

# Ein Vergleich von Modellen der explorativen Faktorenanalyse

(Comparison of Factor Models in Exploratory Factor Analysis)

von

**Raphael Wilhelm Hanke**

(580092)

Bachelorarbeit eingereicht bei:

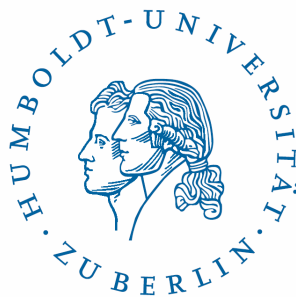
Erstgutachter: **Prof. Dr. Wolfgang Karl Härdle**

Zweitgutachter: **Prof. Dr. Brenda Lopez Cabrera**

Betreuer: **Dr. Sigbert Klinke**

zur Erlangung des akademischen Grades

**Bachelor of Science [B.Sc.] in Betriebswirtschaftslehre**



Humboldt-Universität zu Berlin

C.A.S.E. Center for Applied Statistics and Economics

Ladislaus von Bortkiewicz Chair of Statistics

Berlin, 08. Februar 2018

## Abstract

In der Faktorenanalyse beeinflusst der Wissenschaftler den Modellaufbau. Er bestimmt die Kommunalitätenschätzung, die Anzahl der Faktoren, die Extraktions- und die Rotationsmethode sowie die Interpretation. Um die Konsequenzen der Entscheidungen zu messen, wurde begleitend zur Arbeit eine interaktive Applikation zum Modellvergleich entwickelt. Anhand von Persönlichkeitsmerkmalen wird in verschiedenen Faktorenmodellen die Relevanz der Modellkonstruktion demonstriert und Relationen zwischen Modellstruktur und Ergebnissen werden nachgewiesen. Obwohl alle verglichenen Modelle in Faktoren resultieren, die die gleichen Variablen beinhalten, lassen sich Unterschiede in der Achsenlage, in der Einfachheit der Ladungsstruktur und in der Interpretation einzelner Faktoren aufdecken. Abhängigkeiten zwischen präferiertem Modell und untersuchtem Datensatz werden erkenntlich.

In factor analysis the scientist influences model construction. He determines estimation of communalities, the number of factors, extraction- and rotation-method as well as the interpretation. In order to compare factor models and to indentify the impact of decisions an interactive application is developed. A comparison of several models using data of personality items confirms the importance of model construction, even though all models result in factors that relate to the same variables. Small differences in axis location, simplicity of loading structure and interpretation of factors can be detected. Furthermore a dependency between preferred model construction and the examined data is unveiled.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>iv</b>
<b>Tabellenverzeichnis</b>	<b>v</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Problemstellung . . . . .	1
1.2 Verlauf . . . . .	2
<b>2 Die Faktorenanalyse in der Theorie</b>	<b>3</b>
2.1 Die explorative Faktorenanalyse . . . . .	3
2.1.1 Statistischer Kontext . . . . .	3
2.1.2 Entwicklung und Literaturreview . . . . .	4
2.2 Das Fundamentaltheorem . . . . .	6
2.3 Eignung der Ausgangdaten . . . . .	9
2.4 Wahl der Faktorenanzahl . . . . .	11
2.5 Das Kommunalitätsproblem . . . . .	13
2.6 Die Faktorenextraktion . . . . .	15
2.7 Die Faktorenrotation . . . . .	18
2.7.1 Das Kriterium der Einfachstruktur . . . . .	19
2.7.2 Orthogonale Rotationen . . . . .	20
2.7.3 Oblique Rotationen . . . . .	23
2.7.4 Diskussion zum Rotationsproblem . . . . .	25
2.8 Interpretation der Faktorladungen . . . . .	26
<b>3 Implementierung der Problemstellung in interaktiver Applikation</b>	<b>28</b>
3.1 Zielsetzung . . . . .	28
3.2 R & Shiny . . . . .	28
3.3 Umsetzung und Ergebnis . . . . .	29
<b>4 Ein Vergleich von Faktorenmodellen anhand von Persönlichkeitsitems</b>	<b>32</b>
4.1 Datensatz, Vorbereitung & Deskriptive Analyse . . . . .	32
4.2 Kommunalitätenschätzung und Auswahl der Faktorenanzahl . . . . .	36
4.3 Kriterien für ein Vergleich von Faktorenmodellen . . . . .	38
4.4 Test und Interpretation verschiedener Modelle . . . . .	40
<b>5 Fazit</b>	<b>49</b>

Literaturverzeichnis	51
A Tabellen	54
B CD	60

## Abbildungsverzeichnis

1	Graphen zur Demographie . . . . .	34
2	Heatmap der empirischen Korrelationsmatrix . . . . .	35
3	Scree-Plot von empirischer und reduzierter Korrelationsmatrix mit Parallelanalyse (Eigenwertverläufe normalverteilter Zufallsvariablen in rot) und Kaiserkriterium ( $\gamma = 1$ ) . . . . .	37
4	Diagramm der unrotierten Faktorladungen nach Hauptachsen-Extraktion (links) und Maximum-Likelihood-Extraktion (rechts) . . . . .	41
5	Heatmap der Faktor-Kongruenz-Matrix (Varimax auf y-Achse, Promax auf x-Achse) . . . . .	44
6	Heatmap der Faktor-Kongruenz-Matrix (Promax auf y-Achse, Oblimin auf x-Achse) . . . . .	48

## Tabellenverzeichnis

1	Mit SMC geschätzte Kommunalitäten der Variablen gerundet auf 3 Nachkommastellen . . . . .	36
2	Unrotierte Faktorladungen nach Hauptachsen-Extraktion . . . . .	40
3	Faktorladungen nach Varimax-Rotation (links) und Promax-Rotation (rechts)	43
4	Korrelationsmatrix der Faktoren nach obliquer Promax-Rotation . . . . .	45
5	Faktorladungen nach Promax-Rotation (links) und Direct-Oblimin (rechts) . .	46
6	Korrelationen nach Promax-Rotation (links) Direct-Oblimin (rechts) und Maximum-Likelihood-Extraktion . . . . .	47
7	Unrotierte Faktorenlösung nach Maximum-Likelihood Extraktion . . . . .	54
8	Faktorladungen nach Varimax-Rotation und Hauptachsen-Extraktion . . . . .	55
9	Faktorenmuster nach Promax-Rotation und Hauptachsen-Extraktion inklusive Korrelationsmatrix der Faktoren . . . . .	56
10	Faktorenstruktur nach Promax-Rotation und Hauptachsen-Extraktion . . . .	57
11	Faktorenmuster nach Promax-Rotation und Maximum-Likelihood-Extraktion .	58
12	Faktorenmuster nach Direct-Oblimin-Rotation und Maximum-Likelihood-Extraktion	59

# 1 Einleitung

## 1.1 Problemstellung

Thematik der vorliegenden Arbeit ist ein Vergleich von Modellen der explorativen Faktorenanalyse, mit der Intention Beziehungen zwischen Modell Aufbau und Ergebnissen zu identifizieren.

Besonders im Kontext der Psychologie und den Sozialwissenschaften ist die Methode der Faktorenanalyse geläufig und weit verbreitet. Die Faktorenanalyse ist eine sehr komplexe Analysemethode. Sie bedarf an vielen Stellen des Prozesses eines Eingriffs des Wissenschaftlers, der in der Pflicht ist, methodologische Entscheidungen zu treffen. Neben der Wahl einer faktorenanalytischen Methode ist die Anzahl der extrahierten Faktoren, die Art der Kommunalitätsschätzungen sowie eine Entscheidung über die Wahl einer angemessenen Extraktions- und Rotationsmethode zu treffen. Weitere subjektive Eingriffsmöglichkeiten stehen dem Anwender in der Interpretation der Resultate offen. Problematisch ist, dass das Verfahren der Faktorenanalyse, verstärkt durch die Verbreitung elektronischer Datenverarbeitung und statistischer Software, oftmals konzeptionslos eingesetzt wird. Bei einer komplexen Analysemethode mit infinitiven Möglichkeiten das Resultat zu beeinflussen, fehlt oftmals das Hintergrundwissen und der Überblick über die möglichen Optionen. Mit Blick auf die Interpretation des Ergebnisses ist vielen Anwendern die Bedeutung getroffener Entscheidungen nicht bewusst. Bereits Cattell (1978) betonte, dass der Beschluss über die Rotationsmethode "in half of the cases it is evidently done in ignorance of the issue than by deliberate intent". Die von Cattell angesprochene Problemstellung ist nach wie vor aktuell. In statistischer Datenverarbeitungssoftware sind in Entscheidungsprozessen häufig Defaultwerte implementiert, die den Anwender zu einer unreflektierten Übernahme treiben. Gleichzeitig entsteht das Problem der richtigen Bewertung und Auswertung faktorenanalytischer Forschung.

Die folgende Arbeit behandelt die Fragestellung über die sachgerechte Modellierung von Faktorenanalysen. Verschiedene Modelle werden analysiert und verglichen. Im Fokus steht die Frage nach der Gewichtung und dem Einfluss bestimmter im Verlauf der Faktorenanalyse getroffener Entscheidungen in Bezug auf das Ergebnis. Wie stark und in welcher Form unterscheiden sich die Ergebnisse verschieden aufgebauter Faktorenmodelle? In welchem Maße sind Beziehungen zwischen Modellstruktur und Resultat greifbar? Lässt sich allgemeingültig ein favorisiertes Modell bestimmen?

## 1.2 Verlauf

Die Einleitung führt zunächst in das Themengebiet der Faktorenanalyse ein, reflektiert ihre geschichtliche Entwicklung und ordnet sie in einen statistischen Kontext ein.

Nach einer theoretischen Einführung in die Grundlagen der Faktorenanalyse wird ihr Analyseverlauf beschrieben und in den jeweiligen Teilschritten ein Überblick über mögliche Herangehensweisen vermittelt. Durch Bezüge zur Fachliteratur schließt sich eine Diskussion über die Reichweite und die Konsequenzen getroffener Entscheidungen an, wobei der Fokus auf der Auswahl der Faktorenrotation liegt.

Dem theoretischen Teil der Arbeit folgt die Dokumentation der Implementierung der Faktorenanalyse und der mit ihr verbundenen Problemstellung in eine interaktive Applikation, die die zuvor erarbeiteten Erkenntnisse visualisiert und einen einfachen Vergleich verschiedener Faktorenmodelle ermöglicht. Nach einer Einführung in R und das verwendete Erweiterungspaket "Shiny" wird die Zielsetzung dem Ergebnis mit seinen Komponenten und technischen Implikationen gegenübergestellt.

Im Anschluss folgt die praktische Umsetzung durch einen exemplarischen Vergleich verschiedenen konstruierter Faktorenmodelle anhand eines ausgewählten Datensatzes. Ich stelle den Analyseprozess dar, Outputs und Zusammenhänge werden visualisiert und analysiert mit dem Ziel, Strukturen aufzudecken und Resultate zu liefern.

Abschließend erfolgt in einem Fazit ein Resümee der Ergebnisse. Die durch die Analyse gewonnenen Erkenntnisse resultieren in Schlussfolgerungen und Empfehlungen für die Konstruktion von Faktorenmodellen.



## 2 Die Faktorenanalyse in der Theorie

Kapitel 1 dient als theoretische Einführung in die explorative Faktorenanalyse. Zunächst von Interesse ist der Modellansatz der Faktorenanalyse, im Anschluss ihre einzelnen Teilschritte im Detail.

### 2.1 Die explorative Faktorenanalyse

Die explorative Faktorenanalyse ist ein multivariates statistisches Verfahren, das eine Vielzahl von Variablen gemäß ihrer korrelativen Beziehung in wenige Variablengruppen klassifiziert. Das Ziel einer Faktorenanalyse ist es, eine Struktur zu identifizieren, die dem betrachteten Variablensatz unterliegt. Zusammenhänge innerhalb der Variablen werden durch eine kleinere Gruppe homogener Faktoren erklärt (Bühner, 2011). Es ist eine Methode zur Variablenreduktion, die für die wechselseitigen Beziehungen vieler Variablen ein einfaches Erklärungsmodell sucht. Das Ergebnis der Analyse sind Faktoren, die Aufschluss über Zusammenhänge zwischen den Variablen geben. Bortz (2013) bezeichnet einen Faktor als eine, aus den Korrelationen zwischen den Variablen konstruierte, synthetische Variable. Sie beschreibt einen latenten Sachverhalt bzw. ein Konstrukt, das den betrachteten Variablen zu Grunde liegt und sich durch eine möglichst hohe Korrelation mit allen Variablen auszeichnet. Laut Härdle and Simar (2007) verleihen die vorerst unbeobachtbaren Faktoren dem Wissenschaftler ein besseres Verständnis der Daten als die vollständigen Beobachtungen.

#### 2.1.1 Statistischer Kontext

Die Faktorenanalyse gehört zu den Verfahren der multivariaten Statistik. Überla (1968) definiert ein multivariates statistisches Verfahren als eine Methode, die an jedem Objekt (z.B. Person) mehrere Variablen misst und alle diese Größen für eine Reihe von Personen in der statistischen Analyse gleichzeitig betrachtet und ausgewertet. Es wird generell zwischen Dependenz- und Interdependenzanalysen unterschieden. In Dependenzanalysen erfolgt eine Partitionierung der Datenmatrix, in der ein Teil der Variablen anderen gegenübergestellt wird. Dependenzen werden gesucht und mit Hilfe unabhängiger Variablen werden Aussagen über eine Abhängige getroffen. In Interdependenzanalysen bleibt diese Partitionierung aus, und das Beziehungsgefüge zwischen gleichrangigen Variablen wird untersucht (Hüttner, 2002; Wirtschaftslexikon, 2017a). Die Faktorenanalyse ist den Interdependenzanalysen zuzuordnen. Abzugrenzen ist eine Faktorenanalyse von einer Clusteranalyse in der Hinsicht, dass sie die Variablen gemäß ihrer korrelativen Beziehung in Gruppen klassifiziert (Bortz, 2013), während

eine Clusteranalyse auf die Elemente gerichtet ist und auf die Bildung von intern möglichst homogener Gruppen der Beobachtungsobjekte abzielt. Des Weiteren muss die explorative Faktorenanalyse von der konfirmativen Faktorenanalyse abgegrenzt werden. Laut Iacobucci (2001) dienen sie verschiedenen Zwecken: Explorative Faktorenanalysen zielen allein auf die Identifikation von Strukturen ab. Die Motivation und ihre Ziele sind in Punkt 2.1 beschrieben. Eine konfirmative Faktorenanalyse dagegen dient zum Test bzw. zur Überprüfung einer a priori an den Datensatz unterstellten dimensional Struktur. Der Wissenschaftler sucht nach Bestätigung seiner Annahmen und hat bereits klare Annahmen vor der Durchführung einer Analyse. Verifizierungen angenommener Hypothesen erfolgen mit Hilfe von Strukturgleichungsmodellen (Iacobucci, 2001). Ich werde mich im Folgenden auf Modelle der explorativen Faktorenanalyse beziehen.

### **2.1.2 Entwicklung und Literaturreview**

Die Faktorenanalyse hatte ihr Blütezeit im 20. Jahrhundert. Entwickelt wurde die Faktorenanalyse von Charles Edward Spearman, der 1904 in "General Intelligence, objectivley determined and measured" die Zwei-Faktoren-Theorie zur Beschreibung von Intelligenz aufstellte. Sie umfasste einen allgemeinen Intelligenzfaktor, den "g-factor", und aufgabenspezifische Faktoren, die sich auf die Lösung bestimmter Aufgaben bezogen (Bortz, 2013). Einen hohen Beteiligungsgrad an der heutigen Faktorenanalyse hatte Louis Leon Thurstone (1947), der mit der multiplen Faktorenanalyse den Grundstein legte. Ebenfalls schuf er mit dem Kriterium der Einfachstruktur, das zentrale Prinzip für die Faktorenrotation. Nach seiner Analysemethode wurden zahlreiche psychologische Untersuchungen durchgeführt. Bis in die 50er Jahre wurde die Faktorenanalyse manuell durchgeführt. Später in den 60er und 70ern wurden mathematisch aufwendigere Methoden entwickelt, was stark mit dem technischen Fortschritt zusammenhing. Die zuvor verwendeten mechanischen Tischrechenmaschinen wurden bis Ende der 70er Jahren fast vollständig von elektronischen Taschenrechnern und Computern verdrängt (Darton, 1980). Beteiligt am Fortschritt analytischer Rotationsmethoden waren Wissenschaftler wie Henry Felix Kaiser (Covarimin 1958, Varimax 1960) und John B. Carroll (Quartimax 1953, Quartimin 1957, Covarimin 1960). Erwähnenswert ist außerdem Raymond Bernard Cattell, der unter anderem den heutzutage verbreiteten Scree-Test (1966) zur Bestimmung der Faktorenanzahl entwickelte. Die Entwicklung wurde besonders von Psychologen durch human- und sozialwissenschaftliche Fragestellungen im Bereich der Persönlichkeitsforschung vorangetrieben (Bortz, 2013). Als multivariate Analysemethode ist die Faktorenanalyse, mit denen in

der Vergangenheit entwickelten und im Laufe der Zeit sich etablierten Methoden, ein fester Bestandteil der heutigen statistischen Lehre. Hervorzuheben sind die Lehrbücher von Überla (1968), Gorsuch (1983), Härdle and Simar (2007), Bortz (2013) und Backhaus et al. (2015).

Bedeutende Weiterentwicklungen in der heutigen Zeit gibt es wenige, jedoch ist die korrekte Anwendung der Faktorenanalyse auf aktuelle Fragestellungen in Anbetracht ihrer vielen auf subjektiven Entscheidungen basierenden möglichen Verläufen stets von hoher Relevanz.

## 2.2 Das Fundamentaltheorem

Wie lassen sich aus den Ausgangsdaten rechnerisch Faktoren ermitteln? Das Common Factor Modell von Thurstone beschreibt die grundlegende Annahme der Faktorenanalyse. Es stellt den Zusammenhang zwischen Faktorladungsmatrix und Korrelationsmatrix dar, indem es die Reproduzierbarkeit von Korrelationsmatrix der Variablen durch Faktorladungsmatrix und Faktorkorrelationsmatrix beweist (Backhaus et al., 2015). Im Kontext der Faktorenanalyse bezeichnet eine **Faktorladung**  $q_{jl}$  die Korrelation eines l-ten Faktors F mit der j-ten Variablen X. Das im Folgenden aufgeführte Standardmodell und dessen Notation ist angelehnt an Härdle and Simar (2007). Das orthogonale Faktorenmodell beschreibt, wie die observierten p Variablen  $x_j = (x_1, \dots, x_p)^T$  eines Datensatzes von einer kleineren Anzahl latenter Faktoren  $f_l = (f_1, \dots, f_k)$  mit  $k < p$  abhängen. Zunächst wird von voneinander unabhängigen Faktoren ausgegangen. Ausgangspunkt ist die Annahme, dass jeder Beobachtungswert einer Ausgangsvariablen  $x_j$ , bzw. der standardisierten Variablen  $z_j$ , als eine Linearkombination hypothetischer Faktoren darstellbar ist (Backhaus et al., 2015).

$$\begin{aligned} x_j &= q_{j1} \cdot f_1 + q_{j2} \cdot f_2 + \dots + q_{jl} \cdot f_l + u_j \\ x_j &= \sum_{l=1}^k q_{jl} \cdot f_l + u_j \end{aligned} \tag{1}$$

In Matrixschreibweise kann Gleichung 1 wie folgt dargestellt werden:

$$X = QF + U \tag{2}$$

X ist ein  $p \times 1$  Vektor von beobachtbaren Variablen

Q ist eine  $p \times k$  Matrix mit Ladung der j-ten der Variablen auf den l-ten Faktor

F ist ein  $k \times 1$  Vektor des l-ten gemeinsamen Faktors

U ist ein  $p \times 1$  Vektor des j-ten spezifischer Faktor bzw. Fehlerterm

Innerhalb der Definitionsgleichung 2 ist nur X beobachtbar, die weiteren Komponenten Q, F und U sind unbekannt und werden im Verlauf der Faktorenanalyse ermittelt.

