

# Einführung in faktorenanalytische Verfahren für die Persönlichkeits- und Differentielle Psychologie

Christian Blötner, Lisa-Marie Segbert und Andreas Mokros

Wintersemester 2023 / 2024

**Disclaimer:** Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

## 1 Vorbemerkungen

Nach der Prüfungsordnung für den B.Sc.-Studiengang Psychologie wird empfohlen, Modul M7 nach Modul M6a zu belegen: “Trotz der Wahlfreiheit in der Abfolge des Studiums wird für den Studienabschnitt 2 ‘Grundlagen und Forschungspraxis’ allerdings empfohlen, das Modul M6a vor dem Modul M7 zu studieren, da in Modul M6a Grundlagen der Test- und Fragebogenkonstruktion vermittelt werden, die in dem Modul M7 vertieft werden.” (entnommen von der Modulseite: <https://www.fernuni-hagen.de/KSW/portale/bscpsy/studium/ws-202324/modul-7/>)

Das vorliegende Dokument wird daher nicht auf alle Einzelheiten der Faktorenanalyse eingehen, sondern einen Fokus auf diejenigen Aspekte legen, die für das grundlegende Verständnis persönlichkeits- und differentialpsychologischer Forschung sowie diagnostischer Praxis essentiell sind. Einige wesentliche Inhalte der Faktorenanalyse werden dabei nur rekapituliert. Es wird daher empfohlen, im Falle von Unklarheiten die Prüfungsliteratur des Moduls M6a zu konsultieren.

In der Statistik werden verschiedene Methoden zur Dimensionalitätsreduktion verwendet. Die *Hauptkomponentenanalyse* reduziert Daten korrelierter Variablen auf weniger unkorrelierte Hauptkomponenten. Die Reduktion findet statt, ohne Messfehler zu berücksichtigen und ohne die Aufdeckung einer internen Struktur anzustreben (Schmitt, 2011). Ziel der *Hauptachsenfaktorenanalyse* ist die Identifikation einer geringen Anzahl latenter Variablen aus einer größeren Anzahl von manifesten Variablen. Dabei wird das Vorhandensein eines Messfehlers in psychologischen Messungen anerkannt und der Fokus auf die

Identifikation von Variablen gelegt, die diese manifesten Zusammenhänge erklären. Um dem Anspruch Rechnung zu tragen, Zusammenhänge zwischen interessierenden manifesten Variablen zu erklären, behandeln die nachfolgenden Abschnitte nur “echte” Faktorenanalysen (Hauptachsenfaktorenanalysen). Zur Vertiefung der mathematischen Grundlagen sei auf etablierte methodische Lehrbücher verwiesen, wie Bortz und Schuster (2010), Bühner (2021) und Eid et al. (2017).

Weiterhin möchten wir Ihnen anhand exemplarischer Inhalte des Moduls und der Programmiersprache R demonstrieren, wie verschiedene faktorenanalytische Verfahren angewandt und interpretiert werden. Sämtliche Abbildungen im vorliegenden Lehrtext wurden in R erstellt und sind deshalb in englischer Sprache gehalten. Zur Vertiefung von R eignet sich ganz besonders das entsprechende Lehrbuch von Maike Luhmann, welches mittlerweile in der fünften Auflage vorliegt (Luhmann, 2020). Konkrete R-Befehle oder -Programmierungen werden explizit *nicht* Gegenstand der Prüfung werden.

## 2 Exploratorische Faktorenanalyse

### 2.1 Grundbegriffe und Theoriesprache

Nicht nur in unserer Disziplin der Psychologie besteht großes Interesse an Merkmalen, die nicht direkt beobachtet werden können. Sowohl in der Psychologie als auch in anderen Wissenschaften müssen die unsichtbaren, *latenten* Merkmale zunächst anhand geeigneter Verfahren messbar gemacht werden. *Messung* im psychologischen Sinne bedeutet dabei die Zuweisung eines numerischen Relativs zu einem empirischen Relativ (Hagemann et al., 2022, S. 44). Gemeint ist damit der Ausdruck unterschiedlicher Ausprägungen eines Merkmals durch unterschiedliche Zahlen. So lässt sich das häufig kategorial oder diskret verstandene Geschlecht anhand der Zahlen 0, 1 und 2 für die Ausprägungen *männlich*, *weiblich* und *divers* kodieren. Ebenso lassen sich für die drei kategorialen oder diskreten Variablen aber auch die Zahlen 621, -445 und 9 vergeben. Beiden illustrierten Möglichkeiten zur Kodierung ist gemein, dass sie nur qualitative Unterschiede abbilden (Personen mit unterschiedlichen Kodierungen weisen unterschiedliche Ausprägungen im Merkmal auf), aber keine quantitativen Unterschiede (keine der Kodierungen wird als “besser”, “stärker” oder “wichtiger” angesehen). Bei kontinuierlichen Merkmalen wird hingegen eine Rangfolge oder Hierarchie angestrebt, sodass höhere Zahlen höhere Ausprägungen bedeuten, also quantitative Unterschiede im Zentrum der Betrachtung stehen: Jemand, der ein höheres numerisches Relativ im Merkmal *Körpergröße* aufweist, ist empirisch größer als eine Person mit einem niedrigeren zugewiesenen numerischen Relativ.

Im Folgenden behandeln wir schrittweise die geometrischen und statistischen Grundlagen, um das Konzept der Faktorenanalyse zu erörtern. Wir beginnen mit konzeptuellen Überlegungen zum Begriff des Messens im psychologischen Sinn und der Logik des Aufbaus gängiger Messinstrumente. Dazu illustrieren wir Zusammenhänge zwischen zwei Merkmalen sowohl statistisch (als Korrelationskoeffizient) als auch räumlich

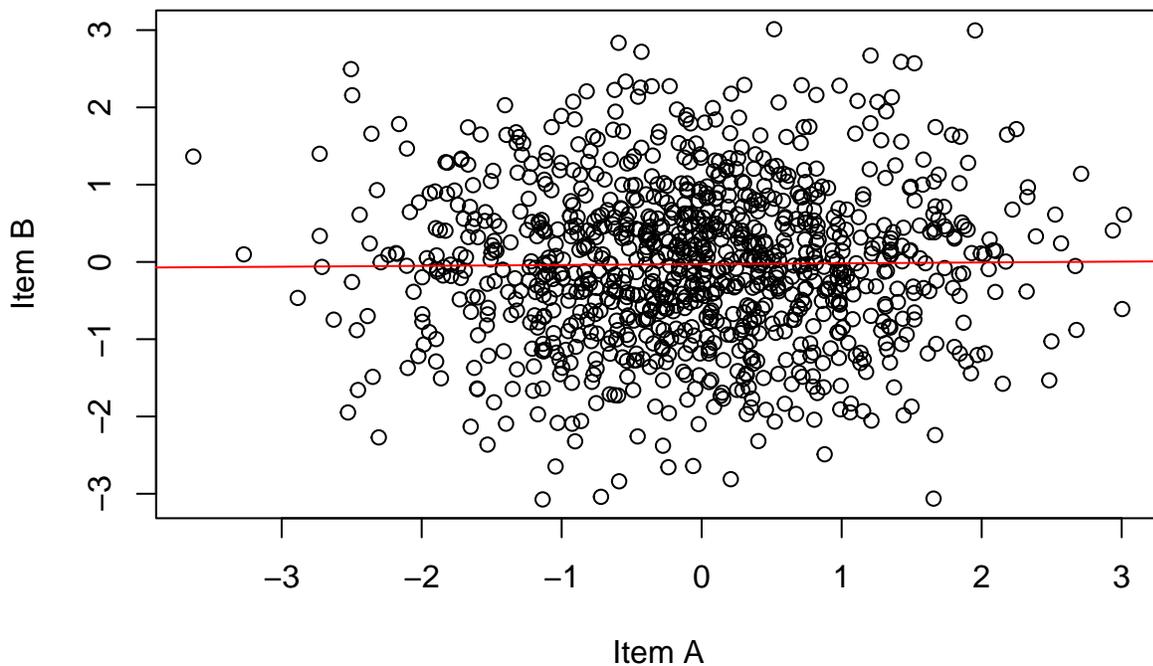
(in einem Streudiagramm). Anschließend werden paarweise Korrelationen zwischen drei Variablen gezeigt, um sowohl statistisch als auch visuell zu verdeutlichen, woran festgemacht werden kann, ob diese Variablen gleiche oder unterschiedliche Merkmale zu erfassen scheinen. Auf diesen Überlegungen aufbauend nähern wir uns nach und nach den Durchführungen konkreter Faktorenanalysen an, wobei die zugehörigen Schritte jeweils einzeln erläutert und kommentiert werden. Die hier behandelten Schritte beziehen sich dabei auf die grundlegende Frage nach der Anzahl der latenten Faktoren im Datensatz, wobei wir drei prominente Verfahren zur Feststellung kurz beleuchten (Eigenwertverlauf, *Scree*-Plot, Parallelanalyse; Abschnitt 2.3), relevante Kennzahlen zur Beurteilung der Güte von Faktoren und Items (Abschnitt 2.4) und die Festlegung darauf, ob und wie stark die extrahierten Faktoren untereinander korreliert sein dürfen (Abschnitt 2.5). Wir schließen das Kapitel mit einer kritischen Reflexion und praktischen Empfehlungen (Abschnitt 2.6).

