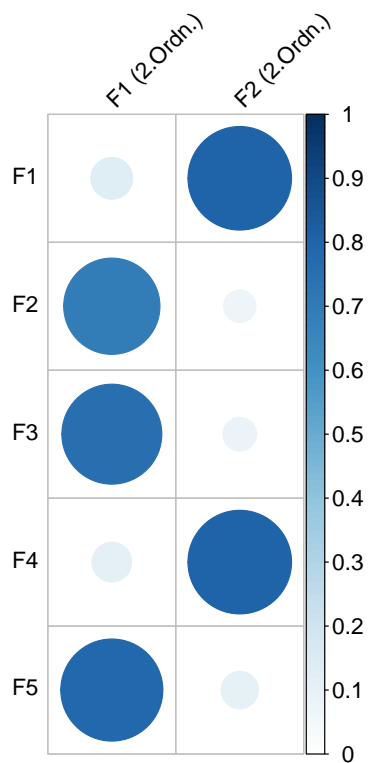
## A.1 Korrelationsmatrix der Faktoren 1.Ordnung



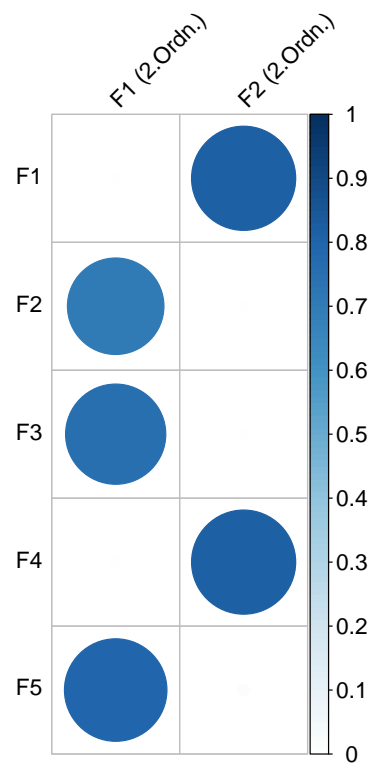
## A.2 Parallelanalyse für die Faktorenanalyse 2. Ordnung



### A.3 Ladungsmatrizen der Faktorenanalyse 2.Ordnung vor und nach einer Varimax-Rotation



(a) Unrotierte Ladungsmatrix



(b) Rotierte Ladungsmatrix

## Erklärung zur Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Seminararbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

München, den 23. September 2015

A handwritten signature in black ink that reads "Simon Reitzner". The signature is written in a cursive style with a prominent, sweeping flourish at the end of the name.

(Simon Reitzner)