Für sie gelten folgende Annahmen:

$$\begin{aligned}
Cov(F, U) &= 0 \\
Var(F) &= I_k \\
Cov(U_i, U_j) &= 0, \quad i \neq j \\
E[F] &= 0 \\
E[U] &= 0 \\
F &\sim (0, I) \\
U &\sim (0, \Psi), \quad \text{wobei} \quad Var(U) = \Psi = diag(\psi_{11}, \dots, \psi_{pp})
\end{aligned} \tag{3}$$

Die Varianz von  $X$  besteht aus zwei Komponenten; der **Kommunalität**  $h_j^2 = \sum_{l=1}^k q_{jl}^2$  und der **spezifischen Varianz**  $\psi_{jj}$ . Sie definiert sich wie folgt:

$$\sigma_{x_j x_j} = Var(X_j) = \sum_{l=1}^k q_{jl}^2 + \psi_{jj} \tag{4}$$

Das Quadrat der Ladung, einer Variablen  $j$  auf den Faktor  $l$ , bestimmt den gemeinsame Varianzanteil zwischen der  $j$ -ten Variablen und dem  $l$ -ten Faktor. Durch die Summe der quadrierten Ladungen einer Variablen über alle  $l$  Faktoren, erhält man die Kommunalität. Die Kommunalität  $h_j^2$  beschreibt, in welchem Ausmaß die Varianz der Variablen  $j$  durch alle Faktoren aufgeklärt wird (Bortz, 2013). Die spezifische Varianz  $\psi_{jj}$  beinhaltet potentielle Messfehler und die von jeder Variablen nicht durch die Faktoren erklärte Residualvarianz (Backhaus et al., 2015; Dorton, 1980). Die  $(p \times p)$  **Kovarianzmatrix** stellt sich in Matrixschreibweise wie folgt dar:

$$Cov(X) = \frac{1}{n-1} X^T X = \frac{1}{n-1} (QF + U)(QF + U)^T = \dots = QR_F F Q^T + \Psi \tag{5}$$

$R_{FF} = C$  verkörpert dabei die Korrelationsmatrix der Faktoren. Zunächst wird von orthogonalen zueinanderstehenden, unkorrelierten Faktoren ausgegangen, wodurch  $C$  eine Einheitsmatrix ( $I$ ) darstellt und wegfällt.

$$Cov(X) = QQ^T + \Psi = \Sigma \tag{6}$$

Ziel der Faktorenanalyse ist es die Ladungen  $Q$  und die spezifische Varianz  $\Psi$  zu finden, durch die sich die Gesamtvarianz erklären lässt. Schätzungen von  $Q$  und  $\Psi$  lassen sich aus der Kovarianzmatrix ableiten (Härdle and Simar, 2007).

Das Modell ist skaleninvariant und wird häufig durch eine z-Standardisierung mit  $Z = \frac{X - \bar{X}}{S_X}$  normalisiert, wodurch  $E(Z) = 0$ , und  $Var(Z) = 1$ . Die zu faktorisierten Variablen werden dadurch bezüglich ihrer Metrik vergleichbar gemacht (Bortz, 2013). Im folgenden wird von standardisierten Werten ausgegangen, es gilt:  $Cov(Z) = R_{zz}$  und  $R_{zz} = \Sigma = QQ^T + \Psi$ . Für standardisierte Variablen sind Korrelationsmatrix und Kovarianzmatrix identisch. Angenommen, es existiere ein Modell in dem k latente Faktoren gefunden werden, würde zur Interpretation der Faktoren  $F_j$  die Korrelationsmatrix  $R_{ZF} = V$  der Faktoren mit den standardisierten Z-Variablen errechnet werden.

Gleichung 5 beschreibt das Fundamentaltheorem der Faktorenanalyse und dient als Ausgangspunkt für die klassischen Lösungsversuche (Darton, 1980). Im Fall von einander unabhängiger Faktoren reduziert es sich zu Gleichung 2.2. "Es besagt, dass die Korrelationsmatrix durch das Faktorenmuster und die Korrelationen zwischen den Faktoren reproduzierbar sein muss" (Überla, 1968, S.53). Den weiteren Verlauf beschreibt Überla wie folgt:

"Jede Faktoranalyse beginnt mit der Datenmatrix X bzw der normalisierten Z. Aus ihr wird die Korrelationsmatrix  $\Sigma$  errechnet. In die Diagonale der Korrelationen werden sodann Schätzungen der Kommunalitäten eingesetzt und man erhält die (reduzierte) Korrelationsmatrix  $\Sigma'$ . Dieses sog. Kommunalitätenproblem, das in der Festlegung geeigneter Schätzwerte  $\hat{h}^2$  besteht, ist das erste Problem, das im Verlauf einer Faktorenanalyse auftritt. Zwischen  $\Sigma'$  und Q liegt das sogenannte Faktorenproblem. Aus  $\Sigma'$  werden nach bestimmten Methoden Faktoren extrahiert, man erhält als Ergebnis die Matrix Q. Die Spalten der Matrix Q sind orthogonal und in einer durch die Extraktionsmethode bestimmten willkürlichen Position, in Bezug auf die Variablen. Es sind viele andere Matrizen Q möglich, die in Bezug auf die Reproduktion von  $\Sigma'$  nach Gleichung gleich gut sind. Aus ihnen muss eine ausgewählt werden d.h. es tritt das sogenannte Rotationsproblem auf." (Überla, 1968, S.63).

Nach Lösung des Rotationsproblems nach einer von mehreren Methoden ist eine Interpretation der Faktoren möglich.

## 2.3 Eignung der Ausgangsdaten

Die Eignung der Ausgangsdaten für eine Faktorenanalyse spiegelt sich in der Korrelationsmatrix wieder, da das Ziel der Faktorenanalyse die Identifizierung latenter Größen ist, die die Zusammenhänge der Variablen beschreiben. Diese Zusammenhänge zwischen einzelnen Variablen können durch die Korrelationsrechnung messbar gemacht werden (Backhaus et al., 2015). Anhand statistischer Prüfkriterien kann die ausreichende Korrelation sowohl individuell als auch der gesamten Korrelationsmatrix gemessen werden.

### Signifikanzprüfung der Korrelationen

Nach der Formulierung der Null-Hypothese (HO) nicht zusammenhängender Variablen, wird diese für jeden Korrelationskoeffizienten anhand des Signifikanzniveaus geprüft. Je niedriger der p-value, also die Irrtumswahrscheinlichkeit, desto eher kann die Nullhypothese abgelehnt werden (Backhaus et al., 2015).

### Bartlett-Test auf Sphärizität

Der von Bartlett 1950/1951 entwickelte Test prüft die Nullhypothese unkorrelierter Variablen in der Grundgesamtheit. Wird diese nicht abgelehnt, sind die Abweichungen der Korrelationsmatrix von einer Einheitsmatrix auf den Zufall zurückzuführen. Vorausgesetzt werden normalverteilte Variablen. Da die Prüfgröße Chi2-verteilt ist, steigt die Wahrscheinlichkeit einer Ablehnung der Nullhypothese mit der Größe der Stichprobe (Rinkenburger, 2011).

### Inverse der Korrelationsmatrix

Durch die Invertierung der Korrelationsmatrix können Schlüsse über die Eignung der Ausgangsdaten gezogen werden. In der Inversen sollten die Werte außerhalb der Diagonalen möglichst nahe bei Null liegen (Backhaus et al., 2015).

### Anti-Image-Kovarianz-Matrix

Die Korrelationsmatrix kann in das Image, dem durch eine Regression erklärbaren Teil, und das Anti-Image zerlegt werden. Die Anti-Image Matrix, die den von den restlichen Variablen unabhängigen Teil betrachtet, wird auf möglichst kleine Werte außerhalb der Diagonalen untersucht. Sind weniger als 25 % ungleich Null ( $>0.09$ ) sind die Ausgangsdaten geeignet (Rinkenburger, 2011).

### **Kaiser-Meyer-Olkin Kriterium (Measuring Sampling Adequacy - MSA)**

Das Verfahren entwickelt eine Prüfgröße, die als Indikator für die Zusammengehörigkeit einzelner Variablen fungiert. Der MSA-Score jeder Variablen liegt zwischen 0 und 1. Durch die gleichmäßige Gewichtung kann ebenfalls ein Score für die vollständige Korrelationsmatrix berechnet werden. Kaiser and Rice (1974) legten Intervalle für die Bewertung des Scores fest:

< 0.5: inakzeptabel (inacceptable)

0.5 - 0.6 schlecht (miserable)

0.6 - 0.7 mäßig (mediocre)

0.7 - 0.8 mittelprächtigt (middling)

0.8 - 0.9 recht gut (meritorious)

0.9 - 1.0 fabelhaft (marvelous)

Rinkenburger (2011) suggeriert die Berücksichtigung mehrerer Kriterien für eine angemessene Beurteilung der Ausgangsdaten. In späteren Analysen wird mit dem Kaiser-Meyer-Olkin Kriterium das laut Backhaus et al. (2015) beste Kriterium verwendet, sowie der Bartlett-Test. Nach Fabrigar et al. (1999) spielt auch die Stichprobengröße eine große Rolle. Sie suggerieren unter normalen Bedingungen eine Stichprobe von mindestens 200 Beobachtungen, merken jedoch an, dass unter schlechten Bedingungen selbst 400-800 Beobachtungen nicht ausreichend sein können.



## 2.4 Wahl der Faktorenanzahl

Bei den meisten empirischen Untersuchungen lässt sich die Gesamtvarianz aller Variablen durch eine Faktorenanzahl erklären, die erheblich kleiner ist als die Anzahl der Variablen. Es stellt sich die Frage nach einem Kriterium für die Bestimmung der richtigen Anzahl der bedeutsamen Faktoren, die extrahiert werden sollen. Die Entscheidung kann auf theoretischen Überlegungen beruhen oder mit statistischen Kennzahlen begründet werden (Rinkenburger, 2011). Im folgendem werden Methoden zur Bestimmung der Faktorenanzahl vorgestellt. Aus der Korrelationsmatrix bzw. der reduzierten Korrelationsmatrix ist es möglich, den Eigenvalue zu errechnen, der dem Wissenschaftler hilft, die durch die latenten Konstrukte reproduzierte Varianz und die Anzahl relevanter Faktoren zu bestimmen. Der **Eigenwert**  $\lambda_l = \sum_{j=1}^p q_{jl}^2$  eines Faktors  $l$  gibt an, wie viel von der Gesamtvarianz aller  $p$  Variablen durch diesen einen Faktor erfasst wird (Bortz, 2013).

### Kaiser-Gutman-Kriterium

Nach dem Kaiser-Guttman-Kriterium werden die  $q$  Faktoren berücksichtigt, deren Eigenwerte größer 1 sind. Jeder Faktor klärt bei Anwendung des Kaiser-Guttman-Kriteriums Varianzanteile mindestens einer Variable auf, vorausgesetzt, es handelt sich um standardisierte Variablen mit Varianz = 1 (Voss, 2009). Die Berechnung der Eigenwerte erfolgt aus der empirischen Korrelationsmatrix und nicht aus der oft fälschlich verwendeten reduzierten Korrelationsmatrix, die in Kapitel 2.5 eingeführt wird (Fabrigar et al., 1999). In der Praxis ist das Kaiser-Gutman-Kriterium das am häufigsten verwendete Kriterium (Rinkenburger, 2011). Faktoren, deren Varianzerklärungsanteil niedriger ist als der einer Variablen, werden nicht berücksichtigt. Bortz (2013) weist drauf hin, dass besonders bei großen Variablenzahlen das Kaiser-Gutman-Kriterium dazu tendiert, die Anzahl der bedeutsamen Faktoren zu überschätzen, wodurch sich Probleme der Interpretierbarkeit der Faktoren ergeben können.

### Scree-Test

Der von Catell (1966) entwickelte Test basiert auf dem Eigenwertdiagramm (auch Scree-Plot), welches die Eigenwerte der Faktoren in abnehmender Höhe in Abhängigkeit ihrer Rangnummern abträgt und verbindet (Bortz, 2013). Ein Knick in der Konstanz des Eigenwertverlaufs signalisiert die Anzahl der zu extrahierenden Faktoren. Es ist die Stelle mit der größten Differenz zwischen den Eigenwerten (Voss, 2009). Die vor dem Knick liegenden Faktoren werden als bedeutsam betrachtet. Bei dem Scree-Test handelt es sich um eine subjektive Methode, bei

der die Ergebnisse von der Interpretation des Auswerters abhängen. Gorsuch (1983) kritisiert Interpretationsschwierigkeiten in Fällen nicht eindeutiger Lösungen, in denen kein Einbruch oder auch zwei Knicke im Eigenwertverlauf zu beobachten sind. Angewandt werden kann der Scree-Test sowohl auf den Eigenwertverlauf der empirischen als auch auf den der reduzierten Korrelationsmatrix. Fabrigar et al. (1999) betonen jedoch den engeren Bezug der reduzierten Korrelationsmatrix zum theoretischen Hintergrund des Common Factor Modells.

### **Parallelanalyse**

Die Parallelanalyse wurde von J.L.Horn (1965) entwickelt und besteht aus einem Vergleich des Eigenwertverlaufs der reduzierten Korrelationsmatrix (Fabrigar et al., 1999) mit dem Eigenwertverlauf der Korrelationen von normalverteilten Zufallsvariablen. Die Anzahl der Variablen und Fälle der Zufallsvariablen müssen dabei mit denen des Datensatzes übereinstimmen (Klopp, 2010). Graphisch wird die Faktorenanzahl über den Schnittpunkt der Eigenwertverläufe bestimmt. Die Faktoren, deren Eigenwerte über denen der zufällig generierten Faktoren liegen, werden berücksichtigt. Die beibehaltenen Faktoren sollen mindestens soviel Varianz aufklären, wie sich bei der Analyse von Zufallszahlen ergibt. Bortz (2013) vermerkt den rechnerischen Aufwand, der mit der Ermittlung der Zufallseigenwerte verbunden ist und bezieht sich auf regressionsanalytische Ansätze sowie auf Tabellen von Lautenschlager, die die Durchführung einer Parallelanalyse vereinfachen können.

Es existieren weitere Verfahren zur Bestimmung der Faktorenanzahl. Bühner (2011) stellt den Minimum-Average-Partial-Test sowie den Likelihood-Quotienten Test vor. Durch den Wissenschaftler könnte ebenfalls ein durch die Faktoren zu erreichender, erklärter Anteil der Gesamtvarianz festgelegt werden, sowie, falls eindeutige Erkenntnisse über die Struktur vorliegen, a priori eine zu extrahierende Faktorenanzahl. Da verschiedene Kriterien zu unterschiedlichen Resultaten führen, sollten in der Praxis immer mehrere Verfahren berücksichtigt werden. Nach der Bestimmung ist ein Test auf die Replizierbarkeit bzw. die Stabilität des Ergebnisses anhand von Stichproben der gleichen Grundgesamtheit möglich. Der Wissenschaftler steht vor der Entscheidung über den Grad der Vereinfachung. Eine Entscheidung zu Gunsten zu vieler Faktoren hat geringere Auswirkungen als eine für zu wenige Faktoren (Fava and Velicer, 1992). Im Zweifelsfall sollten daher eher zu viele als zu wenige Faktoren extrahiert werden.

## 2.5 Das Kommunalitätsproblem

Die Notwendigkeit der reduzierten Korrelationsmatrix und simultan das Kommunalitätsproblem ergibt sich durch das in Kapitel 2.2 beschriebene Faktoranalytische Modell. Es impliziert eine Unterscheidung der Gesamtvarianz in dem von den Faktoren erklärten Teil der Varianz, der Kommunalität  $h_j^2$  und einer Restkomponenten, der spezifischen Varianz bzw. Uniqueness  $\psi_{jj}$ . Die Kommunalität  $h_j^2 = \sum_{l=1}^k q_{jl}^2$ , definiert sich als die Summe der quadrierten Ladungen einer Variablen über alle  $l$  Faktoren, wobei das Quadrat der Ladung den gemeinsamen Varianzanteil zwischen Variable  $j$  und Faktor  $l$  bestimmt. Als Folge der Standardisierung liegt die Kommunalität im Wertebereich zwischen 0 und 1. Für sie ergibt sich meistens ein Wert kleiner 1, da die Gesamtvarianz der Variablen i.d.R. nicht vollständig durch die Faktoren reproduziert werden kann. Im Fall von  $h_j^2 = 1$ , wird eine Variable  $j$  restlos durch die Faktoren erklärt und die spezifische Varianz ist gleich 0 (Überla, 1968).

Ziel einer Faktorenanalyse ist die Erklärung der Korrelationen der manifesten Variablen, weshalb nur die Kommunalitäten im Interesse des Wissenschaftlers liegen. Da ihr endgültiger Wert erst im Zuge der Faktorenextraktion bestimmt wird, müssen sie vorerst geschätzt werden. Die spezifische Varianz ist eine Diagonalmatrix, die nur die diagonalen Elemente der Korrelationsmatrix beeinflusst. Die geschätzten Kommunalitäten ersetzen die Diagonalelemente von  $\Sigma$ . Aus  $\hat{\psi}_{jj} = 1 - h_j^2$  ergibt sich die **reduzierte Korrelationsmatrix**:

$$\Sigma' = \Sigma - \hat{\Psi} \quad (7)$$

Der Anwender hat an dieser Stelle eine Eingriffsmöglichkeit, indem er die Schätzmethode festlegt. Werden die Kommunalitäten zu hoch geschätzt, wird ein Anteil der spezifischen Varianz in die gemeinsame hineingezwängt und das Faktorenmuster verändert sich. Bei einer zu niedrigen Wahl geht gemeinsame Varianz für die Faktorenbestimmung verloren (Überla, 1968). Laut Backhaus et al. (2015) hat in den meisten Studien eine ungenaue Bestimmung der Kommunalitäten keine große Auswirkung, da es sich bei ihr nur um einen Startwert für die spätere Bestimmung in der Extraktion handelt. Je größer die Zahl an Variablen ist, desto unwichtiger sind exakt geschätzte Kommunalitäten, da ihr prozentualer Anteil als Element der Diagonalen abnimmt. Ansätze zur Schätzung der Kommunalität gibt es viele. Überla (1968) merkt an, dass allein Thurstone 12 verschiedene Lösungstechniken angibt. Im folgenden werden drei in der Praxis verbreitete Möglichkeiten zur Bestimmung der Anfangskommunalität  $\hat{h}_j^2$  vorgestellt.

Als Schätzung der Kommunalität  $\hat{h}_j^2$  kann der **höchste Korrelationskoeffizient von  $X_j$**  mit allen anderen Variablen verwendet werden. Es ist ein von Thurstone entwickeltes und bewährtes Verfahren, das durch seine Einfachheit überzeugt. Da der ergebende Wert keine besondere Beziehung zur Kommunalität aufweist und nur die Beziehung mit einer anderen Variablen berücksichtigt, eignet er sich lediglich als erste grobe Schätzung (Überla, 1968). Die Korrelation zweier Items stellt die Mindestschätzung der Reliabilität dar, berücksichtigt dabei aber zu wenige Informationen (Bühner, 2011).

Die **Einsen** können ebenfalls als **Anfangskommunalität** in der Diagonalen der Korrelationsmatrix beibehalten werden. Es ist ein Extremfall, der im Rahmen der Hauptkomponentenanalyse näher betrachtet wird (Backhaus et al., 2015). Es wird vom Anwender angenommen, dass die gesamte Varianz der Ausgangsvariablen erklärt werden soll. Da nicht mehr die gemeinsame, sondern die gesamte Varianz der Items analysiert wird, hat es jedoch Auswirkungen auf die Interpretation.

Das **Quadrat des multiplen Korrelationskoeffizienten von  $X_j$**  mit allen anderen Variablen (auch SMC = Squared Multiple Correlation) eignet sich ebenfalls als Wahl für einen Schätzwert der Kommunalität. Die Methode liefert eindeutige und für jede Korrelationsmatrix bestimmbare Werte. Es ist die empfehlenswerteste Methode, jedoch in der rechnerischen Bestimmung aufwendiger (Überla, 1968). Bühner (2011) bestätigt sie als beste Schätzung für Kommunalitäten, da sie der ursprünglichen Annahme entspricht, Korrelationen der Items durch die Faktoren erklären zu können. Da sie die meiste Information berücksichtigt, werde ich sie in meiner späteren Analyse verwenden.

## 2.6 Die Faktorenextraktion

Die Wahl des Faktorextraktionsverfahrens ist eng mit der Schätzung der Kommunalitäten verbunden (Backhaus et al., 2015). Für die meisten Verfahren ist eine vorherige Schätzung der Kommunalität nötig und sie setzen direkt an sie an. Im Zuge der Extraktion wird die Position der Faktorenachsen bestimmt. Es sind Annäherungen, deren Ziel es ist, ein bestmögliches Ergebnis der Schätzung von Faktorladungen  $\hat{Q}$  und der spezifischen Varianz  $\hat{\Psi}$  zu erzielen, um die empirische Korrelationsmatrix zu reproduzieren. Als geschätzte empirische Korrelationsmatrix ergibt sich  $S = \hat{Q}\hat{Q}^T + \hat{\Psi}$  (Härdle and Simar, 2007). In einem Faktorenmodell existiert keine eindeutige Lösung zur Reproduktion der reduzierten Korrelationsmatrix. Es existieren unendlich viele äquivalente Lösungen für  $\hat{Q}$ , die der Gleichung  $\hat{\Sigma}' = \hat{Q}\hat{Q}^T$  genügen. Ein eindeutiges Ergebnis lässt sich nur durch die Einführung von Restriktionen erreichen, die die verschiedenen Extraktionsmethoden implizieren (Überla, 1968). Am meisten verbreitet sind die Hauptachsen- und die Hauptkomponentenanalysen, die jedoch modelltheoretische Unterschiede aufweisen, die berücksichtigt werden müssen. Abgesehen von ihnen, werde ich im folgenden Teil die Grundgedanken der Maximum-Likelihood-Methode und der Alpha Faktorenanalyse darlegen.