## 2.2 Faktorraum

Zum Zweck psychologisch verstandener Messungen kommen zum großen Teil Fragebogen zum Einsatz (entweder im Selbst- oder im Fremdbereich). Es wird dabei davon ausgegangen oder zumindest angestrebt, dass die beobachteten Antworten auf die Elemente eines Fragebogens (sog. *manifeste* Indikatoren oder Items) ausdrücken, wie stark das latente Merkmal vorliegt. In den seltensten Fällen werden wir uns dabei mit Messinstrumenten zufriedengeben, die nur ein Item beinhalten (sog. *Single-Item-Measures*), sondern längere Skalen nutzen wollen. Dies ist darauf zurückzuführen, dass psychologische Merkmale meist sehr komplex sind; sie lassen sich daher oft nicht prägnant mit einer einzigen Frage zufriedenstellend abbilden. Ferner bestehen bei einem Messinstrument mit nur einem Item nur wenige Differenzierungsmöglichkeiten: Angenommen, wir legten Versuchspersonen ein Item vor, das sie fragt, wie gerne sie sich in Gesellschaft anderer Menschen befinden (*Extraversion*) und nutzten dabei ein fünfstufiges Antwortformat (1 = *überhaupt nicht gerne*, 5 = *sehr gerne*). Ausgehend von den möglichen Antworten gäbe es nur fünf Abstufungen, wobei davon auszugehen ist, dass das Bedürfnis, sich in Gesellschaft anderer Personen aufzuhalten, nuancierter ist als diese bloß fünfstufige Einteilung. Um ein weiteres Spektrum von Differenzierungsmöglichkeiten zu gewährleisten, werden interessierende Merkmale meist mit mehreren Items gemessen. Das *NEO-Persönlichkeitsinventar - revidierte Fassung* (Ostendorf & Angleitner, 2004) erfasst Extraversion etwa mit 48 Items. Durch die Aufsummierung aller Items je zu erfassender Dimension erhöht sich so die Anzahl möglicher Ausprägungen des Merkmals auf 193 mögliche Werte (Minimum = 48, wenn bei allen 48 Items die Antwortoption 1 gewählt wird; Maximum = 240, wenn bei allen 48 Items die Antwortoption 5 gewählt wird). Ebenso ist die Berechnung eines Mittelwerts über alle Items einer Dimension möglich, wodurch nicht nur natürliche Zahlen zur Feststellung der Ausprägung herangezogen werden, sondern auch rationale Zahlen (sog. "Bruchzahlen", z.B. 2.54). Zugleich lässt sich durch die Aufnahme differenzierterer Fragen auch die Messgenauigkeit im Sinne der Reliabilität erhöhen (Moosbrugger & Kelava, 2020). Eine womöglich trivial erscheinende, aber zum Zwecke der Erhöhung der Messgenauigkeit hochrelevante und notwendige

Eigenschaft der einbezogenen Items besteht darin, dass sie untereinander substantiell korreliert sind: Wir gehen davon aus, dass mehrere Items das gleiche Merkmal messen, wenn sie hoch miteinander korreliert sind und unterschiedliche Merkmale messen, wenn sie höchstens moderat korreliert sind. Auf konzeptueller Ebene wird dann davon ausgegangen, dass ein dahinterliegendes *latentes Konstrukt* (etwa Extraversion) die Korrelationen zwischen den Items verursacht (Costello & Osborne, 2005). Mathematisch spricht man von einem gemeinsamen, allen beteiligten Items zugrundeliegenden *Faktor*. Perfekte Korrelationen ( $r = 1$ ) würden wir dabei jedoch weder erwarten noch wünschen. Dies begründet sich einerseits mit der Komplexität der menschlichen Psyche und andererseits mit der Überlegung, dass kein Mehrwert in der Einbeziehung von Items liegt, die *exakt* das Gleiche abbilden: Perfekt korrelierte Items enthalten dieselbe Information und sind dadurch redundant. Grafisch lassen sich die manifesten Indikatoren oder Items in einem mehrdimensionalen Raum anordnen, wobei jedes Item eine Dimension dieses  $n$ -dimensionalen Raums aufspannt (d.h. zwei Achsen bei zwei Items, drei Achsen bei drei Items etc.). Zur Veranschaulichung beginnen wir mit einem einfachen Fall und schauen uns zunächst nur zwei Items in einem entsprechend zweidimensionalen Raum an. Dieser Raum wird *Faktorraum* genannt.

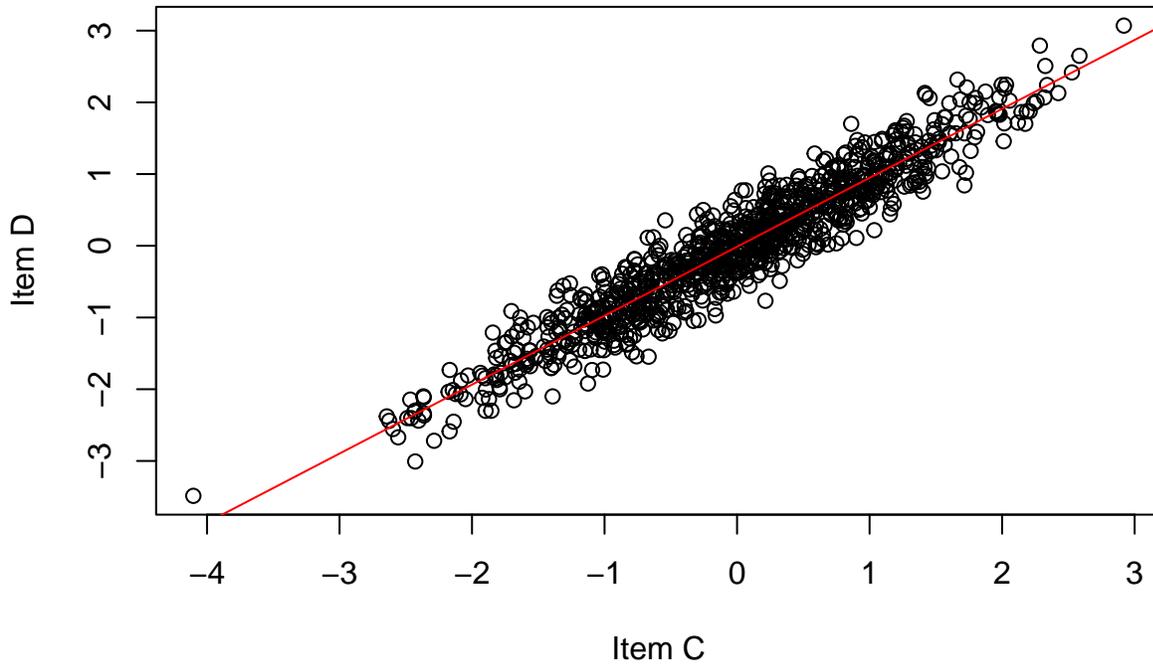
Abbildung 1



In Abbildung 1 sind die Angaben von 1,000 Personen zu den beiden Items **Item A** und **Item B** dargestellt. Jeder Kreis steht für eine Person. Der Abbildung kann entnommen werden, dass die Punktwolke recht breit ist und keiner bestimmten Systematik in der Beantwortung der Items zu folgen scheint. Die eingezeichnete Regressionsgerade verdeutlicht dies insofern, als aus dem Antwortverhalten zu **Item A** nicht ohne Weiteres auf das Antwortverhalten zu **Item B** geschlossen werden kann. Die Korrelation zwischen den beiden Items beträgt  $r = .01$ . Wir schließen daher, dass **Item A** und **Item B** unterschiedliche Merkmale erfassen und

deshalb keinen gemeinsamen Faktor konstituieren. Nachfolgend schauen wir uns zwei deutlich stärker korrelierte Items an.

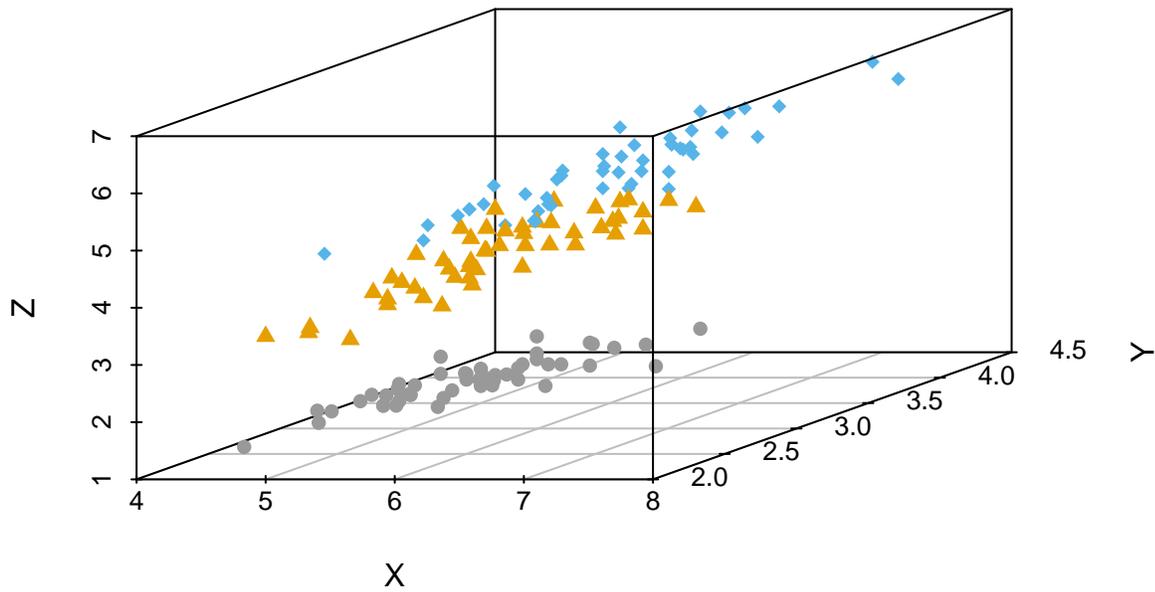
Abbildung 2



In diesem neuen Zahlenbeispiel korrelieren die beiden Items zu  $r = .95$  miteinander (siehe auch die engere Punktwolke, die insgesamt deutlich näher an der Regressionsgeraden liegt), was darauf schließen lässt, dass beide Items dasselbe latente Merkmal erfassen. Beachten Sie jedoch, dass wir dieses überspitzte Beispiel nur zu Veranschaulichungszwecken verwendet haben. Im Falle realer Messinstrumente, die über die beiden Items hinaus noch weitere Fragen enthalten, würden Forschende bei derart hochkorrelierten Items eines der beiden Items ausschließen, da es offensichtlich nur wenig Information enthält, die das jeweils andere Item nicht auch enthält.