### Hauptachsenanalyse

Die Hauptachsenanalyse bestimmt die Faktoren durch Zerlegung der reduzierten Korrelationsmatrix anhand einer Eigenwertdekomposition und anschließender Reproduktion, in deren Verlauf die Kommunalitäten und der spezifische Varianzanteil bestimmt werden (Härdle and Simar, 2007). Die reduzierte Korrelationsmatrix gilt als Ausgangspunkt, von dem aus die Ladungsmatrix  $\hat{Q}$  so geschätzt wird, dass sie nach Multiplikation mit ihrer transponierten möglichst genau der reduzierten Korrelationsmatrix entspricht. Da es keine eindeutige Lösung gibt, werden die Achsen auf eine Weise ermittelt, dass sie rechtwinklig zueinanderstehen und sukzessiv maximale Varianzanteile der Items aufklären. Durch die Methode der kleinsten Fehlerquadrate wird die Lage der Faktorenachsen ermittelt. Vergleichbar mit dem Vorgehen in einer Regression, berücksichtigt sie die Abweichung der empirischen Werte von den geschätzten Achsen. Die erste Achse erklärt die maximale Varianz, jede folgende steht orthogonal zu der vorherigen und erklärt einen niedrigeren Varianzanteil (Noack, 2007). Aus der Faktorenlösung können neue Kommunalitäten berechnet werden. Die modifizierten Werte werden in die Hauptdiagonale eingetragen und der Vorgang wird bis zur Konvergenz der Kommunalitäten wiederholt. Abgebrochen wird nach Evaluation der Differenz zweier aufeinanderfolgender

Kommunalitätenschätzungen (Bühner, 2011). In SPSS liegt eine Endschätzung bei einer Differenz von  $<0.001$  oder nach 25 Iterationen vor. Überla (1968) bezeichnet die Hauptachsenmethode als Routineverfahren in der Faktorenanalyse.

### **Hauptkomponentenanalyse**

Hauptkomponenten- und Hauptachsenanalyse unterscheiden sich nicht in ihrer Rechentechnik (Backhaus et al., 2015). Die Dekompositionstrategie der Eigenwertzerlegung wird jedoch auf unterschiedliche Ausgangsmatrizen angewandt (Überla, 1968). Sie unterscheiden sich in ihren theoretischen Modellen und folglich in der Interpretation ihrer Ergebnisse. Die Hauptkomponentenanalyse unterliegt der Annahme, dass die Faktoren die gesamte Varianz der Items vollständig erklären können (Klinke, 2017), d.h. die Existenz spezifischer Varianz wird dabei nicht berücksichtigt. Es werden keine Kommunalitäten für eine reduzierte Korrelationsmatrix geschätzt, sondern es wird die vollständige Korrelationsmatrix mit 1 als Startwert in der Diagonalen verwendet. Erst im Verlauf der Eigenwertzerlegung zur Bestimmung der Faktoren ergeben sich die Kommunalitäten (Härdle and Simar, 2007). Das Ziel ist ausschließlich die Reduktion von Daten (Klopp, 2010). Die Erklärung der Daten durch ein lineares Modell wird angestrebt, wobei keine Aussage über Kausalität notwendig ist (Noack, 2007). Die niedrigeren Kommunalitäten im Ergebnis der Hauptkomponentenanalyse sind als Informationsverlust zu interpretieren, ausgelöst durch eine niedrigere Faktorenanzahl als es Variablen gibt. Sie repräsentieren nicht reproduzierbare Varianzanteile, anstatt der Einzelrestvarianz anderer Extraktionsmethoden (Backhaus et al., 2015). Die Hauptkomponentenanalyse weist modellkonzeptionelle Unterschiede im Vergleich zu anderen faktorenanalytischen Extraktionsmethoden auf, sodass sie je nach Lehrbuch als eigenständiges Verfahren vorgestellt wird. Backhaus et al. (2015) betont, dass die Ergebnisse der Hauptkomponentenanalyse von denen der Hauptachsenanalyse abweichen. Für Überla (1968) beweist sich die Hauptkomponentenmethode als realitätsfernes deskriptives Instrument zur Beschreibung der Gesamtvarianz. Er sieht ihre Methodik jedoch als ein wichtiges Mittel zur Lösung des Faktorenproblems.

### **Maximum-Likelihood Faktorenanalyse**

Maximum Likelihood ist eine "Methode zur Schätzung der Parameter von ökonometrischen Modellen. Die Parameter der Schätzfunktion werden dabei so gewählt, dass sie die Wahrscheinlichkeit, genau die Beobachtungspunkte der vorliegenden Stichprobe zu erhalten" (Wirtschaftslexikon, 2017b), maximieren. Im Kontext der Faktorenanalyse findet die Maximum-

Likelihood-Lösung die wahrscheinlichsten Populationsparameter, die die empirische Korrelationsmatrix, unter der Annahme des faktorenanalytischen Modells, reproduzieren (Noack, 2007). Die Likelihood-Funktion wird bezüglich Kommunalität  $Q$  und spezifischer Varianz  $\Psi$  maximiert. Die Methode setzt eine multivariate Normalverteilung der Stichprobe voraus (Klinke, 2017) und verwendet die reduzierte Korrelationsmatrix als Ausgangspunkt. Als Startwert für die Schätzung der Kommunalitäten wird bei der Maximum-Likelihood Faktorenanalyse normalerweise der multiple Korrelationskoeffizient verwendet (Backhaus et al., 2015).

### Alpha-Faktorenanalyse

Der Reliabilitätskoeffizient  $\alpha$  schätzt beim ' $\alpha$ -factoring' die Reliabilität und die interne Konsistenz der aus allen Testitems gebildete Summenscores. Das Konzept der  $\alpha$ -Generalisierbarkeit bezieht sich auf die Grundgesamtheit eines jeden Items und berechnet sich durch das Quadrat der Korrelation, also dem gemeinsamen Varianzanteil einer beobachteten Variablen mit ihrer tatsächlichen Größe in der Grundgesamtheit (Überla, 1968). Das Ziel in der von Kaiser und Caffrey (1965) entwickelten Methode ist es, Faktoren mit möglichst hoher Generalisierbarkeit zu bestimmen. Extrahiert werden "Faktoren, die eine maximale Korrelation mit den entsprechenden Faktoren der Grundgesamtheit von Variablen aufweisen." (Überla, 1968, S.150). Gorsuch (1983) merkt an, dass es sich bei der Alpha-Faktorenanalyse eher um eine psychometrische und statistische Herangehensweise handelt.

Neben den bisher vorgestellten Extraktionsmethoden existieren alternative Ansätze. In der **Unweighed-Least-Square-Extraktion** werden die Faktoren mit dem Ziel der Minimierung der Differenz zwischen empirischer Korrelationsmatrix und reproduzierter reduzierter Korrelationsmatrix extrahiert (Klinke, 2017). **Minimum-Residual** ignoriert vorerst die Diagonalelemente der Korrelationsmatrix und minimiert die nicht-diagonal Elemente der Residualmatrix durch ein bestmöglichen Fit von reproduzierter zu empirischer Korrelation (Gorsuch, 1983). Information über Stärken oder Schwächen einzelner Methoden sind begrenzt aufzufinden, jedoch werden Hauptachsen- und Maximum-Likelihood-Extraktion verbreitet bevorzugt behandelt, oft abhängig von der Verteilung der Ausgangsdaten.

## 2.7 Die Faktorenrotation

Die Reproduktion der reduzierten Korrelationsmatrix im Zuge der Extraktion hat keine eindeutige Lösung für  $Q$ . Die resultierenden Faktoren sind rein abstrakte und auf mathematischen Kriterien basierende Größen, eine klare Struktur in ihnen ist nicht erkennbar und auch eine sinnvolle Interpretation nicht möglich (Darton, 1980).

Die Rotation der Achsen führt zu interpretierbaren Faktorladungen, ermöglicht durch eine Transformation der anfänglich erhaltenen Faktoren. Sie entspricht einer Drehung des Koordinatenkreuzes im Ursprung, wodurch sich die Position der Variablen in Bezug auf die Faktorenachsen verändert. Die Rotation führt dabei lediglich zu einer Umverteilung der Varianz auf den neuen Achsen, der Erklärungsgehalt der Faktoren bleibt im Zuge der Transformation unverändert (Backhaus et al., 2015). Bortz (2013) zeigt, dass sich nach einer Rotationstransformation "die Koordinaten ... auf den neuen Achsen als Linearkombination der ursprünglichen Koordinaten darstellen lassen". Die Gewichtungskoeffizienten entsprechen dabei den  $\cos$ -Winkel zwischen den ursprünglichen Achsen und den neuen (Bortz, 2013, S.399). Überla (1968) beschreibt den Rotationsprozess als Transformationsprozess, der die nach der Extraktion erhaltene, unrotierte Faktorenlösung  $\hat{Q}$  mit der Transformations- bzw. Rotationsmatrix  $G$  zur Ladungsmatrix  $\dot{Q}$  transformiert, wobei an die Transformation nach  $\dot{Q}$  gewisse Zielvorstellungen und Bedingungen gestellt sind:

$$\dot{Q} = \hat{Q}G, \quad \text{wobei} \quad \hat{\Sigma}' = \hat{Q}\hat{Q}^T = \dot{Q}\dot{Q}^T \quad (8)$$

Generell zu unterscheiden ist zwischen **orthogonalen und obliquen** Rotationsmethoden. Orthogonale und oblique Methoden unterscheiden sich zwar in ihrer Theorie, Umsetzung und Interpretation, erstreben im Zuge der Rotation nach  $\dot{Q}$  jedoch das gleiche Ziel, dem **Kriterium der Einfachstruktur** (Criterion of Simple Structure). Neben analytischen Rotationen gibt es graphische Ansätze, die ich an dieser Stelle nicht vertiefen werde, da es sich bei ihnen um subjektiv bestimmte Methoden handelt (Kaiser, 1958), die entwickelt wurden, als elektronische Rechner noch nicht verbreitet waren.

Nachdem in Kapitel 2.7.1 das "Simple Structure Kriterium" vorgestellt wird, werden orthogonale und oblique Rotationsmethoden im Detail betrachtet. Es stellt sich die Frage, welche Rotationsmethode in welcher Situation und unter welchen Voraussetzungen in einer Faktorenanalyse gewählt werden sollte.



### 2.7.1 Das Kriterium der Einfachstruktur

Thurstones Kriterium der Einfachstruktur, auch "Simple Structure Criteria", von 1947 hat die Entwicklung vieler Rotationsmethoden maßgeblich beeinflusst und sich als das Hauptkriterium der Faktorenanalyse etabliert. Mit seinem Prinzip löste er das Problem der Unbestimmbarkeit klarer Faktoren, indem er eine Zielsetzung an die Rotation extrahierter Faktoren stellte (Gorsuch, 1983). Thurstone rotierte die unrotierten Faktorladungen, um durch eine Neuverteilung der Varianz eine Einfachstruktur innerhalb der Ladungen zu erhalten. Er erzielte mit ihr eine "sparsamere" Faktorenlösung mit Faktorladungen, die sich besser interpretieren lassen (Kieffer, 1998). Er bewies, dass sich durch sein Prinzip, unabhängig von der Anzahl der beteiligten Variablen, eine klare Position der Faktoren identifizieren lässt. Gorsuch (1983) führt in Anlehnung an Thurstone (1947) die 5 Prinzipien der Faktorenrotation auf:

1. "Each variable should have at least one zero loading.
2. Each factor should have a set of linearly independent variables whose factor loadings are zero.
3. For every pair of factors, there should be several variables whose loadings are zero for one factor but not the other.
4. For every pair of factors, a large proportion of variables should have zero loadings on both factors whenever more than about four factors are extracted
5. For every pair of factors, there should only be a small number of variables with non-zero loadings on both."

(Gorsuch, 1983, S.177-S.178)

Die Rotation ist zulässig, da der bisher erklärte Varianzanteil, die Kommunalität, sich nicht verändert. Durch sie wird kein zusätzlicher Informationsgehalt aufgeklärt. Die Rotation umverteilt lediglich die zuvor erklärte gemeinsame Varianz von früher zu später extrahierten Faktoren, mit dem Ziel einer besseren Interpretation (Rinkenburger, 2011).

Neben dem Kriterium der Simple Structure gibt es weitere Prozedere, die sich jedoch nie durchgesetzt haben. Erwähnenswert sind der 'hyperplane count' (Catell, 1952), der eine gewisse Anzahl von Null-Werten in der Ladungsmatrix voraussetzt, 'proportional profiles' von Catell (1955) oder das 'criterion rotation' von Eysenck (1947) (Gorsuch, 1983). Das neueste Rotationskriterium stammt von Ertel (2016). Mit seinem "Complex Structure Modelling"

strebt er besonders komplexe Ladungsmustern an und entwickelte die Varimin-Rotation, die eine Transformation der extrahierten Faktoren zu einer Komplexstruktur ermöglicht. Fragwürdig ist bei ihr jedoch die Interpretation der komplex rotierten Faktorladungen (Bellmann, 2016). Das Kriterium der Einfachstruktur überzeugt durch seine Einfachheit, dennoch wären laut Gorsuch (1983) andere Prozedere ebenfalls sachgerecht, wenn die persönliche Philosophie der Wissenschaft das Kriterium unterstützen würde. Die von mir in 2.7.2 und 2.7.3 vorgestellten Rotationsmethoden erstreben alle das Ziel der Einfachstruktur.

### 2.7.2 Orthogonale Rotationen

In orthogonalen Rotationen bleibt die Unabhängigkeit zwischen den Faktoren erhalten. Die erhaltenen Faktoren korrelieren nach der Rotation untereinander nicht (Backhaus et al., 2015). Die Transformationsmatrix muss die Bedingung  $GG^T = I$  erfüllen, damit sie orthogonal ist. Während der Rotation ändert sich der Winkel zwischen den k Faktorachsen nicht und sie stehen nach der Rotation nach wie vor rechtwinklig zueinander (Überla, 1968). Nach Härdle and Simar (2007) ist im Fall von zwei Faktoren die Rotationsmatrix G gegeben als:

$$G(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \quad (9)$$

Sie repräsentiert eine Rotation der Koordinatenachsen im Winkel  $\theta$  im Uhrzeigersinn:  $\dot{Q} = QG(\theta)$ . Im Verlauf der orthogonalen Rotation kann eine Kaiser-(Horst)-Normalisierung vorgenommen werden (Klinke, 2017): Durch die Division der Ladungen mit der jeweiligen Kommunalität werden die Variablenvektoren auf eine einheitliche Länge gebracht und erhalten eine gleichmäßige Gewichtung:  $\tilde{q} = \hat{q}_{jl}/\hat{h}_j$  (Härdle and Simar, 2007). Das Ergebnis einer orthogonalen Rotation ist die Faktorladungsmatrix, in der sich die Ladungen der Items auf die Faktoren befinden. Die Ladungen repräsentieren die Korrelation zwischen Variable und Faktor, sowie die semipartiellen standardisierten Regressionsgewichte der Items auf den Faktoren (Bühner, 2011). Im Laufe der Zeit wurden viele orthogonale Rotationsmethoden entwickelt. Im Folgenden werden mit Varimax und Quartimax die am weitesten verbreiteten Prozedere näher betrachtet.

#### Quartimax (Neuhauser und Wrigley, 1954)

Das Vorgehen von Quartimax ist vergleichbar mit einer Vereinfachung der Ladungsstruktur jeder Variablen der Faktorladungsmatrix. Jede Variable soll auf einem Faktor eine hohe Ladung aufweisen. Sie maximiert die Varianz der quadrierten Ladungen in jeder Variable und

minimiert dabei die zur Erklärung benötigte Faktorenanzahl (Kaiser, 1958). Entwickelt wurde die Quartimax Rotation aus einem Ansatz mit einem Punkt und zwei orthogonalen Faktoren. Dieser wird am besten dargestellt, wenn ein Faktor genau durch den Punkt verläuft und der andere zu ihm orthogonal steht. In dem Fall der besten Lösung ist das Kreuzprodukt der quadrierten Ladungen gleich Null. Das Kreuzprodukt ist am größten, wenn eine Variable ähnlich hohe Ladungen auf beide Faktoren aufweist. Am niedrigsten wird es, wenn die Variablen mit einem Faktor hoch korreliert und niedrig mit dem anderen. Die Ladungen werden quadriert, um das Vorzeichen auszugleichen. Der Gedankengang wird übertragen auf ein Modell mit  $k$  Variablen und  $p$  Faktoren. Die Variabilität innerhalb einer Variablen ist am größten, wenn ein Element der quadrierten Faktorladungen gleich groß ist wie die Kommunalität und die restlichen Ladungen der Reihe gleich Null. Die Minimierung des Kreuzproduktes quadrierter Ladungen für eine gegebene Variable, maximiert ihre Ladung auf die kleinstmögliche Faktorenanzahl und minimiert die der restlichen. Da die Kommunalität, die sich durch die Summe jeder Zeile ergibt, weiterhin konstant ist und sich durch die Rotation nicht verändert, entspricht die Minimierung des Kreuzproduktes der Maximierung der quadrierten Kommunalität. Aus der "forth power maximation" ergab sich der Name Quartimax (Gorsuch, 1983; Überla, 1968; Darton, 1980). Die Rotationsmethode maximiert Gleichung 10:

$$Q = \sum_{l=1}^k \sum_{j=1}^p q_{jl}^4 \quad (10)$$

Die Quartimax Rotation neigt dazu, einen Generalfaktor zu produzieren (Kaiser, 1958) und zusätzliche Faktoren mit hauptsächlich kleineren Ladungen (Klinke, 2017). Die Quartimax-Rotation führt zu der angestrebten Einfachstruktur, jedoch können aufgrund des Generalfaktors Probleme bei der Interpretation der Faktoren entstehen (Kieffer, 1998). Kaiser (1958) kritisierte, dass eher die Varianz innerhalb eines Faktors maximiert werden sollte, als die Varianz innerhalb der Ladungen einer Variablen, woraufhin er die Varimax-Rotation entwickelte.

### **Varimax (Kaiser, 1958)**

Die Varimax-Rotation wurde 1958 von Henry F. Kaiser entwickelt und strebt eine Vereinfachung der Spalten an. Angelehnt an der laut Kaiser verzerrten Quartimax-Rotation, entwickelte er eine Methode deren Skaleninvarianz und Stabilität er in den Vordergrund stellte. Er versuchte das Problem der Interpretierbarkeit von Faktorladungen mittlerer Höhe zu lösen (Kaiser, 1958). Das Ziel waren Faktoren mit einer niedrigen Anzahl von hohen Ladungen und einer hohen Anzahl niedriger, was einer Maximierung der Ladungsvarianz innerhalb eines jeden Faktors entspricht (Bortz, 2013). Die Methode sucht den Winkel  $\theta$  zur Rotation der

Achsen auf die Position, die die Varianz der quadrierten Ladungen  $\hat{q}_{ij}^2$  eines jeden Faktors, maximiert. Maximiert wird V, die Summe der Ladungs-Varianzen über alle Faktoren, wobei zuvor eine Normalisierung mit  $\tilde{q} = \hat{q}_{jl}/\hat{h}_j$  durchgeführt wird.

$$V = \frac{1}{p} \sum_{l=1}^k \left[ \sum_{j=1}^p (\tilde{q}_{jl})^2 - \left\{ \frac{1}{p} \sum_{j=1}^p (\tilde{q}_{jl})^2 \right\}^2 \right] \quad (11)$$

Varimax unterliegt dem Kriterium der Einfachstruktur und ist die wichtigste und am weitesten verbreitete orthogonale Rotationsmethode. Generalfaktoren sind bei ihr ausgeschlossen. Gorsuch (1983) weist drauf hin, dass die Varimax-Rotation jedoch nicht angemessen ist, sollte den Items ein Generalfaktor unterliegen. In dem Fall führt die Varimax-Rotation zu verfälschten Ergebnissen. Beobachtungen mit hoher interner Konsistenz sollten mit anderen Verfahren rotiert werden. Kieffer (1998) suggeriert, die Rotationsmethode anhand der an die Analyse gestellten Erwartungen zu wählen.

Quartimax und Varimax sind die Grundmodelle orthogonaler Rotationen. **Equimax** (Saunders, 1963), **Biquartimax** (Carroll, 1957) und **Orthomax** (Harman, 1960) sind weitere orthogonale Rotationen, die versuchen, einen Kompromiss zwischen ihnen zu finden. Während Biquartimax beide Methoden gleichmäßig gewichtet, ist es in Equimax abhängig von der extrahierten Faktorenanzahl. In Orthomax bestimmt der Anwender die Gewichtung zwischen einer Vereinfachung der Spalten und der Vereinfachung der Zeilen. Varimax ist unter allen die meist verbreitetste Rotation und wird z.B von Gorsuch (1983) präferiert und als Standardmethode deklariert. Überla (1968) merkt an, dass orthogonale Rotationen Daten nur im Sinn der Einfachstruktur approximieren können, wenn die Vektoren tatsächlich orthogonal sind.

### 2.7.3 Oblique Rotationen

Die Orthogonalität zwischen den rotierten Faktorenachsen wird bei obliquen Rotationen aufgehoben und damit werden Korrelationen zwischen den Faktoren zugelassen. Im Verlauf der obliquen Rotation werden zwei Matrizen bestimmt, deren Struktur untersucht wird; unterschieden wird zwischen Faktormuster- ( $Q$ ) und Faktorstrukturmatrix ( $Q_S$ ). Zu Beginn, in Gleichung 5, wurde die empirische Kovarianz mit  $Cov(X) = QR_{FF}Q^T + \Psi$  abgeleitet, in der  $R_{FF} = C$  die Korrelationen zwischen den Faktoren verkörperte. Für die vorerst unkorrelierten Faktoren wurde  $C$  zur Einheitsmatrix ( $I$ ), weshalb bei orthogonalen Rotationen Faktormuster und Faktorenstruktur identisch sind. Im jetzt betrachteten Fall obliquen, schiefwinkliger Faktoren nimmt  $C$  die Korrelationsmatrix der Faktoren an. Die Faktorenstruktur ergibt sich aus  $Q_S = QC$  und enthält nun die Korrelationen der Variablen mit den obliquen Faktoren. Das Faktormuster  $Q$  enthält Ladungen, die innerhalb jeder Zeile als Regressionskoeffizienten zur Erklärung der standardisierten Variable, in Abhängigkeit der Faktoren, zu interpretieren sind. Sowohl Faktormuster als auch Faktorenstruktur müssen in der Interpretation berücksichtigt werden (Gorsuch, 1983). Wie bei den orthogonalen Rotationen ist das Ziel der obliquen Rotation eine Einfachstruktur.

#### Direct Oblimin (Jennrich & Simpson, 1966)

Direct Oblimin resultierte aus den Methoden Quartimin und Kovarimin und dem Versuch, Prinzipien aus orthogonalen Rotationen auf oblique anzuwenden. In Quartimin wird die Summe der Kreuzprodukte der quadrierten Variablenladungen minimiert. Das von Carroll 1957 entwickelte Verfahren führt zu hoch korrelierten Faktoren. Kovarimin (von Kaiser 1958) dagegen minimiert die Kovarianzen zwischen den Faktoren, was zu orthogonalen Vektoren führt. Direct Oblimin minimiert Funktion 12 unter der an die Transformationsmatrix  $G$  gestellte Bedingung  $diag(GG^T) = I$  (Darton, 1980).

$$D = \sum_{l \neq m}^k \left( \sum_{j=1}^p q_{jl}^2 q_{jm}^2 - \frac{\delta}{p} \sum_{j=1}^p q_{jl}^2 \sum_{j=1}^p q_{jm}^2 \right) \quad (12)$$

Der Anwender hat die Möglichkeit, den Grad der Gewichtungen beider Methoden und damit Korrelation der resultierenden Faktoren durch den Parameter  $\delta$  zu bestimmen. Je niedriger  $\delta$ , desto orthogonalere und unkorrelierte Faktoren resultieren. Für  $0 < \delta < 1$  ergeben sich hohe positiv korrelierte Faktoren (Jennrich and Sampson, 1966; Gorsuch, 1983).

Kieffer (1998) rät bei der Durchführung von der Direct Oblimin Methode, trotz ihrer Nähe zur Realität, zu Vorsicht. Er kritisiert den subjektiven Eingriff des Wissenschaftlers in der Methode und betont die möglichen Auswirkungen einer falschen Wahl.

### **Promax (Hendricksen & White, 1964)**

Promax nimmt eine orthogonale Lösung als Basis, die durch eine lineare Transformation zu einer idealen Lösung gebracht wird. Mathematisch geschieht dies durch Potenzieren der Ladungen, normalerweise mit  $k = 2, 4$  oder  $6$ . Alle Ladungen werden niedriger, die absolute Differenz zwischen hohen und niedrigen Ladungen erhöht sich jedoch, da sich die Ladungen in mittlerer und niedrigerer Höhe stärker gegen  $0$  bewegen. Eine höhere Potenz führt zu stärkeren Korrelationen innerhalb der Faktoren. Die optimale Potenz ist jene, die die "simplest structure" mit der geringsten Korrelation aufweist. Die Ausgangsmatrix wird nun mit Hilfe einer "Procrustean"-Rotation (Kieffer, 1998) und des "best least square fits" so nah wie möglich an die Zielmatrix gebracht (Klinke, 2017). Promax ist schnell berechnet und liefert qualitativ hochwertige Ergebnisse mit einer guten Einfachstruktur im Vergleich zu anderen Methoden. Die Ergebnisse entsprechen dem Kriterium der Einfachstruktur in der Regel besser als die orthogonale Ausgangslösung, ausgelöst durch die Lockerung der Winkel zwischen den Faktoren (Gorsuch, 1983).

Neben Faktormuster und Faktorstruktur ist auch die Korrelationsmatrix der Faktoren im Ergebnis einer obliquen Rotation enthalten. Der Grad der Korrelation zwischen Faktoren kann vom Anwender in einigen Methoden beeinflusst werden, wird aber in der Regel so gering wie möglich gehalten. Die erwartete Korrelation zwischen den Faktoren beträgt  $> 0.15$  (Bortz, 2013). Sollte inhaltlich eine hohe Korrelation zwischen zwei Faktoren vorliegen, wird die Anzahl der Faktoren reduziert, da die von ihnen erklärte Information redundant ist. Mit z.B. Maxplane, Direct Oblimin und Biquartimin gibt es viele weitere oblique Rotationsmethoden sowie Abwandlungen der hier vorgestellten Vorfahren. Anders als bei den orthogonalen Rotation gibt es innerhalb der obliquen keine, die in der Wissenschaft bevorzugt wird.

#### 2.7.4 Diskussion zum Rotationsproblem

In der Literatur rund um die Faktorenanalyse werden bezüglich der Rotationsentscheidung verschiedene Meinungen vertreten. Sie wird als Entscheidung angesehen, die auf keinen Fall willkürlich getroffen werden sollte.

Für oblique Rotationen argumentiert Iacobucci (2001) mit einer gültigeren, realitätsgetreueren Repräsentation des untersuchten Sachverhalts. Er bezeichnet die Unabhängigkeit der Faktoren in orthogonalen Rotationen als eine willkürliche Beschränkung. Die Nichtberücksichtigung existenter Korrelationen zwischen zwei Faktoren ist ein Fehler in der Konzeptualisierung und führt zu Informationsverlust (Costello and Osborne, 2005). Fabrigar et al. (1999) betonen, dass die Wahrscheinlichkeit, eine gut interpretierbare Einfachstruktur zu erhalten, bei obliquen Rotationen höher ist. Wenn die einem Datensatz unterliegende Struktur auf miteinander korrelierenden Faktoren basiert, kann eine orthogonale Rotation keine optimalen Ergebnisse liefern. Ebenso Kenntnisse sind über die Korrelationen der Faktoren wertvoll für die weitere Verwendung der Ergebnisse.

Eine Präferenz zu obliquen Rotationen aufgrund der realitätsfernen theoretischen Annahme der Orthogonalität ist weit verbreitet, jedoch sprechen auch Gründe für die orthogonalen Rotationsmethoden. Bortz (2013) behandelt vorzugsweise orthogonale Rotationen, da oblique Rotationen in Faktoren resultieren, die durch ihre Abhängigkeit redundante Informationen beinhalten. Die Funktion der Datenreduktion durch Faktorenanalysen wird dabei aufgegeben. Orthogonale Rotationen sind durch weniger Rechenaufwand einfacher zu konzeptualisieren. Angenommen, es existiere ein Modell, in dem 100 Variablen einer fünf Faktorenstruktur unterlägen, müssten allein durch die Entscheidung für eine oblique Rotationsmethode 550 Parameter mehr geschätzt werden, als für eine orthogonale. Neben den Pattern-Elementen müssten noch die Struktur-Elemente bestimmt werden, außerdem jeweils die Korrelationen zwischen den einzelnen Faktoren beider Matrizen (Kieffer, 1998). Die weitere Verwendung des Ergebnisses der Faktorenanalyse hat ebenso Einfluss auf die Wahl zwischen orthogonaler und obliquen Rotation. Bei Faktoren, deren Faktorwerte als unabhängige Prädiktoren in eine multiple Regressionsanalyse eingehen sollen, würde eine orthogonale Rotation späteren Problemen der Multikollinearität vorbeugen und die Interpretation der betas vereinfachen (Iacobucci, 2001). Ein ähnlicher Fall liegt vor, wenn auf Grundlage der Faktorenanalyse möglichst voneinander unabhängige Skalen konstruiert werden sollen (Klopp, 2010).

Während die Varimax-Rotation von nahezu allen Wissenschaftlern innerhalb der orthogonalen Rotation präferiert wird, existiert für die oblique Rotation keine eindeutige Präferenz (Überla, 1968). Nach Kieffer (1998) ist die Entscheidung zwischen einer orthogonalen und einer obliquen Rotation lediglich abhängig von der Konzeption des Wissenschaftlers gegenüber dem untersuchten Konstrukt. Die Auswahl der Rotationsmethode ist eine, von den intrinsischen Werten des Wissenschaftlers geprägte, an das Ziel der Analyse gestellte Entscheidung. Iacobucci (2001) empfiehlt, mit der Annahme von korrelierten Faktoren, von Anfang an eine oblique Rotation durchzuführen. Sollten sich die Interkorrelationen der Faktoren in der weiteren Analyse paarweise als nicht signifikant herausstellen ( $< |0,3|$ ), kann anschließend ein orthogonales Rotationsverfahren angewandt werden (Klinke, 2017). Hohe Korrelationen zwischen Faktoren, ab  $|0,7|$ , sind ein Indiz für eine zu hohe Faktorenanzahl. Auch Härdle and Simar (2007) betonen die Gleichwertigkeit verschiedener Methoden in der Praxis und sprechen sich für einen Vergleich verschiedener Rotationsmethoden aus. Sie empfehlen Cross-Validations, mit der Unterteilung der Daten in Test- und Trainingsets. Die Ergebnisse der Faktor-Scores, -Ladungen und Kommunalitäten sollten jeweils gleich ausfallen. Kieffer (1998) fokussiert sich auf die Betrachtung der Faktorladungen nach der Rotation. Ist ein großer Unterschied zwischen der orthogonalen und der obliquen Lösung beobachtbar, sollte die oblique Rotation interpretiert werden, in Fällen, in denen sie sich entsprechen, die orthogonalen.

## 2.8 Interpretation der Faktorladungen

Das Ergebnis der Faktorenrotation sind die Faktorladungen, die als Grundlage zur Interpretation dienen.

Durch Kenntnis der Faktorladungen  $Q$  und der spezifischen Varianz  $\Psi$  lässt sich Gleichung 2 des Common Factor Models auflösen und die Faktorwerte  $F$  lassen sich berechnen. Die Faktorwerte beschreiben die Ausprägungen der Beobachtungsobjekte auf die ermittelten Faktoren (Backhaus et al., 2015). Die Ermittlung geschieht in der Regel anhand eines Regressionsmodells. Es existieren verschiedene Methoden für deren Ermittlung, welche hier unberücksichtigt bleiben da sie für die Identifizierung und die Interpretation der Faktoren, auf die sich diese Arbeit fokussiert, nicht von Bedeutung sind.



Nach einer obliquen Rotation sind sowohl Muster- als auch Strukturmatrix zu interpretieren. Während die Strukturmatrix die Korrelationen zwischen Variable und Faktor anzeigt, besteht die Mustermatrix aus den semipartiellen standardisierten Regressionsgewichten der Items auf den Faktoren. Für Ergebnisse orthogonaler Rotationen sind beide Matritzen identisch (Bühner, 2011). Eine Interpretation ist in allen Fällen nur mit fundierten Kenntnissen in dem untersuchten Bereich möglich. Andere mögliche Einflüsse auf hohe Faktorladungen müssen ebenfalls berücksichtigt werden. Nach Gorsuch (1983) kann selbst die Art, nach der Individuen für die Analyse ausgewählt wurden, Variablen beeinflussen und einen Faktor generieren. Die Interpretation der Faktorladungen lässt dem Anwender Spielraum für die Entscheidung, ab welcher Ladungshöhe eine Variable einem Faktor zugeordnet wird. Bühner (2011) berücksichtigt Faktorladungen  $> |0.3|$ . Bortz (2013) gibt alternative Richtlinien zur Faktoreninterpretation an. Nach ihm kann ein Faktor bei Ladungen  $> |0.6|$  von mindestens vier Variablen interpretiert werden, sowie bei zehn Variablen mit Ladungen  $> |0.4|$ . Sollten die Ladungen von weniger als zehn Variablen  $|0.4|$  überschreiten, sollte die Stichprobe mindestens 300 Beobachtungen umfassen. Im Fall von  $n < 300$  kann von zufälligen Ladungsstrukturen ausgegangen werden. Bei Ladungen einer Variablen auf mehrere Faktoren in signifikanter Höhe, bleibt dem Wissenschaftler die Möglichkeit offen, sie in beiden Faktoren zu berücksichtigen, sie auszuschließen oder sich in Abhängigkeit der Ladungshöhen auf die höhere zu beschränken. Backhaus et al. (2015) suggerieren die Berücksichtigung von mehrfach ladenden Items erst ab Ladungen  $> |0.5|$  auf jedem dieser Faktoren. Es gilt jedoch anzumerken, dass eindeutige Regeln nicht existieren und je nach Wissenschaftler bzw. Lehrbuch andere Werte angegeben sind. Essentiell für die Identifikation und die Benennung der Faktoren sind außerdem die Vorzeichen der Faktorladungen. Sie müssen in Betracht gezogen werden, um den Bezug der Variablen zum Faktor eindeutig zu bestimmen.

Das Kapitel hat die Faktorenanalyse in der Theorie veranschaulicht und ihren Aufbau demonstriert. Mit ihren vielen möglichen Verläufen in den jeweiligen Teilschritten der Analyse, hat der Wissenschaftler sowohl Spielraum im Modellaufbau als auch in der Interpretation der Ergebnisse. Die Entscheidung über den sachgerechten Aufbau des Faktorenmodells konnte selbst durch die bisherige Forschung nicht eindeutig festgelegt werden und basiert auf einem hohen Grad Subjektivität. Es stellt sich die Frage, ob sich signifikante Unterschiede in den Ergebnissen verschieden modellierter Faktorenanalysen, angewandt an einem bestimmten Datensatz, nachweisen lassen können.

## 3 Implementierung der Problemstellung in interaktiver Applikation

### 3.1 Zielsetzung

Anhand verschiedener Datensätze zielt die Applikation darauf ab, die zuvor betrachtete Fragestellung über der Modellierung von Faktorenmodellen aufzugreifen und zu visualisieren. Durch subjektive Entscheidungen im Verlauf der Analyse bezüglich u.a. Faktorenanzahl, Extraktions- und Rotationsmethoden ist eine Vielzahl sich unterscheidenden Faktorenmodellen möglich. Das Ziel bestand in der Erstellung eines Instrumentes zum Vergleich verschiedenen konstruierter Faktorenmodelle sowie für die Demonstration der Auswirkung ausgewählter Parameter auf das Ergebnis. Die Interaktivität der Anwendung sollte sich in der Hinsicht zeigen, dass sich der User während des Gebrauchs in einer aktiven Rolle befindet, interagiert und Einfluss auf die Auswahl variabler Parameter im Verlauf der Faktorenanalyse ausüben kann. Die Entscheidungen werden unverzüglich umgesetzt und deren Auswirkungen können direkt nachvollzogen werden. Im Fokus stand die Auswahl von Rotations- und Extraktionsmethoden sowie der Faktorenanzahl und deren Wirkung auf Kommunalitäten, Faktorladungen, Faktorkorrelationen und Eigenvalues. Des Weiteren sollte die Applikation den Ablauf einer Faktorenanalyse mit ihren einzelnen Schritten verfolgen, von der Korrelationsmatrix über die Analyse des Scree-Plots bis zur Interpretation der Faktorladungen.

### 3.2 R & Shiny

Die Analysen dieser Arbeit wurden mit R durchgeführt und die Applikation mit Shiny programmiert. Es folgt eine kurze Einführung in R sowie in das Shiny Paket, von dem ich Gebrauch gemacht habe.

**R** ist eine Programmiersprache und Programmierumgebung für statistische Berechnungen und die Erzeugung von Graphiken. Neben einer Vielzahl von statistischen Methoden stellt es auch graphische Techniken zur Verfügung. R ist kostenfrei erhältlich und läuft auf einer Vielzahl von UNIX Plattformen, Windows und MacOS. Anders als eine rein statistische und unflexible Sammlung von Werkzeugen, wie viele andere Software zur Datenanalyse, ist R eher eine dynamische und kohärente Programmierumgebung in die statistische Techniken implementiert sind. Als ein Open Source Programm haben User die Möglichkeit durch die Definition neuer

Funktionen R in seiner Funktionalität zu weiterzuentwickeln. Ebenso kann es durch online abrufbare Pakete erweitert werden (R Core Team, 2017). **Shiny** ist ein Erweiterungspaket von R, in dessen Rahmen es möglich ist, interaktive Web Applikationen oder Dashboards aus der R Umgebung zu programmieren. Es erlaubt, eine Analyse und ihre Ergebnisse in eine interaktive Anwendung zu implementieren. Für Shiny Applikationen sind keine Kenntnisse in HTML, CSS oder JavaScript erforderlich, sie lassen sich jedoch trotzdem mit CSS, HTML und JavaScript erweitern (Chang et al., 2017; RStudio, Inc, 2013).

### 3.3 Umsetzung und Ergebnis

#### Aufbau

Die Applikation besteht aus zwei Teilen **'Factor Analysis'** und **'Model Comparison'**. Die Navigationsleiste am oberen Bildschirmrand ermöglicht den Zugang zu den beiden Hauptoberflächen der Anwendung. Um sie so nutzerfreundlich wie möglich zu gestalten, ist jeder Teil aufgeteilt in zwei Teilbereiche, dem Sidebarpanel und dem Mainpanel. Der Sidebarpanel kontrolliert die Inputvariablen und ist für den Benutzer durchgehend sichtbar. Der Mainpanel ist wiederum unterteilt in eine Reihe von Tabs, die die einzelnen Schritte der Faktorenanalyse representieren bzw. verschiedene Ergebnisse ausgeben. Mainpanel befindet sich im Zentrum und erfährt die größte Aufmerksamkeit des Betrachters. Ausgegeben werden Graphen, Tabellen oder R Outputs in Abhängigkeit der Inputparameter.

#### Datensätze

Es besteht eine Wahl zwischen 2 Datensätzen: **'bfi'** und **'happy'**. Die Auswahl des Datensatzes erfolgt durch ein Dropdown Menu im unteren Bereich des Sidebarpanels. Der **'bfi'** Datensatz beschreibt einen Persönlichkeits Self-Report mit 25 Items von 2799 Befragungsubjekten. Erhoben wurden die Daten als Teil des "Synthetic Aperture Personality Assessment (SAPA)" einem Online-Personal-Selbst-Beurteilungsprojekt. Eine detailliertere Analyse des Datensatzes folgt in Kapitel 4. Zusätzlich steht mit **'happy'** ein, aus dem World-Happiness-Report erstellter Datensatz mit 155 Beobachtungen zur Verfügung. Er stammt aus Kaggle und beinhaltet Scores für sieben, auf die Happiness eines Landes einfließende Variablen.

## Ausführung

Das erste Mainpanel, **'Factor Analysis'**, visualisiert und folgt dem Verlauf einer Faktorenanalyse. Als Startseite wird dem Benutzer eine kurze Beschreibung des ausgewählten Datensatzes angezeigt, sowie eine Übersicht mit den ersten zehn Zeilen. Optional können nach Belieben weitere Zeilen angezeigt werden. Der Benutzer kann auf dem Sidebarpanel die Anzahl der relevanten Faktoren wählen, die im Verlauf der Faktorenanalyse extrahiert werden sollen. Die Auswahl wird durch einen Slider betätigt, der neben einer minimalen Anzahl auch ein Maximum sowie ein Defaultwert besitzt. Die Extraktionsmethode wird durch ein Dropdown Menu ausgewählt, welches dem Nutzer eine gute Übersicht der möglichen Extraktionsmethoden liefert. Auszuwählen ist zwischen Principle-Axis, Alpha-Factoring, Maximum-Likelihood, Minres und ULS. Im letzten Schritt der Wahl der Rotationsmethode, stehen dem Benutzer mit Varimax, Quartimax, Promax, Oblimin und Varimin eine Reihe von Rotationen zur Auswahl. Eine Anzeige der unrotierten Lösung ist ebenfalls möglich. Die Auswahl im Interface erfolgt ebenfalls durch ein Dropdown Menu. Nach der angezeigten Heatmap der Korrelationsmatrix als Ausgangspunkt der Faktorenanalyse wird im nächsten Tab **'Number of Factors'** ein Scree Plot abgebildet. Zusätzliche Informationen liefern die Eigenwerte der empirischen sowie der reduzierten Korrelationsmatrix, die neben der Parallelanalyse und dem Scree-Test als Grundlage der Entscheidung über die Anzahl extrahierter Faktoren dienen können. Der folgende Tab **'Factor Loadings'** zeigt die Faktorladungen. Der Nutzer kann anhand eines Sliders das Cut-off Level zur Auswahl eines unteren Grenzwerts für die angezeigten Faktorladungen bestimmen. Der letzte Tab **'Additional Information'** zeigt zusätzliche Kennzahlen zur ausgeführten Faktorenanalyse an. Neben der Summe der Kommunalitäten und dem Anteil der vom Faktorenmodell erklärten Varianz wird im Falle obliquer Rotationen die Korrelationsmatrix der resultierenden Faktoren ausgegeben.

Ein Zugang zum zweiten Teil der Applikation, dem Vergleich verschiedener Faktorenmodelle, wird dem Benutzer durch die Auswahl "Model Comparison" in der oberen Navigationsleiste ermöglicht. Im Sidebarpanel müssen nun zwei verschiedene Modelle zum Vergleich ausgewählt werden. Die zur Auswahl stehenden Parameter sind wie zuvor jeweils die Faktorenanzahl, die Extraktionsmethode sowie die Rotationsmethode. Im ersten Tab werden anhand von Ladungsdiagrammen die ausgewählten Modelle einander zunächst visuell gegenübergestellt. Die Diagramme beinhalten sowohl die Ladungen der Variablen auf die Faktoren als auch die Korrelationen der Faktoren untereinander. Die Lage der einzelnen Faktorachsen kann modell-

übergreifend im Tab **'Factor Congruence'** verglichen werden. Die Kongruenz Koeffizienten werden in einer Matrix ausgegeben und durch eine Heatmap visualisiert. Eine ausführlichere Erklärung der Faktor Kongruenz erfolgt in Kapitel 4.3. Während der Betrachtung der Ergebnisse besteht immer die Möglichkeit einer Anpassung der Inputvariablen.