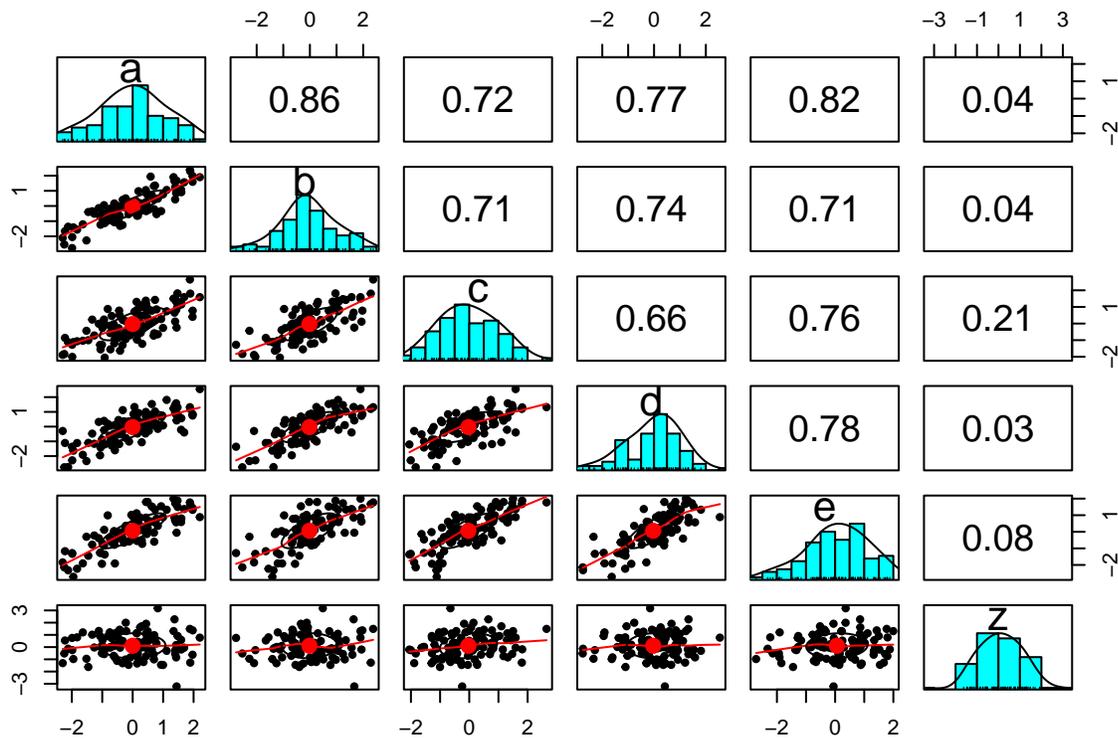
Nachfolgend fügen wir eine dritte Dimension hinzu, indem wir die Korrelationen dreier statt nur zweier Items simultan betrachten (siehe Abbildung 3). Faktorenanalysen werden zwar selten bei derart wenigen Items verwendet, jedoch lassen sich mehr als drei Items nicht mehr vergleichbar einfach in einer gemeinsamen,  $n$ -dimensionalen Grafik visualisieren.

Abbildung 3



Wir sehen in Abbildung 3, dass die Antworten zu Items, die durch orangefarbene Dreiecke und hellblaue Rauten gekennzeichnet sind (Items a und b), meist sehr nahe beieinander liegen, während die Antworten zum hellgrauen Kreisen (Item z) mehrheitlich weit von den orangefarbenen Dreiecken und hellblauen Rauten entfernt sind. Wir gehen aufgrund der relativen räumlichen Nähe der Itemantworten für die Items a, b und z davon aus, dass die Items a und b hoch miteinander korreliert, aber beide relativ unabhängig von Item z sind. Wir nehmen ferner an, dass die durch orangefarbene Dreiecke und hellblaue Rauten visualisierten Itemantworten eine gemeinsame "Ursache" haben (d.h. ein gemeinsames latentes Konstrukt), aber eine andere Ursache aufweisen als die durch hellgraue Kreise dargestellten Itemantworten. Die nachfolgende Grafik verdeutlicht dies auf unterschiedliche Arten.

Abbildung 4



Um den Kontrast zu verstärken, dass einige Items eine gemeinsame Ursache aufweisen und deshalb einen gemeinsamen Faktor abbilden, während ein anderes Item weitgehend unabhängig von den anderen Items ist, haben wir zusätzlich zu den Items a, b und z noch die Items d und e in den Datensatz aufgenommen. In der obigen Abbildung 4 finden sich unterhalb der von oben links nach unten rechts verlaufenden Diagonalen bivariate Streudiagramme, welche die bivariaten Korrelationen zwischen den Items darstellen. Die Histogramme auf der Diagonalen sind für unsere Zwecke nicht von Bedeutung. Abgesehen von der unteren Zeile finden sich durchweg sehr enge Zusammenhänge, da die Punktwolken recht "kompakt" beziehungsweise eng umgrenzt sind (siehe auch die eingezeichneten roten Regressionsgeraden in den Streudiagrammen). Die Korrelationen mit dem letzten Item, welche in der unteren Zeile durch Streudiagramme und Regressionsgeraden visualisiert und in der letzten Spalte anhand von Korrelationskoeffizienten quantifiziert werden, sind vergleichsweise kleiner und mehrheitlich annähernd null. Die Kennzahlen oberhalb der Diagonalen verdeutlichen dies, weil die Items a bis e Interkorrelationen zwischen .66 und .86, mit z jedoch nur Zusammenhänge zwischen .03 und .21 aufweisen. Wir schließen daraus, dass die ersten fünf Items ein gemeinsames Merkmal erfassen (hohe Korrelationen), während das sechste Item ein anderes Merkmal misst als die anderen Items (vorwiegend Nullkorrelationen). Die Beurteilung der Struktur von sechs Items anhand bivariater Korrelationen fällt noch recht leicht. In realen Datensätzen sind die Korrelationen jedoch nicht so stark ausgeprägt wie im vorliegenden, idealisierten Datensatz. Zudem werden in der Regel deutlich mehr Items berücksichtigt. Daher bedarf es für gewöhnlich komplexerer Verfahren, um Fragen nach der Anzahl der Faktoren und der Zugehörigkeit von Items zu Faktoren zu beantworten.

## 2.3 Anzahl zu extrahierender Faktoren

In der Literatur werden viele Verfahren diskutiert, die zur Feststellung der Anzahl zu extrahierender Faktoren herangezogen werden können. Wir beschränken uns hier auf drei recht einfache und weitverbreitete Verfahren, nämlich den Eigenwertverlauf, den *Scree*-Plot und die Parallelanalyse. Wir wenden diese auf einen realen Datensatz an, um aufzuzeigen, dass verschiedene Verfahren unterschiedliche Faktorstrukturen nahelegen können und die inhaltliche Interpretation nicht überflüssig machen. Es sei in der Hinsicht vorweggenommen, dass die drei Verfahren mitunter sehr unterschiedliche Schlussfolgerungen zulassen, sodass die jeweils suggerierten Faktorenzahlen also nie als “objektiv korrekt” bewertet werden können. Die Verfahren sollen daher eher die persönliche Entscheidungsfindung nur unterstützen, müssen aber als nachrangig behandelt werden, wenn eine Interpretation der Struktur inhaltlich und logisch nicht möglich erscheint (siehe auch Abschnitt 2.6 zu Abwägungen).

Es handelt sich hierbei um einen Datensatz, in dem ein Inventar zur Messung der *Big Five* der Persönlichkeit angewandt wird (Datensatz `bfi` aus Revelle, 2023). Jeder der fünf Faktoren wird mit fünf Items erfasst. Anhand des Datensatzes können wir eindrücklich die statistisch suggerierten Schlüsse zur Faktorenzahl (üblicherweise mit  $k$  notiert) mit unseren theoretischen Annahmen zur Anzahl der Faktoren (d.h.  $k = 5$ ) kontrastieren.

### 2.3.1 Eigenwertverlauf

Der Eigenwert eines Faktors  $\eta$  bezeichnet die Summe der quadrierten Ladungen aller Items  $\lambda_i$  auf diesen Faktor. Ladungen sind dabei Semipartialkorrelationen des Items mit dem Faktor (d.h. dass Korrelationen zwischen den Items statistisch kontrolliert werden) und sollten idealerweise größer als  $|.30|$  sein. In dem Fall erklärt der Faktor mindestens 10% der beobachteten Varianz des Items ( $|.30|^2 \approx .10$ ). Demnach quantifiziert der Eigenwert das Potential eines Faktors, Unterschiede im Beantwortungsverhalten abzubilden. Die Idee hinter der Verwendung des Eigenwerte-Kriteriums besteht darin, nur diejenigen Faktoren zu extrahieren, die mehr Varianz aufklären als ein standardisiertes Item. Um das zu gewährleisten, sollen substantielle Faktoren Eigenwerte größer als 1 aufweisen (siehe auch Eid et al., 2017). Allgemein werden von gängigen Statistikprogrammen jedoch so viele Eigenwerte berechnet wie Items vorgegeben werden, im vorliegenden Fall 25. Im Extremfall (d.h. wenn keine Gruppenfaktoren vorliegen) werden so viele Faktoren wie Items nahegelegt.

```
R <- cor(data_BFI[1:25], use = "pairwise")
round(eigen(R)$values, 3)
```

```
## [1] 5.069 2.762 2.153 1.892 1.518 1.079 0.831 0.805 0.714 0.702 0.681 0.649
## [13] 0.631 0.588 0.566 0.545 0.520 0.494 0.483 0.443 0.429 0.407 0.389 0.385
```