Die von mir geschaffene Applikation unterstützt den in Kapitel 4 durchgeführten Vergleich von Faktorenmodellen anhand von Persönlichkeitsitems.

## 4 Ein Vergleich von Faktorenmodellen anhand von Persönlichkeitsitems

Das folgende Kapitel dokumentiert den Verlauf einer Faktorenanalyse und die Durchführung ihrer einzelnen Schritte. Verschieden aufgebaute Modelle werde ich einander gegenüberstellen und ihre Ergebnisse miteinander vergleichen. Werden die Ergebnisse signifikante Unterschiede aufweisen und lassen diese sich auf einen bestimmten Parameter zurückführen? Der Fokus im Vergleich wird auf die Rotationsmethode gelegt, jedoch werden auch Faktorenanzahl und Extraktionsmethode berücksichtigt.

### 4.1 Datensatz, Vorbereitung & Deskriptive Analyse

Der für den Test verschiedener Faktorenmodelle ausgewählte Datensatz, ist Teil des "Synthetic Aperture Personality Assessment" (SAPA), einem online-basierten Persönlichkeitsforschungsprojekt (Revelle, 2017). Die verwendeten Items stammen aus dem "International Personality Item Pool" (Goldberg et al., 2006). Der verwendete Datensatz umfasst 2799 Beobachtungen eines Fragebogens mit 25 Fragen zu Persönlichkeitsausprägungen der Subjekte. Zusätzlich wurden mit Alter, Geschlecht und Bildungsstand 3 zusätzliche Demographische Variablen aufgenommen. Die Antwortmöglichkeiten sind durch eine 6 stufige Likert-Skala skaliert, mit:

*1 = Very Inaccurate, 2 = Moderately Inaccurate, 3 = Slightly Inaccurate,  
4 = Slightly Accurate, 5 = Moderately Accurate, 6 = Very Accurate.*

Es folgen die 25 Variablen und ihre verwendeten Abkürzungen:

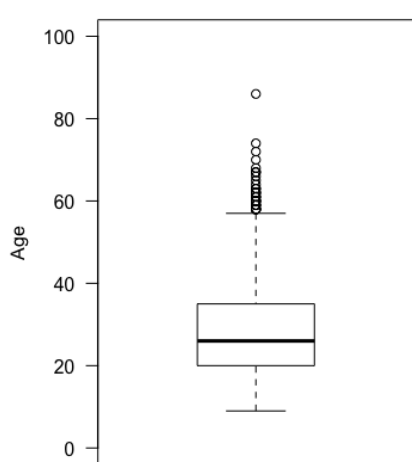
- *A1: Am indifferent to the feelings of others.*
- *A2: Inquire about others' well-being.*
- *A3: Know how to comfort others.*
- *A4: Love children.*
- *A5: Make people feel at ease.*
- *C1: Am exacting in my work.*
- *C2: Continue until everything is perfect.*

- *C3: Do things according to a plan.*
- *C4: Do things in a half-way manner.*
- *C5: Waste my time.*
- *E1: Don't talk a lot.*
- *E2: Find it difficult to approach others.*
- *E3: Know how to captivate people.*
- *E4: Make friends easily.*
- *E5: Take charge.*
- *N1: Get angry easily.*
- *N2: Get irritated easily.*
- *N3: Have frequent mood swings.*
- *N4: Often feel blue.*
- *N5: Panic easily.*
- *O1: Am Full of ideas.*
- *O2: Avoid difficult reading material.*
- *O3: Carry the conversation to a higher level.*
- *O4: Spend time reflecting on things.*
- *O5: Will not probe deeply into a subject.*

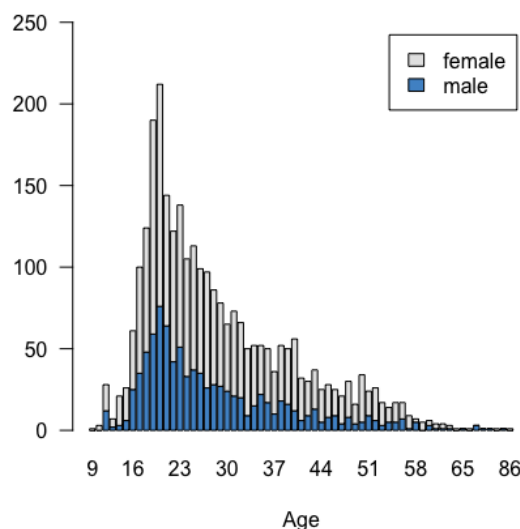
Den Fragen des IPIP unterliegt eine inhaltliche Struktur. Jede der insgesamt 3320 Variablen aus dem Item Pool ist ein Konstrukt dreier unterschiedlicher Konstruktgruppen zugeordnet, die als Giant3, Big6 und Little12 bezeichnet werden. In den 25 Ausgewählten Variablen ist nach dieser Zuordnung ein 3, 5 oder 8 Faktorenmodell möglich. Fraglich ist, ob sich diese Struktur durch eine Faktorenanalyse identifizieren lässt und welche Rolle generell die Auswahl der Extraktions- bzw. der Rotationsmethode spielt.

## Deskriptive Analyse: Demographische Daten der befragten Population

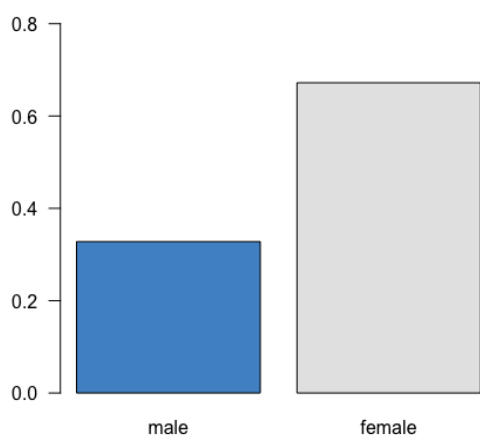
Befragt wurden 2799 Personen zwischen 9 und 68 Jahren. Das arithmetische Mittel des Alters beläuft sich auf 28.79 Jahre, während 50% zwischen 20 und 35 Jahre alt sind und 84,2% jünger als 40. Von den Befragten sind 32,8% Männer und 67,2% Frauen. Der Bildungsstand der befragten Personen wurde in Anlehnung an den akademischen Werdegang in 5 Klassen eingeteilt. 8.7% besuchen momentan die Highschool, 11,3% haben einen Highschoolabschluss, 48.5% studieren am College, 15.3% sind Hochschulabsolventen, (=College Graduates) und 16.2% aller Befragten sind Teil eines Graduiertenprogrammes (=Graduate Degree).



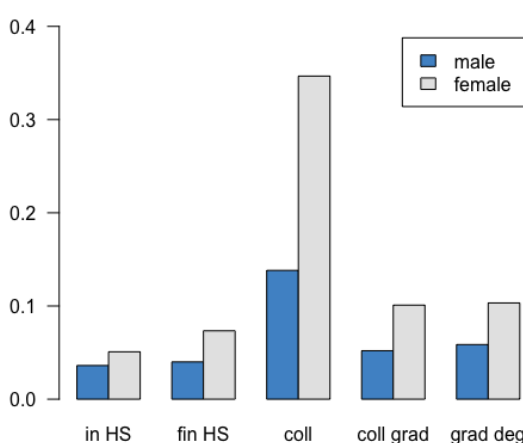
(a) Boxplot des Alters



(b) Absolute Häufigkeit des Alters in unterteilt in Geschlecht



(c) Geschlecht der Beobachtungsobjekte



(d) Barplot zum Bildungsstand unterteilt nach Geschlecht

**Abbildung 1:** Graphen zur Demographie



## Cleaning und Modifizierung des Datensatzes

Vorbereitend zur Faktorenanalyse habe ich die demographischen Variablen entfernt. Außerdem wurde eine Beobachtung entnommen, in der Inkonsistenzen zwischen Alter und Bildungstand aufgetreten sind. Anschließend erfolgte eine z-Standardisierung aller verbleibenden Variablen durch Subtraktion des arithmetischen Mittels jeder Variablen und folgender Division mit ihren Standardabweichungen.

## Die Korrelationsmatrix und ihre Eignung

Im Anschluss wurde die Korrelationsmatrix der standardisierten Variablen gebildet. Sie gilt als Ausgangssituation der Faktorenanalyse. Durch den listenweisen Fallausschluss wurden nur vollständige Beobachtungen berücksichtigt. Die Struktur der Korrelationsmatrix wird in Abbildung 2 verdeutlicht.

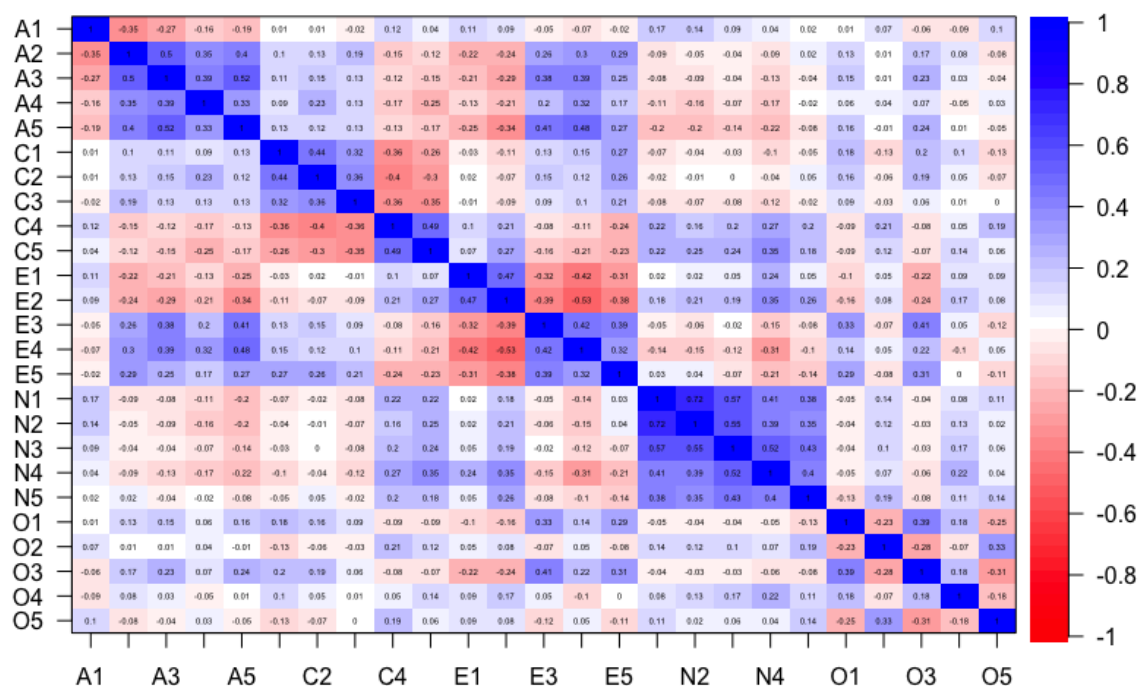


Abbildung 2: Heatmap der empirischen Korrelationsmatrix

Abbildung 2 zeigt eine Heatmap, in der die Korrelationen innerhalb eines Farbspektrums von Blau für eine hohe positive Korrelation von 1 bis nach Rot für eine negative Korrelation von -1 dargestellt werden.

Ein weißes Feld steht für korrelative Unabhängigkeit ( $R = 0$ ), zusätzlich ist in jedem Feld der entsprechende Wert eingetragen. Nach Ermittlung der Korrelationsmatrix wird die Eignung der Ausgangsdaten bzw. der Korrelationsmatrix anhand des Kaiser-Meyer-Olkin Kriteriums überprüft, welches die MSA-Werte (Measure of Sampling Adequacy) jeder Variablen misst. Mit einem arithmetischen Mittel der MSA-Werten von 0.85 und Scores zwischen 0.75 bis 0.90 für die jeweiligen Variablen, ist sie für die Faktorenanalyse nach Kaiser and Rice recht gut (meritorious) geeignet. Zusätzlich zum Kaiser-Meyer-Olkin-Kriterium wird ein Bartlett-Test auf Sphärizität auf ein Sample von  $10\% = 280$  durchgeführt. Mit der Nullhypothese nicht existierender Korrelationen beläuft sich das Chi-Quadrat sich auf 2019.251 mit 300 Freiheitsgraden. Mit  $p=4.301772e-252$  ist diese signifikant, woraufhin die  $H_0$  unkorrelierter Variablen in der Grundgesamtheit abgelehnt wird und weiterhin von korrelierten Variablen ausgegangen wird.

## 4.2 Kommunalitätenschätzung und Auswahl der Faktorenanzahl

### Kommunalitätenschätzung

Die Kommunalitäten der Variablen werden anhand der Squared Multiple Korrelation (SMC) geschätzt. Die SMC's ersetzen als Startwert des Iterationsprozesses die Diagonalelemente der Korrelationsmatrix, wodurch die reduzierte Korrelationsmatrix resultiert. Ihre Replikation ist das Ziel des weiteren Verlaufs und impliziert die Identifikation der Faktoren und ihren Faktorladungen. Tabelle 1 zeigt die sich ergebenden Werte der mit dem quadrierten multiplen Korrelationskoeffizienten geschätzten Kommunalitäten.

A1	A2	A3	A4	A5	C1	C2	C3	C4	C5	E1	E2	E3
0.201	0.393	0.433	0.270	0.418	0.294	0.358	0.255	0.415	0.376	0.332	0.458	0.398

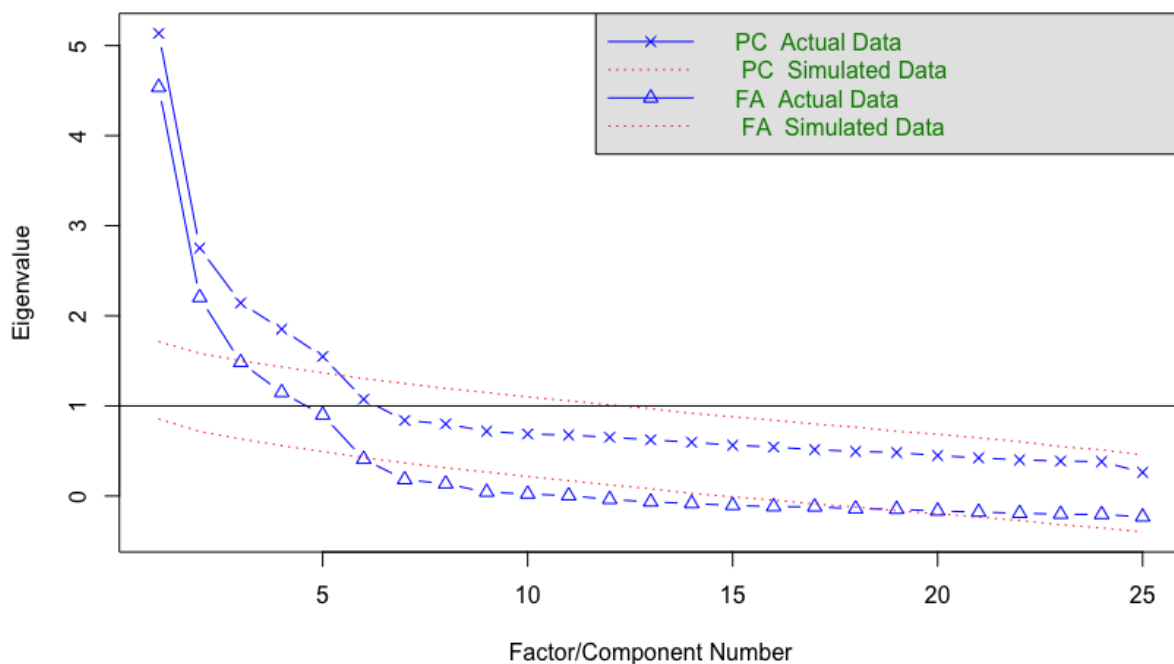
  

E4	E5	N1	N2	N3	N4	N5	01	02	03	04	05
0.472	0.367	0.590	0.573	0.476	0.439	0.317	0.256	0.215	0.353	0.178	0.232

**Tabelle 1:** Mit SMC geschätzte Kommunalitäten der Variablen gerundet auf 3 Nachkommastellen

## Eigenwerte und die Anzahl der Faktoren

Die Eigenwerte spiegeln den durch die einzelnen Faktoren erklärten Varianzanteil wieder. Berücksichtigt werden sowohl die Eigenwerte der reduzierten als auch der empirischen Korrelationsmatrix. Sie dienen zur Bestimmung der Faktorenanzahl. Der **Scree-Plot** in Abbildung 3 bildet absteigend den Eigenvalue in Abhängigkeit der Faktorennummer ab. Der Scree-Test sowie die Parallellanalyse und das Kaiserkriterium können gemeinsam in ihr dargestellt werden. **PC Actual Data** steht für den Eigenwertverlauf der empirischen Korrelationsmatrix und **FA Actual Data** für den Eigenwertverlauf der mit den SMCs reduzierten Korrelationsmatrix. Beide Eigenwertverläufe müssen je nach Methode berücksichtigt werden.



**Abbildung 3:** Scree-Plot von empirischer und reduzierter Korrelationsmatrix mit Parallelanalyse (Eigenwertverläufe normalverteilter Zufallsvariablen in rot) und Kaiserkriterium ( $y = 1$ )

Der Scree-Tests suggeriert die Auswahl eines Modells mit fünf Faktoren. Zwischen dem fünften und dem sechsten Faktor ist die Differenz zwischen den Eigenwerten für beide Verläufe optisch am größten, wodurch an dieser Stelle der Knick entsteht, der die Faktorenanzahl kennzeichnet. Auch zwischen Faktor zwei und drei ist jeweils ein Einbruch zu erkennen, wobei bei 25 Variablen ein Zweifaktorenmodell der Zielsetzung jedoch nicht entspricht. Nach Vergleich der Eigenwertverläufe der reduzierten Korrelationsmatrix mit dem der normalverteilten Zufallsvariablen, ergeben sich im Rahmen der Parallelanalyse fünf relevante Faktoren. Laut dem

Kaiser-Kriterium, welches Faktoren mit Eigenwert  $> 1$  der empirischen Korrelationsmatrix in Betracht zieht, werden sechs Faktoren ausgewählt.

Die Anzahl der extrahierten Faktoren kann durch die gewählte Methode beeinflusst werden. Nach einem Vergleich verschiedener Methoden zur Ermittlung der Faktorenanzahl stellt sich heraus, dass sich je nach Kriterium ein Modell mit zwei, drei und sechs Faktoren rechtfertigen lässt. In Anbetracht des Erklärungsgehaltes des sechsten Faktors mit Eigenwert ( $= 1.1132565$ ) knapp über 1 und der tendenziellen Überschätzung der Faktorenanzahl durch das Kaiser-Kriteriums, halte ich ein Faktorenmodell mit fünf Faktoren für angebracht.

### 4.3 Kriterien für ein Vergleich von Faktorenmodellen

Wie können unterschiedlich aufgebaute Faktorenmodelle am besten verglichen werden? Zur Interpretation und Identifizierung der Faktoren sind die Faktorladungen essenziell und müssen daher auch bei dem Vergleich als entscheidende Parameter im Fokus stehen. Bewertet werden kann die Höhe der Ladungen verschiedener Modelle, ihre Polung sowie die Eindeutigkeit der Ladungen mit Blick auf die Einfachstruktur. Angelehnt an Bühner (2011) werden Faktorladungen  $> |0.3|$  als signifikant angesehen und berücksichtigt. Im Falle von Mehrfachladungen folge ich der Konvention von Backhaus et al. (2015) und der Notwendigkeit von jeweils Ladungen  $> |0.5|$ . Aussagekräftig in einem Vergleich sind ebenfalls die Eigenwerte, an denen der Erklärungsgehalt eines jeden Faktors gemessen und verglichen werden kann sowie, wenn existent, die Faktorinterkorrelationen.

Der von Burt 1948 entwickelte **”Tucker Index of Factor Congruence”**, ermöglicht einen Lagevergleich der Faktorachsen. Die Kongruenz beschreibt den Cosinus des Winkels zwischen jeweils zwei Faktoren verschiedener Faktorenmodelle. Es resultiert eine Matrix, die jeweils den Congruence-Koeffizienten  $\kappa_{xy}$  zwischen -1 und 1, für jede modellübergreifende Faktorkombination ausgibt (Revelle, 2017). Errechnet werden kann dieser mit den Faktorladungen:

$$\kappa_{xy} = \frac{\sum Q_x Q_y}{\sqrt{\sum Q_x^2 \sum Q_y^2}} \quad (13)$$

wobei  $Q_x$  und  $Q_y$  die Spaltenvektoren der Ladungen zweier unterschiedlich aufgebauter Faktorenmodelle  $x$  und  $y$  beschreiben. Während bei  $\kappa = 1 = \cos(0)$  Faktoren in die gleiche Richtung zeigen, verlaufen sie bei  $\kappa = -1$  in die entgegengesetzte Richtung. Orthogonal zueinander stehen sie bei  $\kappa = 0 = \cos(90)$ . Auch Bortz (2013) befasste sich mit der Bewertung

der Ähnlichkeit von Faktorenstrukturen mit Hilfe des Kongruenzmaßes. Ein Ähnlichkeitskoeffizient von über 0.9 signalisiert für ihn eine hohe Übereinstimmung der Faktorstruktur. Wie zuvor die Korrelationsmatrix, kann die resultierende Matrix durch eine Heatmap visualisiert werden.

In dem Vergleich verschiedener Faktorenmodelle fokussiere ich mich auf die Extraktions- und Rotationsmethode. Da die von mir betrachteten Rotationsmethoden alle eine Einfachstruktur anstreben, sollten folglich nicht alle Modelle, unabhängig verschiedener Extraktionen und Rotationen, zum gleichen Ergebnis führen? Berücksichtigt werden muss außerdem die Schiefwinkligkeit der Faktoren nach einer obliquen Rotation. Führen oblique Rotationsmethoden aufgrund der Lockerung der anfänglichen Orthogonalität automatisch zu besseren Ergebnissen? Gleichzeitig ist natürlich hier der Grad der Korrelation zwischen den neu gewonnenen Faktoren zu berücksichtigen. Der Einfluss der Extraktionsmethoden muss ebenfalls untersucht werden. Unterscheiden sich die Ergebnisse in Anbetracht ihrer verschiedenen Grundgedanken und theoretischen Hintergründen? In dem folgenden Kapitel werden verschiedene Modelle unter Berücksichtigung der vorherigen Überlegungen aufgestellt und verglichen.

#### 4.4 Test und Interpretation verschiedener Modelle

Zunächst wird die **unrotierte Faktorenlösung** betrachtet. Extrahiert werden die Faktoren anhand einer Hauptachsenanalyse. Tabelle 2 zeigt die unrotierten Faktorladungen.

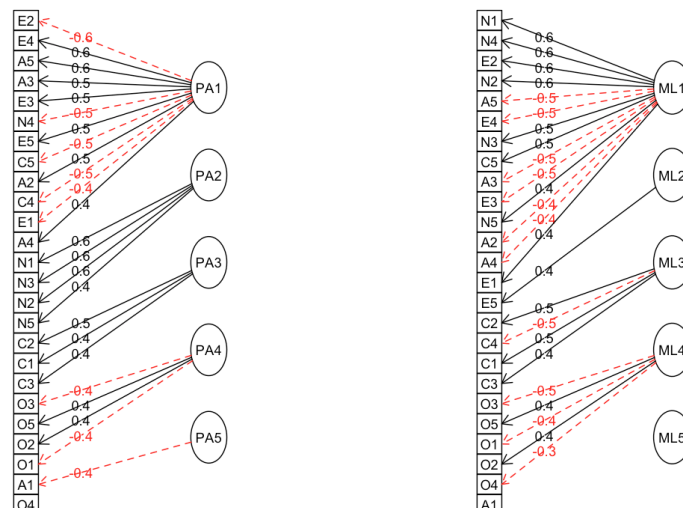
Variable	PA1	PA2	PA3	PA4	PA5	h2	u2	com
A1	-0.22	-0.01	0.10	0.01	<b>-0.36</b>	0.19	0.81	1.86
A2	<b>0.46</b>	0.28	-0.14	0.13	<b>0.35</b>	0.45	0.55	3.02
A3	<b>0.53</b>	0.30	-0.22	0.11	0.30	0.52	0.48	2.81
A4	<b>0.40</b>	0.12	-0.09	0.27	0.18	0.28	0.72	2.55
A5	<b>0.57</b>	0.19	-0.24	0.06	0.18	0.46	0.54	1.87
C1	<b>0.32</b>	0.12	<b>0.45</b>	0.11	-0.05	0.33	0.67	2.15
C2	<b>0.33</b>	0.19	<b>0.48</b>	0.27	0.01	0.45	0.55	2.78
C3	<b>0.31</b>	0.07	<b>0.36</b>	0.30	0.02	0.32	0.68	3.01
C4	<b>-0.46</b>	0.12	<b>-0.44</b>	-0.18	0.05	0.45	0.55	2.48
C5	<b>-0.49</b>	0.15	-0.28	-0.27	0.12	0.43	0.57	2.60
E1	<b>-0.40</b>	-0.19	0.28	0.11	0.25	0.35	0.65	3.20
E2	<b>-0.61</b>	-0.05	0.26	0.05	<b>0.32</b>	0.54	0.46	1.93
E3	<b>0.53</b>	<b>0.33</b>	-0.13	-0.17	-0.08	0.44	0.56	2.13
E4	<b>0.58</b>	0.19	<b>-0.33</b>	0.12	-0.18	0.53	0.47	2.17
E5	<b>0.51</b>	0.28	0.09	-0.03	-0.23	0.40	0.60	2.10
N1	<b>-0.44</b>	<b>0.62</b>	0.03	0.09	-0.26	0.65	0.35	2.24
N2	<b>-0.42</b>	<b>0.61</b>	0.07	0.05	-0.20	0.60	0.40	2.09
N3	<b>-0.41</b>	<b>0.61</b>	0.04	0.05	-0.02	0.55	0.45	1.78
N4	<b>-0.52</b>	<b>0.40</b>	0.09	-0.05	0.20	0.49	0.51	2.30
N5	<b>-0.34</b>	<b>0.42</b>	-0.01	0.20	0.13	0.35	0.65	2.62
O1	<b>0.33</b>	0.21	0.19	<b>-0.36</b>	-0.01	0.31	0.69	3.16
O2	-0.20	0.09	-0.28	<b>0.37</b>	-0.01	0.26	0.74	2.61
O3	<b>0.41</b>	0.28	0.15	<b>-0.44</b>	0.01	0.46	0.54	2.95
O4	-0.06	0.26	0.20	-0.24	0.28	0.25	0.75	3.87
O5	-0.21	-0.04	-0.25	<b>0.42</b>	-0.10	0.30	0.70	2.34
SS loadings	4.49	2.25	1.51	1.19	0.93			
Prop var	0.18	0.09	0.06	0.05	0.04			
Prop explained	0.43	0.22	0.15	0.11	0.09			

**Tabelle 2:** Unrotierte Faktorladungen nach Hauptachsen-Extraktion

In der vorherigen und in den später dargestellten Tabellen wurden die Werte der Faktorladungen **Loadings** auf zwei Dezimalstellen gerundet. Gerundete Faktorladungen mit einem absoluten Wert  $> |0.3|$  werden für eine übersichtlichere Darstellung hervorgehoben, kleinere Werte aufgrund ihrer nicht signifikanten Größe nicht berücksichtigt. In den R Outputs steht **SS loadings** für den Eigenwert eines Faktors. Er ergibt sich aus der Summe der quadrier-

ten Faktorladungen jeder Spalte und entspricht der Varianz, die durch diesen Faktor erklärt wird. **Prop Var** ist der Anteil der Gesamtvarianz, der durch einen Faktor erklärt wird und errechnet sich durch die Division der Eigenwerte durch die Anzahl der Variablen. Der relative Anteil jedes Faktors an dem erklärten Varianzanteil ist unter **Prop explained** aufgeführt.

Die Summe aller Kommunalitäten beträgt 10.37059. Durch die Normierung beträgt die Varianz jeder Variablen 1, woraufhin geschlussfolgert werden kann, dass  $10.37059/25 = 0.4148236 = 41.48\%$  der gemeinsame Varianzanteil ist, der durch ein Faktorenmodell erklärt werden kann. Die restlichen 58.52% sind die spezifische Varianz bzw. Fehlerterm. Mit 10.37059 entspricht die Summe der Eigenwerte der reduzierten Korrelationsmatrix der Summe der Kommunalitäten. Wie erwartet ist in der unrotierten Faktorlösung keine klare Struktur zu erkennen. Die Ladungen sind nicht eindeutig zuzuordnen, da die Items in niedriger und mittlerer signifikanter Höhe auf multiple Faktoren laden. Keine Ladung überschreitet einen Wert von  $|0.62|$ . Der erste Faktor ist von vielen Items beladen und gleicht einem Generalfaktor. Die Eigenwerte der Faktoren bestätigen die Aufklärung sukzessiv maximaler Varianzanteile im Verlauf der Hauptachsen-Extraktion. 43% der erklärten Varianz wird durch den ersten Faktor aufgeklärt. Die Faktoren 2-5 erklären jeweils nur 9%, 6%, 5% und 4% der Gesamtvarianz. Ausgehend von der unrotierten Faktorlösung nach der Hauptachsenanalyse können **verschiedene Extraktionsmethoden verglichen** werden.



**Abbildung 4:** Diagramm der unrotierten Faktorladungen nach Hauptachsen-Extraktion (links) und Maximum-Likelihood-Extraktion (rechts)

Hauptachsen-Extraktion, MinRes und Unweighted Least Squares extrahieren exakt die gleichen Faktoren. Eine Begründung könnte der gemeinsame Least-Squares Ansatz liefern, der den Methoden zu Grunde liegt. Zu Unterschieden kommt es nach der Extraktion durch die Maximum-Likelihood Methode. Sie hat mathematisch im Vergleich zu der Hauptachsenanalyse einen anderen Ansatz. Tabelle 7 im Appendix zeigt die unrotierte Faktorenlösung nach Maximum-Likelihood-Extraktion und Abbildung 4 veranschaulicht den beobachteten Unterschied. In beiden Fällen entsteht ein Generalfaktor, jedoch mit unterschiedlichen Ladungsstrukturen. Die Auswirkung der Extraktionsmethode wird im späteren Verlauf wieder aufgegriffen.

Es bedarf einer Umverteilung der Varianz auf die Faktoren durch eine Faktorenrotation. Nach einer durchgeführten **Varimax-Rotation** (links in Tabelle 3) wurde eine Einfachstruktur der Faktorladungen nach Thurstone nur bedingt erreicht. Mehrfachladungen treten bei vier Variablen (A2, A5, E5, N4) auf. Obwohl die niedrigeren Ladungen in allen Fällen nah am Cutoff-Wert von 0.3 liegen, ist eine exakte Zuordnung in diesen Fällen trotzdem schwierig zu treffen. Die signifikanten Ladungen liegen in einer Spanne von  $|0.37| - |0.77|$ . Es lässt sich in den verbleibenden Variablen des Datensatzes eindeutig eine ihm unterliegende Struktur eines fünf Faktorenmodells identifizieren. Die 21 Persönlichkeitsitems lassen sich, wie es der Aufbau des "International Personality Item Pool" nahelegt, in fünf Konstruktsgruppen unterteilen. Inhaltlich interpretiert werden können die 5 Faktoren nach dem Entwurf des 'International Personality Item Pool' durch die Konstrukte *Agreeableness*, *Conscientiousness*, *Extraversion*, *Emotional Instability* und *Openness* (Goldberg et al., 2006). Beispielsweise dem Konstrukt 'emotionaler Instabilität' unterliegen 'häufige Stimmungsschwankungen' (N3), 'ein Drang zu Irritationen' (N2), 'panisches Verhalten' (N5) und 'Wut' (N1).

Im **Vergleich zur unrotierten** Faktorenlösung sind die Eigenwerte der Faktoren nach der Varimax-Rotation ausgeglichener. Die erklärte Varianz wurde von früher zu später extrahierten Faktoren umverteilt, die Kommunalitäten bleiben folglich unverändert. Die Faktoren eins und zwei haben mit 24% und 26% den höchsten Anteil an der erklärten Varianz. An der Gesamtvarianz tragen alle Faktoren mit Erklärungsanteilen von zwischen 11% und 6% bei. Die Lösung nach der Varimax-Rotation dient als Ausgangspunkt für einen Vergleich verschiedener Rotationsmethoden.



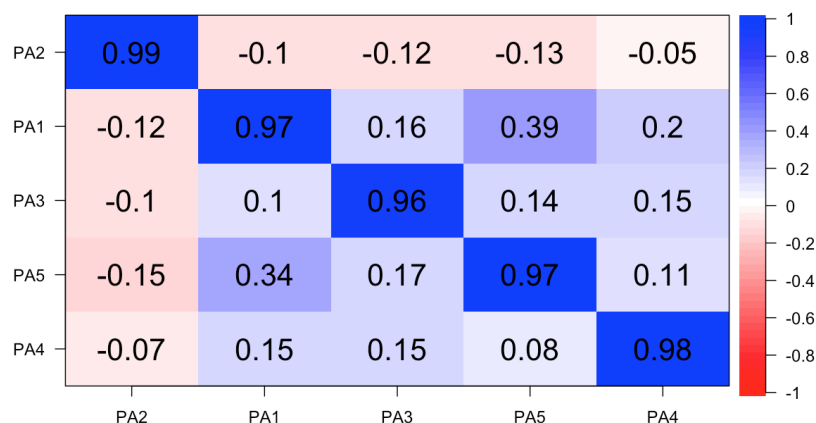
Hauptachsen-Extraktion

	Varimax					Promax					
Variable	PA2	PA1	PA3	PA5	PA4	PA2	PA1	PA3	PA5	PA4	h2
A1	0.12	0.04	0.02	<b>-0.41</b>	-0.08	0.16	0.12	0.07	<b>-0.44</b>	-0.06	<i>0.19</i>
A2	0.03	0.21	0.14	<b>0.62</b>	0.07	0.05	0.11	0.07	<b>0.60</b>	0.00	<i>0.45</i>
A3	0.01	<b>0.32</b>	0.11	<b>0.64</b>	0.06	0.03	0.24	0.00	<b>0.61</b>	-0.01	<i>0.52</i>
A4	-0.06	0.18	0.23	<b>0.42</b>	-0.11	-0.03	0.01	0.19	<b>0.40</b>	-0.16	<i>0.28</i>
A5	-0.11	<b>0.39</b>	0.09	<b>0.53</b>	0.06	-0.08	<b>0.33</b>	-0.03	<b>0.49</b>	0.00	<i>0.46</i>
C1	0.01	0.06	<b>0.54</b>	0.02	0.20	0.07	-0.03	<b>0.56</b>	-0.04	0.15	<i>0.33</i>
C2	0.09	0.03	<b>0.65</b>	0.11	0.11	0.16	-0.09	<b>0.69</b>	0.04	0.04	<i>0.45</i>
C3	-0.02	0.02	<b>0.55</b>	0.12	0.00	0.04	-0.09	<b>0.59</b>	0.07	-0.06	<i>0.32</i>
C4	0.25	-0.06	<b>-0.61</b>	-0.04	-0.11	0.19	0.06	<b>-0.62</b>	0.03	-0.06	<i>0.45</i>
C5	0.30	-0.17	<b>-0.55</b>	-0.06	0.03	0.22	-0.07	<b>-0.55</b>	0.02	0.09	<i>0.43</i>
E1	0.04	<b>-0.57</b>	0.04	-0.10	-0.07	-0.03	<b>-0.63</b>	0.16	0.00	-0.04	<i>0.35</i>
E2	0.25	<b>-0.68</b>	-0.09	-0.10	-0.04	0.15	<b>-0.71</b>	0.05	0.02	0.00	<i>0.54</i>
E3	0.02	<b>0.54</b>	0.08	0.27	0.27	0.09	<b>0.54</b>	-0.05	0.17	0.22	<i>0.44</i>
E4	-0.10	<b>0.65</b>	0.10	0.30	-0.08	-0.01	<b>0.66</b>	-0.03	0.20	-0.14	<i>0.53</i>
E5	0.03	<b>0.50</b>	<b>0.32</b>	0.09	0.21	0.13	<b>0.49</b>	0.24	-0.03	0.15	<i>0.40</i>
N1	<b>0.77</b>	0.07	-0.04	-0.22	-0.08	<b>0.81</b>	0.19	0.03	-0.25	-0.09	<i>0.65</i>
N2	<b>0.75</b>	0.03	-0.03	-0.19	-0.02	<b>0.78</b>	0.14	0.04	-0.22	-0.03	<i>0.60</i>
N3	<b>0.73</b>	-0.06	-0.07	-0.03	0.00	<b>0.74</b>	0.01	0.00	-0.03	-0.01	<i>0.55</i>
N4	<b>0.59</b>	<b>-0.33</b>	-0.17	0.00	0.07	<b>0.53</b>	-0.30	-0.09	0.06	0.08	<i>0.49</i>
N5	<b>0.54</b>	-0.15	-0.03	0.10	-0.15	<b>0.53</b>	-0.13	0.04	0.14	-0.16	<i>0.35</i>
O1	0.01	0.22	0.12	0.06	<b>0.50</b>	0.03	0.20	0.05	0.00	<b>0.48</b>	<i>0.31</i>
O2	0.19	0.00	-0.10	0.09	<b>-0.45</b>	0.20	0.04	-0.06	0.12	<b>-0.47</b>	<i>0.26</i>
O3	0.02	0.30	0.08	0.13	<b>0.59</b>	0.05	0.28	-0.02	0.05	<b>0.57</b>	<i>0.46</i>
O4	0.23	-0.18	-0.01	0.16	<b>0.37</b>	0.19	-0.22	-0.01	0.18	<b>0.38</b>	<i>0.25</i>
O5	0.10	-0.01	-0.06	-0.02	<b>-0.54</b>	0.11	0.04	-0.01	0.01	<b>-0.54</b>	<i>0.30</i>
SS loadings	2.69	2.44	1.98	1.78	1.48	2.67	2.54	1.95	1.75	1.46	
Prop Var	0.11	0.10	0.08	0.07	0.06	0.11	0.10	0.08	0.07	0.06	
Prop explained	0.26	0.24	0.19	0.17	0.14	0.26	0.24	0.19	0.17	0.14	

**Tabelle 3:** Faktorladungen nach Varimax-Rotation (links) und Promax-Rotation (rechts)

Fraglich ist, ob das theoretische Modell wirklich unkorrelierte Faktoren annimmt, da nur in diesem Fall mit einer Varimax-Rotation und ihren orthogonal zu einander stehenden Faktoren als Resultat gearbeitet werden sollten. Wechselseitige Beziehungen zwischen Persönlichkeitskonstrukten entsprechen durchaus der Realität, weshalb ebenfalls oblique Rotationsmethoden betrachtet werden müssen. Anhand einer **Promax-Rotation werde ich der orthogonalen Rotation eine oblique gegenüberstellen** und die Unterschiede in ihren Ladungsstrukturen sowie in der Lage ihrer Faktorenachsen analysieren.

Der Tucker Index of Factor Congruence eignet sich für die Analyse der Achsenlage. Abbildung 6 zeigt eine Heatmap der Kongruenzmatrix:



**Abbildung 5:** Heatmap der Faktor-Kongruenz-Matrix (Varimax auf y-Achse, Promax auf x-Achse)

Die Diagonale der Congruence-Matrix zeichnet sich durch Koeffizienten  $0.95 < \kappa < 1$  aus. Sie vergleicht die Lage der varimaxrotierten und promaxrotierten Faktoren, wobei bei  $\kappa = 1$  die Faktoren in die gleiche Richtung zeigen. Es handelt es sich dementsprechend nur um kleine Differenzen von maximal 16,26 Grad ( $= \arccos(0.96)$ ) bis 8.11 Grad ( $= \arccos(0.99)$ ) zwischen den resultierenden Achsen der verschiedenen Faktormodelle.

Tabelle 3 (rechts) zeigt die Mustermatrix nach Durchführung der obliquen Promax-Rotation. Nach der Promax-Rotation ist eine eindeutig interpretierbare Struktur zu erkennen. Durch sie ergibt sich ein eindeutigeres Bild der Relevanz jeder Variablen in den extrahierten Faktoren. Die relative Wichtigkeit einer gegebenen Variablen für die extrahierten Faktoren kann verglichen werden, ohne den Einfluss der restlichen Variablen. Das Problem der Mehrfachladungen hat sich mit der Promax-Rotation größtenteils aufgelöst. Nur Variable A5, die sowohl auf

Faktor eins als auch auf Faktor fünf lädt, wird nicht weiter berücksichtigt. Das schon nach der Varimax-Rotation aufgefundene Fünf-Faktoren Modell zeichnet sich mit einer Beteiligung von mehr Variablen eindeutiger ab als zuvor. Die Variablen A2, E5 und N4 können nun eindeutig einem Faktor zugeordnet werden. Dem zuvor angesprochenem Konstrukt emotionaler Instabilität ist nun auch 'depressives Verhalten' (N4) zuzuordnen. Die Promax-Rotation führt zu einer besseren Einfachstruktur mit Koeffizienten zwischen  $|0.38|$  -  $|0.81|$ . Ein eindeutiger Trend in der Höhe der Koeffizienten ist jedoch nicht zu erkennen.

Tabelle 10 im Anhang zeigt die Strukturmatrix. Auffällig sind die Mehrfachladungen, die eine hohe Korrelation der Faktoren vermuten lassen. Da oblique Rotationen aufgrund der gelockerten Orthogonalitätsbedingung **Korrelationen** zwischen den Faktoren zulassen, müssen auch diese genauer betrachtet werden. Die Korrelationen sind in Tabelle 4 abgebildet.