## [25] 0.268

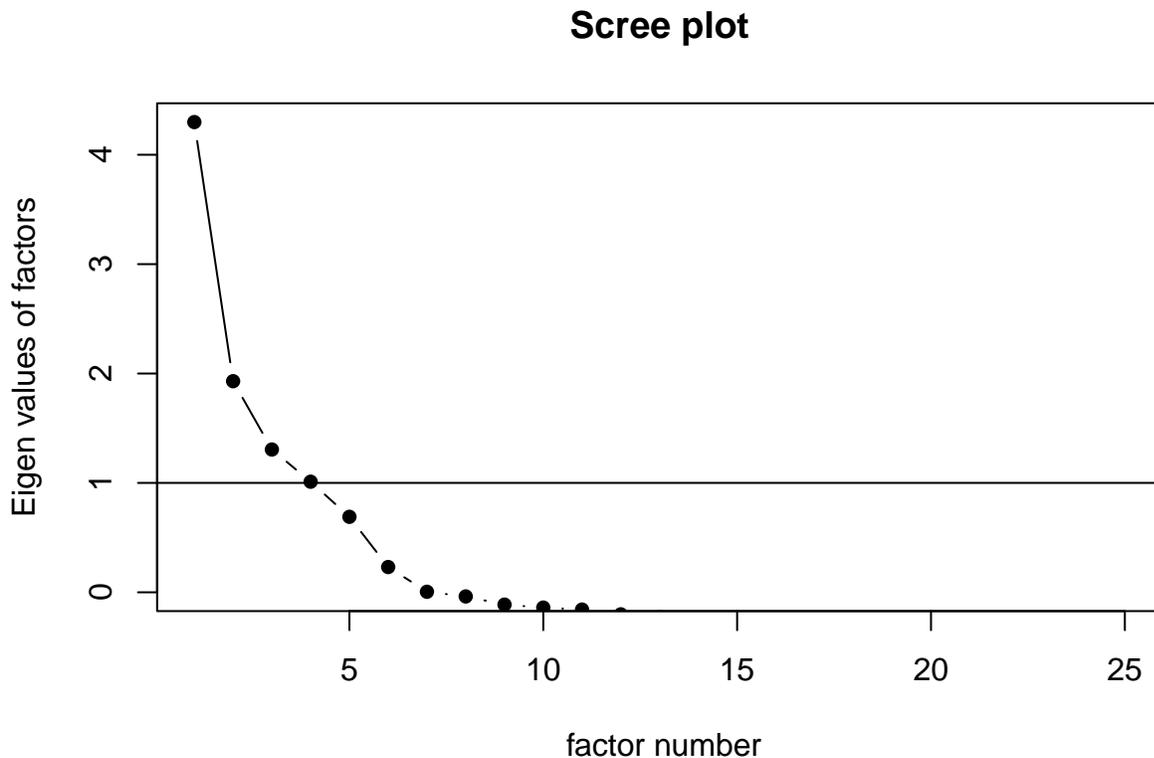
Im vorliegenden Fall müssten gemäß Eigenwerte-Kriterium sechs Faktoren extrahiert werden, da die ersten sechs Faktoren Eigenwerte über 1 aufweisen. Allerdings neigt das Eigenwerte-Kriterium zur Überschätzung der wahren Anzahl zugrundeliegender Faktoren (Zwick & Velicer, 1986) und erscheint insgesamt die am wenigsten belastbaren Ergebnisse zu liefern (Costello & Osborne, 2005). Das Eigenwerte-Kriterium sollte daher durch elaboriertere Verfahren ergänzt werden.

### 2.3.2 Scree-Plot

Der *Scree*-Test ist eine grafische Umsetzung des Eigenwertverlaufs. Hierbei werden auf der X-Achse so viele Faktoren abgetragen wie Items analysiert werden. Das heißt, dass im Extremfall jedes Item als eigener Faktor beziehungsweise als eigenes latentes Konstrukt aufgefasst werden muss. Auf der Y-Achse werden die Eigenwerte abgetragen. Der Verlauf der beobachteten Eigenwerte der Faktoren wird durch eine verbindende Linie angezeigt. Im Scree-Plot wird dann die Stelle gesucht, an der diese Linie einen markanten Knick aufweist. Es wird dann diejenige Zahl an Faktoren extrahiert, die links vom Knick liegen. Aufgrund der Subjektivität des Schlusses auf das Vorliegen oder Nicht-Vorliegen eines solchen Knicks, ist der Scree-Plot ungeeignet zur alleinigen Festlegung der Faktorenzahl. Übereinstimmend damit fanden Crawford und Coopman (1979) nur sehr geringe Beurteilerübereinstimmungen bei der Anwendung von *Scree*-Tests.

Abbildung 5

```
scree(data_BFI[1:25], pc = FALSE)
```



Im hier veranschaulichten Fall in Abbildung 5 wäre es etwa denkbar, zwischen zwei und vier Faktoren zu extrahieren (wobei auch andere Interpretationen der Lokation des Knicks denkbar sind). Ganz offensichtlich ist diese Interpretation höchst subjektiv, was die generelle Eignung des *Scree*-Plots als alleiniges Kriterium zur Festlegung der Anzahl zu extrahierender Faktoren infragestellt.

### 2.3.3 Parallelanalyse

Die Parallelanalyse stellt eine Art inferenzstatistische Umsetzung der Festlegung der Faktorenzahl dar. Hierbei handelt es sich um den Vergleich des beobachteten Eigenwertverlaufs mit einem simulierten Eigenwertverlauf, der auf der gleichen Anzahl von Items und Versuchspersonen basiert. Es werden üblicherweise mindestens 1,000 Datensätze simuliert und die jeweils pro Datensatz gefundenen Eigenwerte gemittelt.

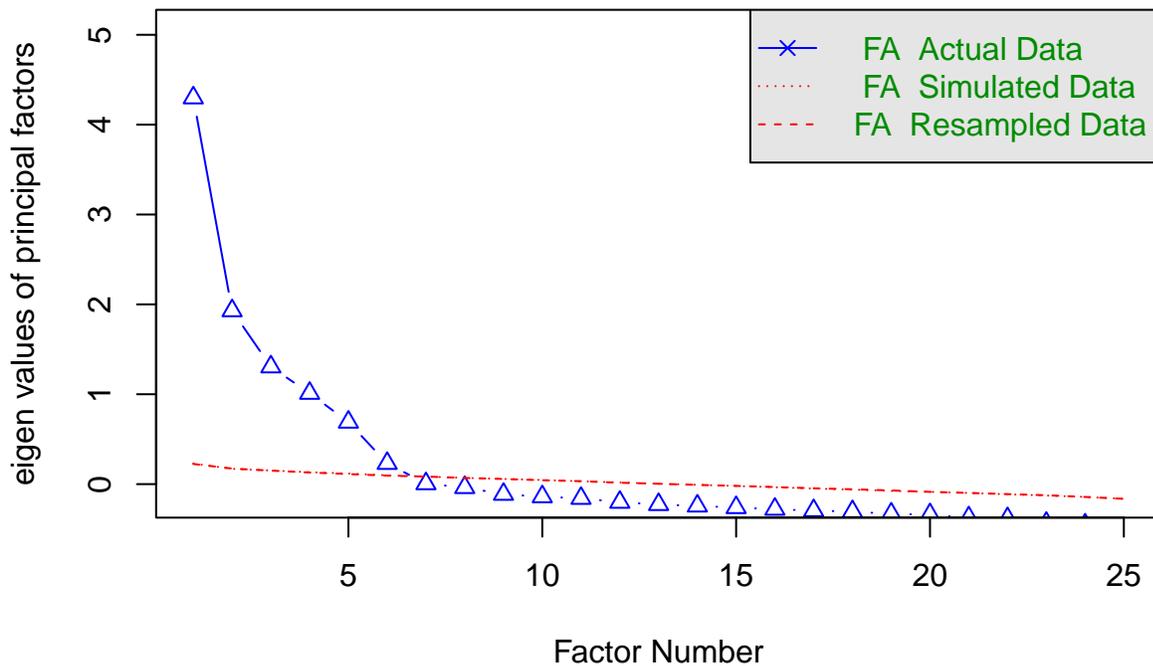
Bei simulierten Daten werden für gewöhnlich geringe Korrelationen zwischen Variablen erwartet, da es sich um Zufallswerte handelt, während die beobachteten Daten auf Items basieren, die gezielt mehr oder weniger hoch miteinander korrelieren sollten. In Analogie zur Inferenzstatistik sollten die realen Daten überzufällig starke Zusammenhänge aufweisen (Schmitt, 2011). Es wird daher nur diejenige Anzahl an Faktoren im realen Datensatz extrahiert, die größere Eigenwerte aufweisen als die Eigenwerte des korrespondierenden

simulierten Verlaufs (der im Wesentlichen einen Verlauf von Zufallszahlen darstellt). Die unkomplizierte Interpretierbarkeit dieser Faktoren lässt sich jedoch auch mit der Parallelanalyse nicht garantieren, da mit steigendem Stichprobenumfang der simulierte Eigenwertverlauf flacher wird. Konkret finden sich bei einem größeren Stichprobenumfang tendenziell stärkere simulierte Eigenwerte, die weniger steil abfallen und damit dazu neigen, eine höhere Zahl zu extrahierender Faktoren nahezu legen (siehe Bühner, 2021, für eine ausführliche Diskussion des Ansatzes und für Evidenz aus Simulationsstudien zur Präzision der Parallelanalyse). Nachfolgend finden Sie die grafische Darstellung und die verbale Ausformulierung der Parallelanalyse für ein *Big Five*-Inventar, für welches sechs Faktoren nahegelegt werden (siehe Abbildung 6).