	PA2	PA1	PA3	PA5	PA4
PA2	1.00	-0.24	-0.22	-0.01	0.02
PA1	-0.24	1.00	0.39	0.35	0.15
PA3	-0.22	0.39	1.00	0.25	0.19
PA5	-0.01	0.35	0.25	1.00	0.16
PA4	0.02	0.15	0.19	0.16	1.00

**Tabelle 4:** Korrelationsmatrix der Faktoren nach obliquer Promax-Rotation

Nach der Promax-Rotation weisen die Faktoren signifikante Korrelationen auf, zwischen Faktor 1 und 3 von 0.39 sowie zwischen 1 und 5 von 0.35. Niedrigere Korrelationen bestehen zwischen 3 und 5 mit 0.25, 2 und 1 mit -0.24 und 2 und 3 mit -0.22. Die Mehrfachladungen in der Strukturmatrix erklären sich durch die hohen Korrelationen zwischen den Faktoren. Aus den Korrelationen lässt sich schlussfolgern, dass von einem theoretischen Modell mit korrelierten Faktoren ausgegangen werden sollte, um den Sachverhalt möglichst realitätsgetreu darzustellen.

Die möglichen **Auswirkungen der Extraktionsmethode** können an der Promax-Rotation überprüft werden. Bestätigen sich die in der unrotierten Faktorenlösung beobachteten Unterschiede? Tabelle 5 zeigt links die Faktorladungen der Promax-Rotation nach der Maximum-Likelihood-Extraktion. Das vorherige Ergebnis nach Hauptachsen-Extraktion und Promax-Rotation (Tabelle 3, rechts) kann mit dem jetzigen, nach der Maximum-Likelihood-Extraktion,

verglichen werden. Bei einem Vergleich der Outputs wird ersichtlich, dass das Modell mit der Maximum-Likelihood-Extraktion in vergleichbaren Ergebnissen resultiert. Es weist eine ähnlich gute Einfachstruktur mit lediglich einer Mehrfachladung auf, wobei diese sich nun in Variable N4 befindet und nicht wie zuvor in Variable A5. Auch in der Ausprägung der Faktorladungen sowie in den entstandenen Korrelationen gibt es keine signifikanten Unterschiede.

Maximum-Likelihood-Extraktion											
	Promax					Oblimin					
Variable	ML2	ML1	ML3	ML5	ML4	ML2	ML3	ML5	ML1	ML4	h2
A1	0.15	0.11	0.04	<b>-0.39</b>	-0.04	0.20	0.04	<b>-0.36</b>	-0.14	-0.04	<i>0.15</i>
A2	0.05	0.10	0.08	<b>0.57</b>	0.00	-0.02	0.09	<b>0.60</b>	0.01	0.03	<i>0.40</i>
A3	0.04	0.20	0.01	<b>0.62</b>	-0.01	-0.03	0.03	<b>0.67</b>	-0.07	0.04	<i>0.51</i>
A4	-0.02	0.08	0.20	<b>0.43</b>	-0.17	-0.06	0.20	<b>0.46</b>	-0.04	-0.15	<i>0.29</i>
A5	-0.09	0.28	-0.04	<b>0.55</b>	0.01	-0.14	0.00	<b>0.58</b>	-0.17	0.06	<i>0.48</i>
C1	0.06	-0.05	<b>0.54</b>	-0.01	0.16	0.06	<b>0.53</b>	0.00	0.05	0.16	<i>0.32</i>
C2	0.14	-0.11	<b>0.66</b>	0.08	0.06	0.13	<b>0.64</b>	0.11	0.13	0.06	<i>0.43</i>
C3	0.04	-0.10	<b>0.58</b>	0.09	-0.06	0.04	<b>0.56</b>	0.11	0.08	-0.06	<i>0.32</i>
C4	0.15	0.03	<b>-0.64</b>	0.05	-0.04	0.12	<b>-0.64</b>	0.06	0.04	-0.03	<i>0.47</i>
C5	0.19	-0.09	<b>-0.56</b>	0.01	0.09	0.14	<b>-0.57</b>	0.01	0.16	0.10	<i>0.43</i>
E1	-0.05	<b>-0.65</b>	0.16	-0.01	-0.04	-0.09	0.10	-0.10	<b>0.57</b>	-0.11	<i>0.37</i>
E2	0.12	<b>-0.71</b>	0.05	-0.02	-0.01	0.06	-0.03	-0.09	<b>0.67</b>	-0.07	<i>0.55</i>
E3	0.07	<b>0.48</b>	-0.07	0.24	0.25	0.06	-0.02	0.30	<b>-0.34</b>	<b>0.31</b>	<i>0.44</i>
E4	-0.01	<b>0.61</b>	-0.04	0.27	-0.11	0.00	0.01	<b>0.36</b>	<b>-0.53</b>	-0.05	<i>0.52</i>
E5	0.16	<b>0.48</b>	0.24	-0.01	0.17	0.18	0.27	0.08	<b>-0.39</b>	0.22	<i>0.40</i>
N1	<b>0.86</b>	0.22	0.04	-0.27	-0.10	<b>0.85</b>	0.01	-0.09	-0.09	-0.05	<i>0.71</i>
N2	<b>0.83</b>	0.17	0.05	-0.25	-0.04	<b>0.82</b>	0.02	-0.08	-0.04	0.01	<i>0.66</i>
N3	<b>0.72</b>	0.00	-0.01	-0.02	-0.01	<b>0.66</b>	-0.06	0.10	0.14	0.03	<i>0.53</i>
N4	<b>0.49</b>	<b>-0.32</b>	-0.10	0.05	0.08	<b>0.41</b>	-0.16	0.09	<b>0.42</b>	0.08	<i>0.48</i>
N5	<b>0.50</b>	-0.16	0.03	0.15	-0.15	<b>0.44</b>	-0.02	0.22	0.25	-0.14	<i>0.34</i>
O1	0.02	0.15	0.03	0.02	<b>0.51</b>	-0.01	0.06	0.02	-0.06	<b>0.53</b>	<i>0.32</i>
O2	0.17	0.02	-0.08	0.15	<b>-0.45</b>	0.16	-0.10	0.21	-0.03	<b>-0.44</b>	<i>0.24</i>
O3	0.04	0.24	-0.04	0.08	<b>0.59</b>	0.01	0.00	0.09	-0.10	<b>0.63</b>	<i>0.47</i>
O4	0.16	-0.26	-0.02	0.17	<b>0.38</b>	0.08	-0.04	0.15	<b>0.36</b>	<b>0.38</b>	<i>0.26</i>
O5	0.10	0.02	-0.03	0.05	<b>-0.52</b>	0.11	-0.05	0.10	-0.07	<b>-0.52</b>	<i>0.27</i>
SS loadings	2.65	2.39	1.94	1.88	1.49	2.49	2.05	2.1	2.07	1.64	
Prop Var	0.11	0.10	0.08	0.08	0.06	0.10	0.08	0.08	0.08	0.07	
Prop explained	0.26	0.23	0.19	0.18	0.14	0.24	0.20	0.20	0.20	0.16	

**Tabelle 5:** Faktorladungen nach Promax-Rotation (links) und Direct-Oblimin (rechts)

Zuvor hatte sich die oblique Rotation gegenüber der orthogonalen bewiesen, weshalb dem Ergebnis mit **Direct-Oblimin** eine weitere oblique Rotation gegenübergestellt wird. Die re-

sultierenden Ladungen sind in Tabelle 5 rechts abgebildet.

Trotz einer schlechteren Einfachstruktur im Vergleich zu Promax aufgrund von Mehrfachladungen bei E3, E4, N4 und O4, zeigt sich die Vorteilhaftigkeit von Direct-Oblimin gegenüber der orthogonalen Varimax-Rotation. Die Varianz wurde gleichmäßiger auf die Faktoren verteilt. Jeder Faktor erklärt zwischen 7 und 10% der Gesamtvarianz, wobei Faktoren drei, fünf und eins jeweils 20% der erklärten Varianz ausmachen.

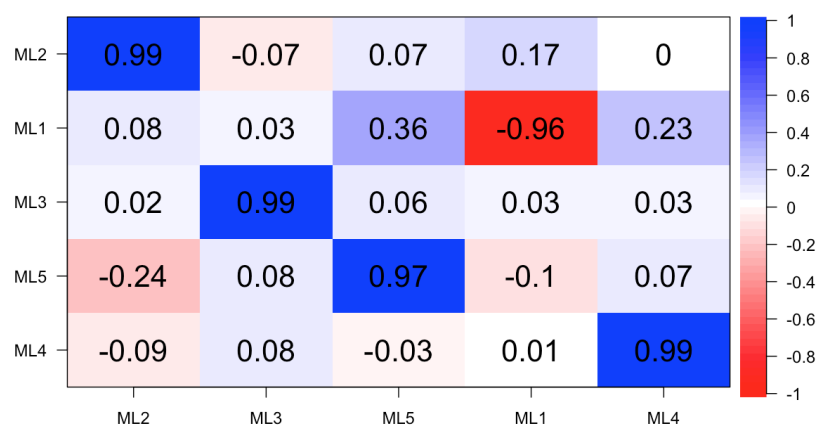
Tabelle 6 zeigt die Korrelationen der obliquen Faktoren nach Promax- und Direct-Oblimin-Rotation.

	ML2	ML1	ML3	ML5	ML4		ML2	ML3	ML5	ML1	ML4
ML2	1.00	-0.27	-0.23	0.01	0.03	ML2	1.00	-0.21	-0.03	0.24	-0.01
ML1	-0.27	1.00	0.38	0.34	0.15	ML3	-0.21	1.00	0.20	-0.22	0.20
ML3	-0.23	0.38	1.00	0.24	0.21	ML5	-0.03	0.20	1.00	-0.31	0.23
ML5	0.01	0.34	0.24	1.00	0.19	ML1	0.24	-0.22	-0.31	1.00	-0.17
ML4	0.03	0.15	0.21	0.19	1.00	ML4	-0.01	0.20	0.23	-0.17	1.00

**Tabelle 6:** Korrelationen nach Promax-Rotation (links) Direct-Oblimin (rechts) und Maximum-Likelihood-Extraktion

Obwohl nach beiden Methoden die gleichen Faktoren miteinander korreliert sind, ergibt die Oblimin-Rotation im Betrag niedrigere Korrelationen. Zwischen den Faktoren fünf und eins besteht nach der Direct-Oblimin-Rotation mit -0.31 die einzige signifikante Korrelation. Im Vergleich dazu, beträgt die Korrelation zwischen Faktor fünf und Faktor eins nach der Promax-Rotation 0.34. Auffällig sind die vermehrt negativen Korrelationen der Faktoren nach der Oblimin-Rotation. Die höchste Korrelation nach der Promax-Rotation beträgt 0.38 zwischen den Faktoren drei und eins. Nach Direct-Oblimin hingegen hat sie mit -0.22 ebenfalls einen negativen Wert. Aufschluss über die Ursache kann die Kongruenzmatrix in Abbildung 6 geben. Sie zeigt, dass sich Faktor eins im Direkt-Oblimin-Modell entgegengesetzt zu Faktor eins im Promax-Modell positioniert hat. Zwischen ihnen liegt ein Winkel von 163,7398 Grad. Die Lage der anderen Faktorenachsen beider Modelle entsprechen einander. Die Maximale Abweichung zwischen ihnen liegt bei Faktor 5 bei 14,06987 Grad. Zwischen den Faktoren zwei, drei und vier liegen jeweils 8,10961 Grad. Die entgegengesetzte Lage von Faktor eins beeinflusst die Interpretation der Faktoren in Bezug zu den Variablen. Die Ladungen E1, E2, E3, E4 und E5 weisen nach der Oblimin-Rotation jeweils ein umgekehrtes Vorzeichen auf. Die Items mit zuvor positiven Ladungen 'Verantwortungsübernahme' (E5), dem 'einfachen Schluss von

Freundschaften' (E4) sowie dem 'Gewinn von Mitmenschen' (E3) laden nun negativ und die anfänglich negativen bezüglich 'Verschwiegenheit'(E1) und 'Berührungsängsten' (E2) positiv. Während nach der Promax-Rotation (und allen bisher betrachteten Modellen) der erste Faktor das Konstrukt der Extraversion widerspiegelt, legt die oppositionelle Polung nach der Oblimin-Rotation eine Interpretation des Faktors als Intraversion nahe.



**Abbildung 6:** Heatmap der Faktor-Kongruenz-Matrix (Promax auf y-Achse, Oblimin auf x-Achse)

Trotz kleiner Differenzen in den Faktorinterkorrelationen und abgesehen von der inhaltlich konträren Interpretation eines Faktors ausgelöst durch die entgegengesetzte Achsenlage, identifiziert ein Faktorenmodell mit Direct-Oblimin-Rotation nach Maximum-Likelihood-Extraktion Faktoren mit den gleichen Variablen wie zuvor die Promax-Rotation nach sowohl Hauptachsen- und Maximum-Likelihood-Extraktion und die Varimax-Rotation nach Hauptachsen-Extraktion.

## 5 Fazit

Nach der Vorstellung der Theorie hinter der Faktorenanalyse und ihrer Vielfalt in der Modellkonstruktion wurde die Tragweite subjektiv getroffenen Entscheidungen des Wissenschaftlers im Zuge Ihrer Durchführung betrachtet. Durch die Wahl bestimmter methodischer Herangehensweisen greift er in die Analyse ein. Vor einem Entscheidungsproblem steht er bei der Kommunalitätenschätzung, in der Auswahl der Faktorenanzahl, der Extraktionsmethode sowie der Rotationsmethode. Ebenfalls besteht für den Anwender Spielraum in der Bestimmung und Interpretation der Faktoren.

Im Fokus meiner Untersuchungen stand der Vergleich unterschiedlich aufgebauter Modelle, mit dem Ziel, den Einfluss getroffener Modellierungsentscheidungen auf das Ergebnis der Faktorenanalyse zu ergründen und zu beurteilen. Im Rahmen meiner Herangehensweise an die Problematik wurde eine interaktive Applikation mit dem R Package Shiny entwickelt, die den Prozess der Faktorenanalyse visualisiert und einen Vergleich verschieden konstruierter Faktorenmodelle ermöglicht. Sie involviert den Benutzer, der über Faktorenanzahl, Extraktions- und Rotationsmethode entscheidet und so die Möglichkeit hat die Auswirkungen seiner getroffenen Entscheidungen zu erforschen.

Im Verlauf eines Faktorenmodellvergleichs anhand von Persönlichkeitsmerkmalen an ließen sich Strukturen feststellen und sich an diese fortführende Überlegungen anknüpfen. Es gilt anzumerken, dass sie fest in Verbindung zu den verwendeten Daten stehen und sich nicht generalisieren lassen. In der Analyse wurden vier verschiedene Modelle aufgebaut, durch die unabhängig von der Rotations- und Extraktionsmethode Faktoren mit den gleichen Variablen gebildet wurden. In einem Modell unterschied sich die Interpretation eines Faktors.

In einem Vergleich zwischen einer orthogonalen (Varimax) und einer obliquen Rotation (Promax) hat sich die oblique als die bessere Wahl herausgestellt. Ihre Vorteilhaftigkeit bestand in der Einfachstruktur und der eindeutigen Interpretierbarkeit der Faktorladungen. Die Zuordnung einiger Variablen war aufgrund von Mehrfachladungen nach der Varimax-Rotation nicht immer möglich. Trotz eines geringen Unterschiedes in der Achsenlage hat sich die oblique Promax-Rotation gegenüber der orthogonalen Varimax-Rotation als die vorteilhaftere

bewiesen. Abgesehen von der Elimination der Mehrfachladungen, entsprechen korrelierte Faktoren als Ergebnis der Faktorenanalyse dem Streben nach einem realitätsgetreuen Modell. Zu begründen ist es durch konzeptionell inhaltliche Überlegungen von stets korrelierenden Persönlichkeitsmerkmalen. Der Vergleich zweier obliquen Rotationen, Promax und Direct-Oblimin nach jeweils der Maximum-Likelihood-Extraktion, führte zu einer besseren Interpretierbarkeit der Promax Ergebnisse. Die Oblimin rotierten Faktoren wiesen geringere Interkorrelationen auf. Die Lage der Faktorenachsen unterschied sich durch die entgegengesetzte Ausrichtung einer Faktorenachse. Die dadurch ausgelöste, gegensätzliche Polung der Ladungen auf den angesprochenen Faktor hatte eine inhaltlich oppositionelle Faktoreninterpretation zu Folge. Die oblique Promax-Rotation hat sich als die beste Rotationsmethode mit den eindeutigsten Ergebnissen herausgestellt. Die Extraktionsmethode spielte keine Rolle. Zuvor in der unrotierten Faktorenlösung beobachtete Unterschiede, ausgelöst durch verschiedene Extraktionsverfahren, ließen sich nicht bestätigen. Die Ergebnisse nach Maximum-Likelihood-Extraktion und Hauptachsen-Extraktion und jeweils folgender Promax-Rotation unterschieden sich mit Sicht auf ihre Einfachstruktur und ihre Interpretierbarkeit nicht.

Meine Analyse bestätigt, dass es sich in der Modellierung einer Faktorenanalyse nicht um arbiträre Entscheidungen ohne Konsequenzen handelt. Unterschiede in den Ergebnissen verschieden modellierter Analysen wurden nachgewiesen. Beziehungen zwischen Modellstruktur und Resultaten wurden bekräftigt, doch die Frage nach optimalen Faktormodellen und genauen Auswirkungen bestimmter Komponenten in ihrer Modellierung ist zu komplex, als dass sie sich im Rahmen dieser Arbeit exakt beantworten lässt. Präzise Aussagen über die Vorteilhaftigkeit bestimmter Entscheidungen in den Phasen der Faktorenanalyse lassen sich nur durch Bestätigungen anhand wiederholter Experimente an verschiedenen Datensätzen treffen. Meine Ergebnisse implizieren die Abhängigkeit eines Vergleichs verschiedener Faktorenmodelle vom Untersuchungsinhalt. Es ergeben sich neue Fragestellungen, z.B. nach dem Resultat einer Faktorenanalyse, in der es sich in den untersuchten Konstrukten nicht um Persönlichkeitsmerkmale handelt, sondern um Faktoren, deren inhaltliche Korrelation sich nicht eindeutig bestimmen lässt. Zu welchen Ergebnissen würden verschieden modellierte Faktorenanalysen führen? Wie würden verschieden aufgebaute Faktorenmodelle sich in einem solchen Falle unterscheiden? Diese Fragen sind eine Herausforderung für weitere Analysen auf diesem Gebiet.



## Literatur

- BACKHAUS, K., B. ERICHSON, W. PLINKE, AND R. WEIBER (2015): *Multivariate analysemethoden: eine anwendungsorientierte einföhrung*, Springer-Verlag.
- BELLMANN, M. (2016): “Factor analysis of geometric figures with four attributes: A comparison of PCA, varimax and varimin,” *Personality and Individual Differences*, 90, 326–331.
- BORTZ, J. (2013): *Statistik: Für Sozialwissenschaftler*, Springer-Verlag.
- BÜHNER, M. (2011): *Einföhrung in die Test-und Fragebogenkonstruktion*, Pearson Deutschland GmbH.
- CATTELL, R. (1978): *The scientific use of factor analysis in behavioral and life sciences*, Springer Science Business Media.
- CHANG, W., J. CHENG, J. ALLAIRE, Y. XIE, AND J. MCPHERSON (2017): *shiny: Web Application Framework for R*, r package version 1.0.5.
- COSTELLO, A. B. AND J. W. OSBORNE (2005): “Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis,” *Practical assessment, research & evaluation*, 10, 1–9.
- DARTON, R. A. (1980): “Rotation in Factor Analysis,” *Journal of the Royal Statistical Society. Series D (The Statistician)*, 29, 167–194.
- ERTEL, S. (2016): *Komplexität modellieren-Faktorenanalyse am Scheideweg*, Universitätsverlag Göttingen.
- FABRIGAR, L. R., D. T. WEGENER, R. C. MACCALLUM, AND E. J. STRAHAN (1999): “Evaluating the use of exploratory factor analysis in psychological research,” *Psychological methods*, 4, 272.
- FAVA, J. L. AND W. F. VELICER (1992): “The effects of overextraction on factor and component analysis,” *Multivariate Behavioral Research*, 27, 387–415.
- GOLDBERG, L. R., J. A. JOHNSON, H. W. EBER, R. HOGAN, M. C. ASHTON, C. R. CLONINGER, AND H. G. GOUGH (2006): “The international personality item pool and the future of public-domain personality measures,” *Journal of Research in personality*, 40, 84–96.

- GORSUCH, R. L. (1983): *Factor analysis. 2nd*, Hillsdale, NJ: LEA.
- HÄRDLE, W. AND L. SIMAR (2007): *Applied multivariate statistical analysis*, Springer Science & Business Media.
- HÜTTNER, M. (2002): *Grundzüge der Marktforschung*, Walter de Gruyter.
- IACOBUCCI, D. (2001): "Methodological and statistical concerns of the experimental behavioral researcher: factor analysis," *Journal of Consumer Psychology*, 10, 75–82.
- JENNRICH, R. I. AND P. SAMPSON (1966): "Rotation for simple loadings," *Psychometrika*, 31, 313–323.
- KAISER, H. F. (1958): "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, 23, 187–200.
- KAISER, H. F. AND J. RICE (1974): "Little jiffy, mark IV," *Educational and psychological measurement*, 34, 111–117.
- KIEFFER, K. M. (1998): "Orthogonal versus Oblique Factor Rotation: A Review of the Literature regarding the Pros and Cons." .
- KLINKE, S. (2017): "Datenanalyse 1 WS 2017/18 Vorlesungsfolien," .
- KLOPP, E. (2010): "Explorative Faktorenanalyse," .
- NOACK, M. (2007): "Faktorenanalyse," *Online verfügbar unter <http://www.unidue.de/imperia/md/content/soziologie/stein/faktorenanalyse.pdf>, zuletzt geprüft am, 30, 2013.*
- R CORE TEAM (2017): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- REVELLE, W. (2017): "An introduction to the psych package: Part II Scale construction and psychometrics," .
- RINKENBURGER, R. (2011): "Einführung in die explorative Faktorenanalyse," *Theorien und Methoden der Betriebswirtschaft Handbuch für Wissenschaftler und Studierende*, 455–476.
- RSTUDIO, INC (2013): *Easy web applications in R.*, uRL: <http://www.rstudio.com/products/shiny/>.
- SPEARMAN, C. (1904): "General Intelligence, Objectively Determined and Measured," *The American Journal of Psychology*, 15, 201–292.

THURSTONE, L. L. (1947): “Multiple factor analysis.” .

ÜBERLA, K. (1968): *Faktorenanalyse*, Springer Berlin-Heidelberg-New York.

VOSS, A. (2009): “Statistik 2 WS 2009/2010 Vorlesungsfolien,” .

WIRTSCHAFTSLEXIKON, G. (2017a): “Dependenzanalyse,” *Online im Internet:*  
*[<http://wirtschaftslexikon.gabler.de/Archiv/2247/dependenzanalyse-v7.html>],* *abgerufen*  
*am: 14.01.2018.*

——— (2017b): “Maximum-Likelihood-Methode,” *Online im*  
*Internet:[[http://wirtschaftslexikon.gabler.de/Archiv/88974/maximum-likelihood-methode-](http://wirtschaftslexikon.gabler.de/Archiv/88974/maximum-likelihood-methode-v9.html)*  
*v9.html], abgerufen am: 30.11.2017.*

## A Tabellen

Maximum-Likelihood Extraktion								
Variable	ML1	ML2	ML3	ML4	ML5	h2	u2	com
A1	0.23	-0.02	0.08	0.02	-0.30	0.15	0.85	2.08
A2	<b>-0.39</b>	<b>0.35</b>	-0.09	0.08	<b>0.34</b>	0.40	0.60	3.19
A3	<b>-0.46</b>	<b>0.39</b>	-0.18	0.08	<b>0.33</b>	0.51	0.49	3.23
A4	<b>-0.37</b>	0.20	-0.03	0.24	0.23	0.29	0.71	3.18
A5	<b>-0.54</b>	0.30	-0.22	0.04	0.23	0.48	0.52	2.38
C1	-0.27	0.18	<b>0.46</b>	0.04	0.01	0.32	0.68	1.96
C2	-0.26	0.24	<b>0.52</b>	0.16	0.10	0.43	0.57	2.30
C3	-0.27	0.13	<b>0.41</b>	0.24	0.08	0.32	0.68	2.79
C4	<b>0.44</b>	-0.01	<b>-0.49</b>	-0.17	0.07	0.47	0.53	2.27
C5	<b>0.48</b>	0.00	<b>-0.34</b>	-0.26	0.11	0.43	0.57	2.55
E1	<b>0.36</b>	<b>-0.31</b>	0.26	0.04	0.27	0.37	0.63	3.78
E2	<b>0.58</b>	-0.23	0.22	-0.03	<b>0.33</b>	0.55	0.45	2.29
E3	<b>-0.46</b>	<b>0.43</b>	-0.13	-0.18	-0.03	0.44	0.56	2.48
E4	<b>-0.54</b>	<b>0.34</b>	-0.28	0.15	-0.10	0.52	0.48	2.57
E5	<b>-0.41</b>	<b>0.42</b>	0.12	-0.03	-0.21	0.40	0.60	2.61
N1	<b>0.59</b>	<b>0.56</b>	0.03	0.10	-0.17	0.71	0.29	2.23
N2	<b>0.58</b>	<b>0.55</b>	0.07	0.05	-0.14	0.66	0.34	2.15
N3	<b>0.53</b>	<b>0.49</b>	0.02	0.01	0.09	0.53	0.47	2.05
N4	<b>0.58</b>	0.23	0.04	-0.12	0.26	0.48	0.52	1.84
N5	<b>0.42</b>	0.30	-0.01	0.14	0.23	0.34	0.66	2.73
O1	-0.27	0.25	0.16	<b>-0.40</b>	-0.02	0.32	0.68	2.89
O2	0.19	0.05	-0.25	<b>0.37</b>	0.09	0.24	0.76	2.53
O3	<b>-0.34</b>	<b>0.34</b>	0.11	<b>-0.48</b>	-0.01	0.47	0.53	2.77
O4	0.11	0.18	0.15	<b>-0.33</b>	0.28	0.26	0.74	3.32
O5	0.18	-0.06	-0.22	<b>0.44</b>	0.00	0.27	0.73	1.92
SS loadings	4.37	2.34	1.51	1.19	0.95			

**Tabelle 7:** Unrotierte Faktorenlösung nach Maximum-Likelihood Extraktion

Hauptachsen-Extraktion								
Variable	PA2	PA1	PA3	PA5	PA4	h2	u2	com
A1	0.12	0.04	0.02	<b>-0.41</b>	-0.08	0.19	0.81	1.29
A2	0.03	0.21	0.14	<b>0.62</b>	0.07	0.45	0.55	1.37
A3	0.01	<b>0.32</b>	0.11	<b>0.64</b>	0.06	0.52	0.48	1.56
A4	-0.06	0.18	0.23	<b>0.42</b>	-0.11	0.28	0.72	2.17
A5	-0.11	<b>0.39</b>	0.09	<b>0.53</b>	0.06	0.46	0.54	2.04
C1	0.01	0.06	<b>0.54</b>	0.02	0.20	0.33	0.67	1.29
C2	0.09	0.03	<b>0.65</b>	0.11	0.11	0.45	0.55	1.15
C3	-0.02	0.02	<b>0.55</b>	0.12	0.00	0.32	0.68	1.10
C4	0.25	-0.06	<b>-0.61</b>	-0.04	-0.11	0.45	0.55	1.45
C5	0.30	-0.17	<b>-0.55</b>	-0.06	0.03	0.43	0.57	1.79
E1	0.04	<b>-0.57</b>	0.04	-0.10	-0.07	0.35	0.65	1.11
E2	0.25	<b>-0.68</b>	-0.09	-0.10	-0.04	0.54	0.46	1.35
E3	0.02	<b>0.54</b>	0.08	0.27	0.27	0.44	0.56	2.03
E4	-0.10	<b>0.65</b>	0.10	0.30	-0.08	0.53	0.47	1.56
E5	0.03	<b>0.50</b>	<b>0.32</b>	0.09	0.21	0.40	0.60	2.17
N1	<b>0.77</b>	0.07	-0.04	-0.22	-0.08	0.65	0.35	1.21
N2	<b>0.75</b>	0.03	-0.03	-0.19	-0.02	0.60	0.40	1.14
N3	<b>0.73</b>	-0.06	-0.07	-0.03	0.00	0.55	0.45	1.03
N4	<b>0.59</b>	<b>-0.33</b>	-0.17	0.00	0.07	0.49	0.51	1.82
N5	<b>0.54</b>	-0.15	-0.03	0.10	-0.15	0.35	0.65	1.41
O1	0.01	0.22	0.12	0.06	<b>0.50</b>	0.31	0.69	1.55
O2	0.19	0.00	-0.10	0.09	<b>-0.45</b>	0.26	0.74	1.53
O3	0.02	0.30	0.08	0.13	<b>0.59</b>	0.46	0.54	1.66
O4	0.23	-0.18	-0.01	0.16	<b>0.37</b>	0.25	0.75	2.61
O5	0.10	-0.01	-0.06	-0.02	<b>-0.54</b>	0.30	0.70	1.09
SS loadings	2.69	2.44	1.98	1.78	1.48			
Prop Var	0.11	0.10	0.08	0.07	0.06			
Prop explained	0.26	0.24	0.19	0.17	0.14			

**Tabelle 8:** Faktorladungen nach Varimax-Rotation und Hauptachsen-Extraktion

Hauptachsen-Extraktion								
Variable	PA2	PA1	PA3	PA5	PA4	h2	u2	com
A1	0.16	0.12	0.07	<b>-0.44</b>	-0.06	0.19	0.81	1.51
A2	0.05	0.11	0.07	<b>0.60</b>	0.00	0.45	0.55	1.10
A3	0.03	0.24	0.00	<b>0.61</b>	-0.01	0.52	0.48	1.30
A4	-0.03	0.10	0.19	<b>0.40</b>	-0.16	0.28	0.72	1.95
A5	-0.08	<b>0.33</b>	-0.03	<b>0.49</b>	0.00	0.46	0.54	1.82
C1	0.07	-0.03	<b>0.56</b>	-0.04	0.15	0.33	0.67	1.19
C2	0.16	-0.09	<b>0.69</b>	0.04	0.04	0.45	0.55	1.16
C3	0.04	-0.09	<b>0.59</b>	0.07	-0.06	0.32	0.68	1.11
C4	0.19	0.06	<b>-0.62</b>	0.03	-0.06	0.45	0.55	1.24
C5	0.22	-0.07	<b>-0.55</b>	0.02	0.09	0.43	0.57	1.40
E1	-0.03	<b>-0.63</b>	0.16	0.00	-0.04	0.35	0.65	1.14
E2	0.15	<b>-0.71</b>	0.05	0.02	0.00	0.54	0.46	1.10
E3	0.09	<b>0.54</b>	-0.05	0.17	0.22	0.44	0.56	1.61
E4	-0.01	<b>0.66</b>	-0.03	0.20	-0.14	0.53	0.47	1.29
E5	0.13	<b>0.49</b>	0.24	-0.03	0.15	0.40	0.60	1.85
N1	<b>0.81</b>	0.19	0.03	-0.25	-0.09	0.65	0.35	1.34
N2	<b>0.78</b>	0.14	0.04	-0.22	-0.03	0.60	0.40	1.23
N3	<b>0.74</b>	0.01	0.00	-0.03	-0.01	0.55	0.45	1.00
N4	<b>0.53</b>	-0.30	-0.09	0.06	0.08	0.49	0.51	1.73
N5	<b>0.53</b>	-0.13	0.04	0.14	-0.16	0.35	0.65	1.49
O1	0.03	0.20	0.05	0.00	<b>0.48</b>	0.31	0.69	1.36
O2	0.20	0.04	-0.06	0.12	<b>-0.47</b>	0.26	0.74	1.54
O3	0.05	0.28	-0.02	0.05	<b>0.57</b>	0.46	0.54	1.51
O4	0.19	-0.22	-0.01	0.18	<b>0.38</b>	0.25	0.75	2.73
O5	0.11	0.04	-0.01	0.01	<b>-0.54</b>	0.30	0.70	1.09
SS loadings	2.67	2.54	1.95	1.75	1.46			
Prop Var	0.11	0.10	0.08	0.07	0.06			
Prop explained	0.26	0.24	0.19	0.17	0.14			
PA2	1.00	-0.24	-0.22	-0.01	0.02			
PA1	-0.24	1.00	0.39	0.35	0.15			
PA3	-0.22	0.39	1.00	0.25	0.19			
PA5	-0.01	0.35	0.25	1.00	0.16			
PA4	0.02	0.15	0.19	0.16	1.00			

**Tabelle 9:** Faktorenmuster nach Promax-Rotation und Hauptachsen-Extraktion inklusive Korrelationsmatrix der Faktoren

Hauptachsen-Extraktion					
Variable	PA2	PA1	PA3	PA5	PA4
A1	0.12	-0.06	-0.04	<b>-0.40</b>	-0.10
A2	0.00	<b>0.33</b>	0.25	<b>0.66</b>	0.13
A3	-0.04	<b>0.44</b>	0.24	<b>0.69</b>	0.12
A4	-0.10	0.29	0.30	<b>0.46</b>	-0.05
A5	-0.16	<b>0.51</b>	0.24	<b>0.60</b>	0.12
C1	-0.04	0.17	<b>0.55</b>	0.10	0.25
C2	0.03	0.16	<b>0.64</b>	0.19	0.17
C3	-0.07	0.14	<b>0.55</b>	0.18	0.05
C4	<b>0.31</b>	-0.22	<b>-0.64</b>	-0.12	-0.17
C5	<b>0.36</b>	<b>-0.32</b>	<b>-0.60</b>	-0.13	-0.02
E1	0.09	<b>-0.57</b>	-0.09	-0.19	-0.10
E2	0.31	<b>-0.72</b>	-0.25	-0.22	-0.08
E3	-0.03	<b>0.59</b>	0.22	<b>0.38</b>	<b>0.32</b>
E4	-0.17	<b>0.69</b>	0.25	<b>0.40</b>	-0.02
E5	-0.04	<b>0.56</b>	0.42	0.22	0.27
N1	<b>0.76</b>	-0.09	-0.15	-0.20	-0.08
N2	<b>0.74</b>	-0.12	-0.14	-0.18	-0.02
N3	<b>0.74</b>	-0.18	-0.17	-0.04	0.00
N4	<b>0.63</b>	<b>-0.43</b>	-0.29	-0.06	0.04
N5	<b>0.55</b>	-0.22	-0.13	0.07	-0.14
O1	-0.02	0.28	0.21	0.15	<b>0.52</b>
O2	0.19	-0.06	-0.15	0.04	<b>-0.45</b>
O3	0.00	<b>0.36</b>	0.20	0.23	<b>0.62</b>
O4	0.25	-0.16	-0.02	0.16	<b>0.37</b>
O5	0.09	-0.07	-0.12	-0.07	<b>-0.54</b>
SS loadings	2.89	3.40	2.71	2.35	1.68
Prop Var	0.12	0.14	0.11	0.09	0.07
Prop explained	0	0	0	0	0

**Tabelle 10:** Faktorenstruktur nach Promax-Rotation und Hauptachsen-Extraktion

Maximum-Likelihood-Extraktion								
Variable	ML2	ML1	ML3	ML5	ML4	h2	u2	com
A1	0.15	0.11	0.04	<b>-0.39</b>	-0.04	0.15	0.85	1.50
A2	0.05	0.10	0.08	<b>0.57</b>	0.00	0.40	0.60	1.12
A3	0.04	0.20	0.01	<b>0.62</b>	-0.01	0.51	0.49	1.22
A4	-0.02	0.08	0.20	<b>0.43</b>	-0.17	0.29	0.71	1.84
A5	-0.09	0.28	-0.04	<b>0.55</b>	0.01	0.48	0.52	1.56
C1	0.06	-0.05	<b>0.54</b>	-0.01	0.16	0.32	0.68	1.21
C2	0.14	-0.11	<b>0.66</b>	0.08	0.06	0.43	0.57	1.20
C3	0.04	-0.10	<b>0.58</b>	0.09	-0.06	0.32	0.68	1.13
C4	0.15	0.03	<b>-0.64</b>	0.05	-0.04	0.47	0.53	1.14
C5	0.19	-0.09	<b>-0.56</b>	0.01	0.09	0.43	0.57	1.34
E1	-0.05	<b>-0.65</b>	0.16	-0.01	-0.04	0.37	0.63	1.15
E2	0.12	<b>-0.71</b>	0.05	-0.02	-0.01	0.55	0.45	1.07
E3	0.07	<b>0.48</b>	-0.07	0.24	0.25	0.44	0.56	2.15
E4	-0.01	<b>0.61</b>	-0.04	0.27	-0.11	0.52	0.48	1.48
E5	0.16	<b>0.48</b>	0.24	-0.01	0.17	0.40	0.60	2.00
N1	<b>0.86</b>	0.22	0.04	-0.27	-0.10	0.71	0.29	1.37
N2	<b>0.83</b>	0.17	0.05	-0.25	-0.04	0.66	0.34	1.27
N3	<b>0.72</b>	0.00	-0.01	-0.02	-0.01	0.53	0.47	1.00
N4	<b>0.49</b>	<b>-0.32</b>	-0.10	0.05	0.08	0.48	0.52	1.93
N5	<b>0.50</b>	-0.16	0.03	0.15	-0.15	0.34	0.66	1.59
O1	0.02	0.15	0.03	0.02	<b>0.51</b>	0.32	0.68	1.20
O2	0.17	0.02	-0.08	0.15	<b>-0.45</b>	0.24	0.76	1.62
O3	0.04	0.24	-0.04	0.08	<b>0.59</b>	0.47	0.53	1.37
O4	0.16	-0.26	-0.02	0.17	<b>0.38</b>	0.26	0.74	2.62
O5	0.10	0.02	-0.03	0.05	<b>-0.52</b>	0.27	0.73	1.10
SS loadings	2.65	2.39	1.94	1.88	1.49			
ML2	1.00	-0.27	-0.23	0.01	0.03			
ML1	-0.27	1.00	0.38	0.34	0.15			
ML3	-0.23	0.38	1.00	0.24	0.21			
ML5	0.01	0.34	0.24	1.00	0.19			
ML4	0.03	0.15	0.21	0.19	1.00			

**Tabelle 11:** Faktorenmuster nach Promax-Rotation und Maximum-Likelihood-Extraktion



Maximum-Likelihood-Extraktion								
Variable	ML2	ML3	ML5	ML1	ML4	h2	u2	com
A1	0.20	0.04	<b>-0.36</b>	-0.14	-0.04	0.15	0.85	2.00
A2	-0.02	0.09	<b>0.60</b>	0.01	0.03	0.40	0.60	1.05
A3	-0.03	0.03	<b>0.67</b>	-0.07	0.04	0.51	0.49	1.04
A4	-0.06	0.20	<b>0.46</b>	-0.04	-0.15	0.29	0.71	1.65
A5	-0.14	0.00	<b>0.58</b>	-0.17	0.06	0.48	0.52	1.33
C1	0.06	<b>0.53</b>	0.00	0.05	0.16	0.32	0.68	1.23
C2	0.13	<b>0.64</b>	0.11	0.13	0.06	0.43	0.57	1.24
C3	0.04	<b>0.56</b>	0.11	0.08	-0.06	0.32	0.68	1.14
C4	0.12	<b>-0.64</b>	0.06	0.04	-0.03	0.47	0.53	1.10
C5	0.14	<b>-0.57</b>	0.01	0.16	0.10	0.43	0.57	1.35
E1	-0.09	0.10	-0.10	<b>0.57</b>	-0.11	0.37	0.63	1.26
E2	0.06	-0.03	-0.09	<b>0.67</b>	-0.07	0.55	0.45	1.08
E3	0.06	-0.02	0.30	<b>-0.34</b>	<b>0.31</b>	0.44	0.56	3.03
E4	0.00	0.01	<b>0.36</b>	<b>-0.53</b>	-0.05	0.52	0.48	1.79
E5	0.18	0.27	0.08	<b>-0.39</b>	0.22	0.40	0.60	3.06
N1	<b>0.85</b>	0.01	-0.09	-0.09	-0.05	0.71	0.29	1.05
N2	<b>0.82</b>	0.02	-0.08	-0.04	0.01	0.66	0.34	1.03
N3	<b>0.66</b>	-0.06	0.10	0.14	0.03	0.53	0.47	1.16
N4	<b>0.41</b>	-0.16	0.09	<b>0.42</b>	0.08	0.48	0.52	2.45
N5	<b>0.44</b>	-0.02	0.22	0.25	-0.14	0.34	0.66	2.39
O1	-0.01	0.06	0.02	-0.06	<b>0.53</b>	0.32	0.68	1.05
O2	0.16	-0.10	0.21	-0.03	<b>-0.44</b>	0.24	0.76	1.88
O3	0.01	0.00	0.09	-0.10	<b>0.63</b>	0.47	0.53	1.09
O4	0.08	-0.04	0.15	<b>0.36</b>	<b>0.38</b>	0.26	0.74	2.41
O5	0.11	-0.05	0.10	-0.07	<b>-0.52</b>	0.27	0.73	1.23
SS loadings	2.49	2.05	2.1	2.07	1.64			
ML2	1.00	-0.21	-0.03	0.24	-0.01			
ML3	-0.21	1.00	0.20	-0.22	0.20			
ML5	-0.03	0.20	1.00	-0.31	0.23			
ML1	0.24	-0.22	-0.31	1.00	-0.17			
ML4	-0.01	0.20	0.23	-0.17	1.00			

**Tabelle 12:** Faktorenmuster nach Direct-Oblimin-Rotation und Maximum-Likelihood-Extraktion

## B CD

### Inhaltsverzeichnis:

1. `thesis.pdf` - Bachelorarbeit im PDF Format
2. `B.Sc_Shiny.R` - R Code der Shiny App aus Kapitel 3
3. `B.Sc_Script.R` - R Code der Analyse aus Kapitel 4
4. `2017.csv` - Datensatz World Happiness Report

## **Erklärung zur Urheberschaft**

Hiermit erkläre ich, Raphael Wilhelm Hanke, dass ich die vorliegende Arbeit noch nicht für andere Prüfungen eingereicht habe. Ich habe die Arbeit selbständig verfasst. Sämtliche Quellen einschließlich Internetquellen, die ich unverändert oder abgewandelt wiedergegeben habe, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, habe ich als solche kenntlich gemacht. Ich bin mir darüber bewusst, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung eingeleitet wird.

Berlin, 08. Februar 2018

Raphael Wilhelm Hanke