Abbildung 6

```
fa.parallel(data_BFI[1:25], fa = "fa")
```

### Parallel Analysis Scree Plots



```
## Parallel analysis suggests that the number of factors = 6 and the number of components = NA
```

## 2.4 Kennzahlen der Exploratorischen Faktorenanalyse

Bevor wir die Verwendung verschiedener Rotationsverfahren exemplarisch am *Big Five*-Datensatz demonstrieren, sollen zunächst wichtige Kenngrößen erläutert werden. Gleichung 1 illustriert die Berechnung der *Kommunalität*  $h_j^2$ .

$$h_j^2 = \sum_{i=1}^k \lambda_{ij}^2 \quad (1)$$

Die Kommunalität eines Items drückt den Varianzanteil des Items  $j$  aus, der durch die extrahierten Faktoren  $i \dots k$  gemeinsam erklärt wird. Zur Berechnung der Kommunalität werden die Ladungen  $\lambda_{ij}$  des Items  $j$  auf die Faktoren  $i \dots k$  quadriert und anschließend aufsummiert. Dieser Wert kann zwischen 0 und 1 liegen und sollte idealerweise möglichst hoch ausgeprägt sein, damit das Item als relevant erachtet werden kann. Üblicherweise in der Forschung gefundene Kommunalitäten nehmen Werte im Bereich  $.40 \leq h_j^2 \leq .70$  an (Costello & Osborne, 2005). Die Interpretation der Kommunalität als durch die Faktoren  $i \dots k$  erklärten Varianzanteil des Items  $j$  gilt jedoch nur im Fall orthogonaler Faktoren (siehe Abschnitt 2.5), da sich nur hier die Varianzanteile eines Items eindeutig einem Faktor zuordnen lassen. Für die Interpretation der Kommunalität als erklärter Varianzanteil bedarf es im Fall einer obliquen (schiefwinkligen) Rotation (siehe Abschnitt 2.5) der Kenntnis der Varianz-Kovarianz-Matrix der Faktoren (vgl. Hagemann et al., 2022, S. 87). Die *Uniqueness*  $u_j^2$  bildet das Gegenstück der Kommunalität zu Eins (siehe Gleichung 2) und quantifiziert den Varianzanteil, der **nicht** durch die extrahierten Faktoren aufgeklärt werden kann. Dieser Wert sollte entsprechend niedrig sein, da hohe Werte darauf hindeuten würden, dass das Item wenig mit den anderen Items gemeinsam hat.

$$u_j^2 = 1 - \sum_{i=1}^k \lambda_{ij}^2 = 1 - h_j^2 \quad (2)$$

Eine dritte Größe, die hier nicht mathematisch erörtert werden soll, ist der *Komplexitätskoeffizient*. Dieser drückt aus, wie viele Faktoren benötigt werden, um die Varianz des Items zu erklären (Revelle, 2023). Dieser Wert kann beliebig hoch werden, sollte aber idealerweise möglichst nahe bei 1 liegen, um *Einfachstruktur* zu approximieren. Von Einfachstruktur spricht man, wenn alle Items sich eindeutig einem (und nur einem!) Faktor zuordnen lassen. Dies ist der Fall, wenn alle Items hohe Ladungen von je nur einem Faktor aufweisen, während die Ladungen anderer Faktoren auf die Items Null oder zumindest vernachlässigbar sind.

Nachfolgend finden Sie die korrespondierenden Koeffizienten für unser obiges simuliertes Beispiel mit fünf stark korrelierten Items (Items **a**, **b**, **c**, **d** und **e**) und einem vom Rest weitestgehend unabhängigen Item (Item **z**). Wir möchten darauf hinweisen, dass der nachfolgende R-Code in der Standardeinstellung von obliquen Faktoren ausgeht (siehe Abschnitt 2.5). Deshalb ergeben die quadrierten und aufsummierten Faktorladungen je Item nicht exakt die angegebenen Kommunalitäten.

```
fa(data, 2)
```

```
## Loading required namespace: GPArotation
```

```
## Factor Analysis using method = minres
```

```

## Call: fa(r = data, nfactors = 2)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      MR1   MR2   h2    u2 com
## a  0.99 -0.05 0.901 0.0987 1.0
## b  0.85  0.03 0.767 0.2331 1.0
## c  0.06  0.95 0.995 0.0048 1.0
## d  0.83  0.01 0.713 0.2874 1.0
## e  0.70  0.22 0.762 0.2377 1.2
## z -0.29  0.44 0.083 0.9172 1.7
##
##
##              MR1  MR2
## SS loadings          3.03 1.20
## Proportion Var        0.50 0.20
## Cumulative Var        0.50 0.70
## Proportion Explained  0.72 0.28
## Cumulative Proportion 0.72 1.00
##
## With factor correlations of
##      MR1  MR2
## MR1 1.00 0.76
## MR2 0.76 1.00
##
## Mean item complexity = 1.2
## Test of the hypothesis that 2 factors are sufficient.
##
## df null model = 15 with the objective function = 4.7 with Chi Square = 451.89
## df of the model are 4 and the objective function was 0.19
##
## The root mean square of the residuals (RMSR) is 0.02
## The df corrected root mean square of the residuals is 0.04
##
## The harmonic n.obs is 100 with the empirical chi square 1.46 with prob < 0.83
## The total n.obs was 100 with Likelihood Chi Square = 17.83 with prob < 0.0013
##
## Tucker Lewis Index of factoring reliability = 0.88
## RMSEA index = 0.186 and the 90 % confidence intervals are 0.104 0.279

```

```

## BIC = -0.59
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors    MR1 MR2
## Multiple R square of scores with factors          0.98 1.00
## Minimum correlation of possible factor scores      0.95 1.00
## Minimum correlation of possible factor scores      0.90 0.99

```

Hier wurden zwei Faktoren zugrunde gelegt, um den substanziellen bivariaten Korrelationen der Items **a**, **b**, **c**, **d** und **e** Rechnung zu tragen, während Item **z** unabhängig von den anderen Items zu sein scheint. Es zeigt sich, dass die ersten fünf Items hohe Kommunalitäten  $h^2$  (und demnach niedrige Uniqueness-Koeffizienten  $u^2$ ) aufweisen und durch einen Faktor erklärt werden können ( $com$  nahe 1). Die Zuordnung von Item **z** ist nicht ohne Weiteres möglich, da vom Betrag recht ähnliche Ladungen zweier Faktoren vorliegen (Spalten **MR1** und **MR2**). Trotz hoher bivariater Korrelationen von Item **c** mit den Items **a**, **b**, **d** und **e** weist Item **c** eine extreme Ladung des zweiten Faktors, nicht des ersten, auf. Generell werden Items demjenigen Faktor zugeordnet, der die höchste Ladung aufweist (Bühner, 2021). Item **c** würde damit einem anderen Faktor zugeordnet als die anderen Items. Im vorliegenden Beispiel erscheinen die Items **a** und **c** dennoch problematisch, da sie so hohe Ladungen und Kommunalitäten aufweisen, dass sie auch als redundant erachtet werden können: Aufgrund der Höhe der Ladungen erscheint es so, als würden sie die jeweiligen latenten Faktoren im Wesentlichen bereits fast perfekt abbilden. Da ein Item in der Regel nur einen Ausschnitt des Bedeutungsraums eines Faktors abbilden kann und soll, würden Forschende im vorliegenden Fall zur Absicherung der inhaltlichen Breite des latenten Faktors Item **a** entfernen (siehe auch das sog. *Bandwidth-Fidelity-Dilemma*; Salgado, 2017).

Bei den Angaben unterhalb der Ladungsmatrix handelt es sich um summarische Informationen zu den Faktoren und erklärten Varianzen. So bezeichnet **SS loadings** die Summe der quadrierten Ladungen der einzelnen Items auf die Faktoren. Demnach sind die Angaben in der Zeile **SS loadings** mathematisch als Eigenwerte zu verstehen. Sie müssen aber nicht zwingend mit den beispielsweise im Eigenwerte-Verlauf oder im *Scree-Plot* abgetragenen Eigenwerten übereinstimmen. Das liegt daran, dass bei der Berechnung der eigentlichen Eigenwerte die herangezogene Anzahl von Faktoren noch nicht feststeht, in der hier durchgeführten Faktorenanalyse aber schon eine genaue Zahl zu extrahierender Faktoren vorgegeben wird. **Proportion Var** gibt den Anteil der Gesamtvarianz an, der durch den entsprechenden Faktor erklärt wird. **Cumulative Var** entspricht dabei der Summe der zuvor erläuterten einzelnen Varianzaufklärungen. Diese insgesamt aufgeklärte Varianz lässt sich wiederum untergliedern in die jeweiligen Anteile, die die Faktoren daran haben, nämlich in der Zeile **Proportion Explained**. Wir sehen hier, dass 72% (**Proportion Explained**) der insgesamt erklärten Varianz von 70% (Angabe ganz rechts in der Spalte **Cumulative Var**) auf den ersten Faktor zurückgehen und 28% auf den zweiten Faktor.

Die darunter liegende Korrelationsmatrix (`With factor correlations of`) veranschaulicht die Interkorrelationen der Faktoren. Dass in diesem Beispiel überhaupt eine Korrelationsmatrix ausgegeben wurde, ist darauf zurückzuführen, dass das verwendete Programmpaket zur Durchführung der Faktorenanalyse automatisch eine *oblique* Rotation durchgeführt hat (siehe nächster Abschnitt). Im Falle einer explizit anzufordernden *orthogonalen* Rotation entfele die Korrelationsmatrix der Faktoren. Die darunterliegenden Informationen sind für unsere Zwecke nicht von Bedeutung. Interessierte seien auf die Dokumentation des `psych`-Pakets von Revelle (2023) verwiesen.

## 2.5 Rotation: Orthogonale und oblique Faktoren

Das Prinzip der Faktorenrotation liegt darin begründet, dass die Faktoren, die in der Faktorenanalyse gesucht werden, unter Umständen eine komplexere Geometrie aufweisen als es die menschliche Kognition zulässt (z.B. weil der Faktorraum mehr als drei Dimensionen aufweist) oder im Vergleich zu dem Raum, in dem wir Menschen “navigieren”, gedreht ist. Die Faktoren, die sich zunächst als Achsen im  $n$ -dimensionalen Raum gedacht werden können, sollen möglichst nah an den Items liegen, die diese Faktoren konstituieren. Ist dies der Fall, können wir die Beziehungen einfacher interpretieren und fundiert darüber entscheiden, ob ein Item dazu geeignet ist, den Faktor zu erfassen (d.h. das Item liegt räumlich nahe am Faktor) oder ob es nicht dazu geeignet ist, den Faktor zu erfassen (d.h. das Item liegt räumlich weiter vom Faktor entfernt). Prinzipiell ist das Konzept der Rotation auch schon in unserem Eingangsbeispiel zweier hochkorrelierter Items zu sehen gewesen. Hierbei können wir uns die eingezeichnete Regressionsgerade als den zugrundeliegenden Faktor vorstellen, der gut dazu geeignet ist, die Itemantworten zu den Items zu erklären.

Die Punktwolke, die den Daten entspricht, wird durch die Faktorenrotation nicht in ihrer Gestalt geändert. Die Achsen werden lediglich relativ zur Punktwolke gedreht. Durch die tendenzielle Annäherung der Achsen an die Punktwolke erhöht sich die Interpretierbarkeit der Faktorenlösung. Konkret werden zur inhaltlichen Interpretation der Faktoren die Inhalte besonders ladungsstarker Items im Sinne des Gesamtkonstrukts herangezogen. Da die Lageverhältnisse der Items zueinander unverändert bleiben, hat die Rotation keinen Einfluss auf die insgesamt aufgeklärte Varianz der Items. Durch die Auswahl orthogonaler versus obliquen Rotation ändert sich nur die Aufteilung der Gesamtvarianz auf die einzelnen Faktoren. Um die Varianzaufklärung durch alle Faktoren zu ändern, müsste die relative Lage der Items zueinander (im Gegensatz zur Lage der Items relativ zu den Achsen) variiert werden.

Das zuletzt herangezogene Beispiel simulierter Daten war idealisiert, weil in realen Datensätzen üblicherweise deutlich mehr Items vorliegen, die auch weniger stark korreliert sind. Faktorenanalysen “echter” Daten lassen sich außerdem nicht immer einfach interpretieren und beurteilen. Nachfolgend wenden wir daher die Faktorenanalyse auf den bereits vorgestellten *Big Five*-Datensatz an und veranschaulichen die Auswirkungen verschiedener Rotationsverfahren auf die Interpretierbarkeit. Die Dimensionen *Offenheit für Erfahrung* (Kürzel O), *Gewissenhaftigkeit* (Kürzel C), *Extraversion* (Kürzel E), *Verträglichkeit* (Kürzel

A) und *Neurotizismus* (Kürzel N) werden anhand von je fünf Items auf einer sechsstufigen Skala erfasst, 1 = *Very Inaccurate*, 2 = *Moderately Inaccurate*, 3 = *Slightly Inaccurate*, 4 = *Slightly Accurate*, 5 = *Moderately Accurate*, 6 = *Very Accurate*. Die genaue Struktur des Datensatzes kann bei Revelle (2023, S. 38–39) nachvollzogen werden. Der Empfehlung folgend, etwaige Vorüberlegungen zur Faktorenzahl höher zu gewichten als die Befunde der Verfahren zur Feststellung der Faktorenzahl (Abschnitt 2.3), greifen wir nachfolgend die theoretisch fundierte fünffaktorielle Struktur des verwendeten *Big Five*-Inventars auf. Dies erleichtert auch die spätere inhaltliche Interpretierbarkeit. Interessierte seien gerne dazu ermutigt, eine eigene Faktorenanalyse des *bfi*-Datensatzes mit sechs vorgegebenen Faktoren durchzuführen.

### 2.5.1 Unrotierte Lösung

```
fa.bfi_none <- fa(R, nfactors = 5, fm = "ml", rotate = "none")
print(fa.bfi_none, sort = TRUE)

## Factor Analysis using method = ml
## Call: fa(r = R, nfactors = 5, rotate = "none", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##   item  ML1  ML2  ML3  ML4  ML5  h2  u2 com
## N4   19 -0.60  0.22  0.02  0.09  0.29  0.50  0.50  1.8
## N1   16 -0.60  0.57  0.04 -0.07 -0.18  0.72  0.28  2.2
## E2   12  0.59  0.25 -0.17  0.03 -0.33  0.55  0.45  2.2
## N2   17 -0.58  0.55  0.09 -0.03 -0.14  0.66  0.34  2.2
## E4   14  0.56  0.34 -0.28 -0.11 -0.12  0.54  0.46  2.4
## A5    5  0.53  0.29 -0.24 -0.05  0.21  0.48  0.52  2.4
## N3   18 -0.53  0.49  0.02 -0.03  0.10  0.53  0.47  2.1
## C5   10  0.48  0.01  0.35 -0.27 -0.11  0.44  0.56  2.6
## E3   13  0.46  0.44 -0.10  0.20 -0.03  0.46  0.54  2.5
## A3    3  0.45  0.40 -0.22 -0.11  0.30  0.52  0.48  3.4
## N5   20 -0.43  0.29 -0.06 -0.17  0.20  0.34  0.66  2.7
## A2    2  0.38  0.35 -0.15 -0.13  0.31  0.40  0.60  3.5
## A4    4  0.37  0.21 -0.10 -0.29  0.19  0.31  0.69  3.3
## E1   11  0.36  0.32 -0.23  0.08 -0.26  0.36  0.64  3.7
## E5   15  0.42  0.43  0.13  0.05 -0.18  0.42  0.58  2.6
## C4    9  0.44  0.02  0.50 -0.20 -0.05  0.48  0.52  2.3
## C2    7  0.25  0.24  0.49 -0.20  0.14  0.42  0.58  2.7
## C1    6  0.27  0.19  0.46 -0.06  0.06  0.33  0.67  2.1
```

```

## C3      8  0.28  0.13  0.39 -0.27  0.08  0.33  0.67  3.1
## 03     23  0.32  0.34  0.13  0.50  0.05  0.48  0.52  2.8
## 05     25  0.17  0.07  0.24  0.43  0.08  0.29  0.71  2.1
## 01     21  0.27  0.23  0.19  0.40  0.04  0.32  0.68  3.0
## 02     22  0.19 -0.05  0.29  0.39 -0.01  0.28  0.72  2.4
## 04     24 -0.13  0.16  0.12  0.29  0.32  0.24  0.76  3.2
## A1      1  0.21  0.03 -0.12  0.01  0.31  0.16  0.84  2.2
##
##
##          ML1  ML2  ML3  ML4  ML5
## SS loadings          4.39 2.39 1.54 1.27 0.93
## Proportion Var       0.18 0.10 0.06 0.05 0.04
## Cumulative Var       0.18 0.27 0.33 0.38 0.42
## Proportion Explained 0.42 0.23 0.15 0.12 0.09
## Cumulative Proportion 0.42 0.64 0.79 0.91 1.00
##
## Mean item complexity = 2.6
## Test of the hypothesis that 5 factors are sufficient.
##
## df null model = 300 with the objective function = 7.41
## df of the model are 185 and the objective function was 0.61
##
## The root mean square of the residuals (RMSR) is 0.03
## The df corrected root mean square of the residuals is 0.04
##
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##
##          ML1  ML2  ML3  ML4  ML5
## Correlation of (regression) scores with factors 0.95 0.92 0.85 0.82 0.79
## Multiple R square of scores with factors        0.90 0.84 0.73 0.67 0.63
## Minimum correlation of possible factor scores    0.80 0.68 0.45 0.34 0.26

```

In dieser als “unrotierte Lösung” bezeichneten Faktorstruktur erscheint es schwer, auf Anhieb eine Systematik in den Ladungsmustern zu erkennen. Die ersten beiden Faktoren (ML1 und ML2) laden vergleichsweise stark auf Items genuin unterschiedlicher Konstrukte (zuvorderst Neurotizismus, Extraversion, Verträglichkeit und Gewissenhaftigkeit). Der dritte Faktor (ML3) lädt primär auf Items, die Gewissenhaftigkeit abbilden sollen, wobei auch andere Faktoren moderat auf diese Items laden (sog. *Kreuz-* oder *Nebenladungen*). Beim vierten Faktor finden sich vorwiegend starke Ladungen auf Items, die Offenheit erfassen sollen (ML4), wobei auch

hier nicht-vernachlässigbare Ladungen von anderen Faktoren vorliegen. Die Kommunalitäten ( $h^2$ ) und Uniqueness-Koeffizienten ( $u^2$ ) sind recht heterogen und legen geringe Eignung mancher Items nahe. Die Komplexitätskoeffizienten ( $com$ ) suggerieren, dass mehrere Faktoren nötig sind, um die Varianzen einiger Items zu erklären ( $com > 1$ ). Die unrotierte Lösung erreicht daher keine Einfachstruktur.

## 2.5.2 Orthogonale Rotation

Bei der Verwendung eines orthogonalen Rotationsverfahrens, wie der *Varimax*-Rotation (`rotate = "varimax"`), werden Faktoren extrahiert, die untereinander jeweils unkorreliert sind. Das Antwortverhalten zu Items des einen Faktor lässt damit keine Rückschlüsse auf das Antwortverhalten in Hinblick auf Items des anderen Faktors zu. Die orthogonale Rotation weist zudem den Vorteil auf, dass die resultierenden Faktoren leichter interpretiert werden können. Das ist darauf zurückzuführen, dass die Items eindeutiger einem Faktor zugeordnet werden können (verglichen mit dem Verzicht auf eine Faktorenrotation) und Nebenladungen im Vergleich zu den Hauptladungen kleiner werden (Costello & Osborne, 2005).

```
fa.bfi_varimax <- fa(R, nfactors = 5, fm = "ml", rotate = "varimax")
print(fa.bfi_varimax, sort = TRUE)

## Factor Analysis using method = ml
## Call: fa(r = R, nfactors = 5, rotate = "varimax", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##   item  ML2  ML1  ML3  ML5  ML4  h2  u2 com
## N1   16  0.81  0.10 -0.04 -0.21 -0.08 0.72 0.28 1.2
## N2   17  0.78  0.06 -0.02 -0.20 -0.02 0.66 0.34 1.1
## N3   18  0.72 -0.08 -0.07 -0.01  0.00 0.53 0.47 1.0
## N4   19  0.56 -0.37 -0.19  0.00  0.07 0.50 0.50 2.0
## N5   20  0.52 -0.18 -0.06  0.11 -0.15 0.34 0.66 1.6
## E2   12 -0.23  0.67  0.10  0.16  0.07 0.55 0.45 1.4
## E4   14 -0.12  0.60  0.09  0.39 -0.04 0.54 0.46 1.9
## E1   11 -0.03  0.58 -0.03  0.14  0.07 0.36 0.64 1.2
## E3   13  0.02  0.50  0.08  0.33  0.31 0.46 0.54 2.5
## E5   15  0.05  0.50  0.31  0.13  0.22 0.42 0.58 2.3
## C4    9 -0.22  0.10  0.65  0.01  0.08 0.48 0.52 1.3
## C2    7  0.08 -0.01  0.62  0.14  0.13 0.42 0.58 1.2
## C5   10 -0.27  0.19  0.57  0.04 -0.04 0.44 0.56 1.7
## C3    8 -0.03  0.02  0.56  0.12  0.00 0.33 0.67 1.1
## C1    6  0.01  0.04  0.53  0.06  0.22 0.33 0.67 1.4
```

```

## A3   3  0.03  0.28  0.11  0.65  0.05  0.52  0.48  1.4
## A5   5 -0.12  0.34  0.08  0.58  0.07  0.48  0.52  1.8
## A2   2  0.04  0.20  0.14  0.58  0.04  0.40  0.60  1.4
## A4   4 -0.05  0.17  0.23  0.45 -0.13  0.31  0.69  2.0
## A1   1 -0.10 -0.04 -0.02  0.37  0.07  0.16  0.84  1.2
## 03  23  0.02  0.27  0.06  0.15  0.62  0.48  0.52  1.5
## 05  25 -0.07  0.01  0.07 -0.01  0.52  0.29  0.71  1.1
## 01  21 -0.02  0.18  0.11  0.08  0.52  0.32  0.68  1.4
## 02  22 -0.17  0.01  0.11 -0.12  0.47  0.28  0.72  1.6
## 04  24  0.21 -0.22 -0.05  0.13  0.36  0.24  0.76  2.7
##
##
##                               ML2 ML1 ML3 ML5 ML4
## SS loadings                   2.68 2.30 2.01 1.95 1.57
## Proportion Var                 0.11 0.09 0.08 0.08 0.06
## Cumulative Var                 0.11 0.20 0.28 0.36 0.42
## Proportion Explained           0.26 0.22 0.19 0.19 0.15
## Cumulative Proportion          0.26 0.47 0.67 0.85 1.00
##
## Mean item complexity = 1.6
## Test of the hypothesis that 5 factors are sufficient.
##
## df null model = 300 with the objective function = 7.41
## df of the model are 185 and the objective function was 0.61
##
## The root mean square of the residuals (RMSR) is 0.03
## The df corrected root mean square of the residuals is 0.04
##
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##
##                               ML2 ML1 ML3 ML5 ML4
## Correlation of (regression) scores with factors 0.93 0.87 0.86 0.85 0.83
## Multiple R square of scores with factors        0.86 0.75 0.74 0.72 0.69
## Minimum correlation of possible factor scores   0.73 0.51 0.48 0.43 0.39

```

Angewandt auf Vorannahmen zu den *Big Five* der Persönlichkeit ist das Modell korrekt spezifiziert, weil (a) wir die theoretisch postulierte Faktorenanzahl zugrunde gelegt haben ( $k = 5$ ) und (b) gemäß der Theorie orthogonale (also unkorrelierte) Faktoren anstreben. Allerdings stellten sich die *Big Five* empirisch als nicht

unabhängig voneinander heraus (DeYoung et al., 2002; Digman, 1997). Insgesamt erfordert die Anwendung eines orthogonalen Rotationsverfahrens häufig eine starke Rechtfertigung, weil korrelative Unabhängigkeit von Konstrukten vielfach nicht der psychologisch-statistischen Realität entspricht (Costello & Osborne, 2005). So zeigten etwa DeYoung et al. (2002) und Digman (1997), dass sich die eigentlich als unabhängig gedachten *Big Five* zu den Meta-Faktoren *Flexibilität* (hohe Extraversion und hohe Offenheit) und *Stabilität* (niedriger Neurotizismus, hohe Gewissenhaftigkeit, hohe Verträglichkeit) faktorisieren lassen. Um latente Faktoren zu sogenannten *Faktoren höherer Ordnung* zu faktorisieren, muss zunächst die Unabhängigkeitsannahme der Faktoren aufgegeben werden und eine oblique Rotation durchgeführt werden.

### 2.5.3 Oblique Rotation

Bei der obliquen (schiefwinkligen) Rotation wird erlaubt, dass die extrahierten Faktoren korreliert sein dürfen, wobei das Ausmaß ihrer Interkorrelation beliebig hoch sein kann. Die orthogonale Rotation kann daher als Spezialfall der obliquen Rotation angesehen werden, weil eine oblique Rotation auch ergeben kann, dass die extrahierten Faktoren trotz “Erlaubnis” nicht korrelieren. Ein obliques Rotationsverfahren anstelle eines orthogonalen Verfahrens zu nutzen, ist daher nie falsch und häufig tatsächlich angemessener (Bühner, 2021; Costello & Osborne, 2005; Schmitt, 2011). Die oblique Rotation ist daher beim R-Paket `psych` (Revelle, 2023) die Standardeinstellung. Bühner (2021) empfahl als Grenzwert für die Anwendung der obliquen Rotation eine Faktorkorrelation von  $\rho = |.10|$ . Dies entspricht einem geteilten Variananteil von 1%. Eine orthogonale Rotation kann also insbesondere dann angemessen sein, wenn die Faktoren eine Korrelation unterhalb dieses Grenzwertes aufweisen. Der Nachteil bei der obliquen Rotation besteht darin, dass die Item-Faktor-Zuordnung unter Umständen erschwert ist, insbesondere, wenn die Analyse stark korrelierte Faktoren nahelegt. In solchen Fällen ist es plausibel, Items mit zu starken Doppelladungen auszuschließen.

```
fa.bfi_promax <- fa(R, nfactors = 5, fm = "ml", rotate = "promax")
print(fa.bfi_promax, sort = TRUE)

## Factor Analysis using method = ml
## Call: fa(r = R, nfactors = 5, rotate = "promax", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##   item  ML2  ML1  ML3  ML5  ML4  h2  u2 com
## N1   16  0.88  0.23  0.03 -0.26 -0.10 0.72 0.28 1.4
## N2   17  0.84  0.17  0.05 -0.25 -0.03 0.66 0.34 1.3
## N3   18  0.72 -0.01  0.00 -0.01 -0.01 0.53 0.47 1.0
## N5   20  0.50 -0.16  0.01  0.15 -0.16 0.34 0.66 1.6
## N4   19  0.49 -0.34 -0.10  0.07  0.09 0.50 0.50 2.0
## E2   12 -0.11  0.71 -0.05  0.02  0.02 0.55 0.45 1.1
```

```

## E1  11  0.06  0.64 -0.16  0.03  0.04  0.36  0.64  1.2
## E4  14 -0.03  0.61 -0.05  0.29 -0.10  0.54  0.46  1.5
## E3  13  0.08  0.51 -0.06  0.22  0.26  0.46  0.54  2.0
## E5  15  0.16  0.50  0.23  0.00  0.17  0.42  0.58  1.9
## C4   9 -0.14 -0.03  0.67 -0.06  0.03  0.48  0.52  1.1
## C2   7  0.14 -0.13  0.66  0.09  0.06  0.42  0.58  1.2
## C3   8  0.03 -0.09  0.59  0.08 -0.05  0.33  0.67  1.1
## C5  10 -0.18  0.09  0.57 -0.03 -0.09  0.44  0.56  1.3
## C1   6  0.07 -0.05  0.55  0.00  0.17  0.33  0.67  1.2
## A3   3  0.04  0.21  0.03  0.62 -0.01  0.52  0.48  1.3
## A2   2  0.05  0.12  0.08  0.56 -0.02  0.40  0.60  1.2
## A5   5 -0.10  0.28 -0.03  0.54  0.01  0.48  0.52  1.6
## A4   4 -0.02  0.09  0.19  0.43 -0.19  0.31  0.69  1.9
## A1   1 -0.14 -0.11 -0.05  0.41  0.05  0.16  0.84  1.5
## O3  23  0.04  0.27 -0.05  0.07  0.60  0.48  0.52  1.4
## O5  25 -0.09 -0.02  0.03 -0.04  0.53  0.29  0.71  1.1
## O1  21  0.00  0.16  0.04  0.01  0.51  0.32  0.68  1.2
## O2  22 -0.18 -0.02  0.08 -0.15  0.48  0.28  0.72  1.6
## O4  24  0.15 -0.25 -0.03  0.16  0.37  0.24  0.76  2.6
##
##
##          ML2  ML1  ML3  ML5  ML4
## SS loadings      2.67 2.47 1.99 1.86 1.55
## Proportion Var   0.11 0.10 0.08 0.07 0.06
## Cumulative Var   0.11 0.21 0.28 0.36 0.42
## Proportion Explained 0.25 0.23 0.19 0.18 0.15
## Cumulative Proportion 0.25 0.49 0.68 0.85 1.00
##
## With factor correlations of
##          ML2  ML1  ML3  ML5  ML4
## ML2  1.00 -0.28 -0.23 0.03 0.04
## ML1 -0.28  1.00  0.40 0.33 0.12
## ML3 -0.23  0.40  1.00 0.21 0.18
## ML5  0.03  0.33  0.21 1.00 0.15
## ML4  0.04  0.12  0.18 0.15 1.00
##
## Mean item complexity = 1.4

```

```

## Test of the hypothesis that 5 factors are sufficient.
##
## df null model = 300 with the objective function = 7.41
## df of the model are 185 and the objective function was 0.61
##
## The root mean square of the residuals (RMSR) is 0.03
## The df corrected root mean square of the residuals is 0.04
##
## Fit based upon off diagonal values = 0.98
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors ML2 ML1 ML3 ML5 ML4
## Multiple R square of scores with factors 0.93 0.91 0.88 0.87 0.84
## Minimum correlation of possible factor scores 0.87 0.82 0.78 0.75 0.71
## Minimum correlation of possible factor scores 0.74 0.64 0.57 0.50 0.41

```

Im vorliegenden Fall wird eine *Promax*-Rotation auf den *Big Five*-Beispieldatensatz angewandt (`rotate = "promax"`). Im Vergleich zur Durchführung der orthogonalen *Varimax*-Rotation unter ansonsten gleichen Bedingungen sehen wir, dass die zugelassenen Nebenladungen vergleichsweise stärker sind, die Zuordnung von Items zu Faktoren aber trotzdem noch gut möglich ist. Ausgehend von den Ladungsmustern würden wir die Faktoren ML2, ML1, ML3, ML5 und ML4 (in der Reihenfolge) als Neurotizismus, Extraversion, Gewissenhaftigkeit, Verträglichkeit und Offenheit interpretieren (da die Ladungen der entsprechenden Items auf die Faktoren am stärksten sind). Die Faktoren Extraversion (ML1) und Gewissenhaftigkeit (ML3) sind stark miteinander korreliert ( $\rho = .40$ ), während die Faktoren Neurotizismus (ML2) und Offenheit (ML4) weitestgehend unkorreliert sind,  $\rho = .04$ . Alle anderen Korrelationen liegen betragsmäßig dazwischen.

## 2.6 Fazit zur Exploratorischen Faktorenanalyse

Dieser Lehrtext kann und soll nur einen groben Überblick über das Thema Faktorenanalyse bieten, um aufzuzeigen, dass Persönlichkeitstheorien und psychologische Messinstrumente nicht ohne Weiteres und durch unmittelbare Anschauung bewertet werden können. Eine Reihe intuitiver und subjektiver Einflüsse und Entscheidungen sind bei der Durchführung faktorenanalytischer Verfahren unumgänglich (Costello & Osborne, 2005; für einen Direkt-Vergleich der Lösungen verschiedener Ansätze zur Exploratorischen Faktorenanalyse beim selben Datensatz, siehe Schmitt, 2011, S. 314). Exemplarisch sei hier auf das Fehlen objektiver Maßstäbe für den Abbruch der Faktorenextraktion hingewiesen. Wie die Beispiele gezeigt haben, legen verschiedene Verfahren verschiedene Faktorenzahlen nahe. Generell erscheint der Eigenwertverlauf zu liberal, da tendenziell mehr Faktoren als nötig extrahiert werden. Der *Scree*-Plot ist ein äußerst subjektives Verfahren, da unterschiedliche Personen den entscheidenden "Knick" an verschiedenen Stellen

sehen könnten. Die Parallelanalyse erscheint noch das statistisch belastbarste Verfahren zu sein, da es eine quasi-inferenzstatistische Absicherung beinhaltet, damit jedoch auch von den gleichen Problemen betroffen ist wie die Inferenzstatistik im Allgemeinen (d.h. bei einer ausreichend hohen Stichprobe werden auch triviale Faktoren noch als “statistisch bedeutsam” dargestellt). Ferner garantiert das statistisch elaborierte Verfahren nicht zwangsläufig eine unkomplizierte Interpretierbarkeit dieser Faktoren (unter der Annahme, dass die gewählte Faktorenzahl nachfolgend in einer adäquaten schiefwinkligen oder rechtwinkligen Analyse angewandt wird). Ein weiterer Aspekt, der ausführlicher Überlegungen bedarf, betrifft die individuelle Toleranz für die Höhe von Haupt- und Nebenladungen. Sicherlich ist es wünschenswert, im Falle einer mehrfaktoriellen Lösung die Einfachstruktur anzustreben, damit Items und Faktoren möglichst eindeutig im Sinne eines einzigen latenten Merkmals interpretiert werden können. Gerade wenn einzelne Items aber kennzeichnend für mehr als ein Merkmal sind (und dies auch so gewünscht ist), müssen Kreuzladungen auch in Kauf genommen werden. In der Regel werden entsprechende Items zwar ausgeschlossen, um die Interpretierbarkeit zu erhöhen; diese Faustregel sollte jedoch immer mit einer gewissen Differenziertheit betrachtet werden. So könnte etwa das Item “Ich bin gerne unter anderen Menschen” wahlweise zu Verträglichkeit oder zu Extraversion zugeordnet werden, da es für beide *Big Five*-Konstrukte inhaltlich passend erscheint.

Die zuletzt geschilderten Darstellungen zu Aspekten der Exploratorischen Faktorenanalyse sollen verdeutlichen, dass eine Vielzahl von Entscheidungen gravierende Auswirkungen darauf haben, ob und wie die Faktoren zu interpretieren sind, aber auch, ob und wie gut sich eine Faktorenlösung replizieren lässt. Es sollte gezeigt werden, dass die Statistik ein hilfreiches Werkzeug darstellt, aber eigene Überlegungen nicht ersetzen kann (Bühner, 2021). Im Falle eines streng exploratorischen Vorgehens (d.h. gänzliche Abwesenheit jeglicher Vorannahmen zur Dimensionalität vorliegender Item-Sets) erscheinen zunächst viele Entscheidungen gerechtfertigt. So verglichen etwa Harbin et al. (2018) bei der initialen Entwicklung eines Messinstruments zur Erfassung von schulischem *Bullying* zunächst verschiedene Faktorenlösungen, entfernten Items, die über die verschiedenen Strukturen hinweg mehrdeutig oder ungeeignet erschienen (d.h. Kreuzladungen oder durchweg geringe Ladungen aufwiesen) und entschieden sich aufgrund der Abwägung zwischen hoher Varianzaufklärung und guter inhaltlicher Interpretierbarkeit für eine finale Struktur.

Bühner (2021) empfahl über das Vorgehen von Harbin et al. (2018) hinausgehend, im Falle einer nicht interpretierbaren Lösung probeweise einen weiteren Faktor zu extrahieren und die rotierte Ladungsmatrix erneut zu untersuchen. Selbst im Falle theoretischer Vorannahmen ist es dabei nicht per se “verboten”, eine andere Anzahl an Faktoren zu extrahieren, weil dies den theoretisch-empirischen Horizont erweitern und unter Umständen auch die theoretische Vorannahme bestärken kann. Costello und Osborne (2005) empfahlen hingegen, neben der durch die Prüfverfahren nahegelegten Faktorenanzahl auch Lösungen mit einem Faktor mehr sowie einem Faktor weniger zu testen. Wenn etwa der *Scree*-Plot vier Faktoren nahelegt, sollten drei, vier und fünf Faktoren geprüft und die am besten interpretierbare Lösung beibehalten werden. Alle Empfehlungen sprechen sich daher zumindest implizit dafür aus, der Interpretierbarkeit der

Faktorlösung einen größeren Stellenwert beizumessen als einer unter Umständen unreflektierten Anwendung der eingesetzten Prüfverfahren und Verfahrensschritte.

## 3 Einführung in die Konfirmatorische Faktorenanalyse

### 3.1 Allgemeine Anmerkungen und Logik

Vereinfacht gesagt, handelt es sich bei der Konfirmatorischen Faktorenanalyse um ein Modell, in dem konkrete Annahmen darüber vorliegen, welche Kreuzladungen ausgelassen werden dürfen, ohne die inhaltlich-theoretische Grundlage zu verkennen (Schmitt, 2011). Damit unterscheidet sich die Konfirmatorische Faktorenanalyse von der Explorativen Faktorenanalyse dahingehend, dass bei der Konfirmatorischen Faktorenanalyse nicht erst wie bei der Exploratorischen Faktorenanalyse die Struktur in den Daten gesucht wird. Vielmehr bestehen Überlegungen zur spezifischen Zuordnung von Items zu Faktoren im Rahmen der Konfirmatorischen Faktorenanalyse bereits im Vorhinein. Angewandt auf unser *Big Five*-Beispiel sollten die jeweiligen Items etwa nur auf diejenigen Faktoren laden, die sie erfassen sollen, während Ladungen auf fremde Faktoren unterbunden werden. Konfirmatorische Faktorenanalysen streben üblicherweise also eine Einfachstruktur an.

Unter eher dogmatisch orientierten Forschenden wird postuliert, dass Exploratorische Faktorenanalysen nur dann durchgeführt werden sollen, wenn keinerlei Vorannahmen zur Struktur vorliegen und dass Konfirmatorische Faktorenanalysen nur dann genutzt werden sollen, wenn exakte Annahmen zur Dimensionalität (d.h. Anzahl der Faktoren) und zur Item-Faktor-Zuordnung vorliegen (d.h. Zulassen und Auslassen von Ladungen; vgl. Blötner & Mokros, 2023, für eine Abwägung). Empirisch finden sich jedoch häufig auch Ansätze, in denen eine anscheinend “gesicherte” Struktur reanalysiert wird, um den Status Quo zu hinterfragen (siehe etwa Roy et al., 2021).

Zu Beginn einer Konfirmatorischen Faktorenanalyse wird das zu prüfende Modell spezifiziert und festgelegt, welche Ladungen zugelassen werden und welche nicht. Mit diesem Modell geht eine sogenannte *modellimplizierte Varianz-Kovarianz-Matrix* ( $\Sigma_{Modell}$ ) einher, die das notwendige Muster der Varianzen und Kovarianzen der Items unter der Annahme abbildet, dass das aufgestellte Modell korrekt sei. Diese modellimplizierte Varianz-Kovarianz-Matrix wird verglichen mit der *empirischen Varianz-Kovarianz-Matrix* ( $\Sigma_{empirisch}$ ), also dem Muster von Varianzen und Kovarianzen, die sich tatsächlich im untersuchten Datensatz ergeben. Es wird die Nullhypothese geprüft, dass die modellimplizierte Varianz-Kovarianz-Matrix mit der empirischen Varianz-Kovarianz-Matrix übereinstimmt,  $H_0 : \Sigma_{Modell} = \Sigma_{empirisch}$ . Im Gegensatz zum üblichen Vorgehen in der Statistik, in der die Nullhypothese eines Nulleffekts (im weitesten Sinne) für gewöhnlich abgelehnt werden soll, sind wir bei der Durchführung einer Konfirmatorischen Faktorenanalyse daran interessiert, die Nullhypothese *nicht* verwerfen zu müssen. Es handelt sich also um

einen Anpassungstest. Für die Mathematik hinter dieser Prüfung sei auf einschlägige Lehrbücher (bspw. Bühner, 2021; Eid et al., 2017) und entsprechende Kurse zu multivariater Statistik im Masterstudium verwiesen.

### 3.2 Interpretation

Die Nullhypothese der Gleichheit der beobachteten und der modellimplizierten Varianz-Kovarianz-Matrizen wird anhand eines  $\chi^2$ -Tests geprüft. Es wird davon ausgegangen, dass die beiden Matrizen überzufällig stark voneinander abweichen (d.h. dass das aufgestellte Modell die Daten also nicht adäquat widerspiegelt), wenn der beobachtete  $\chi^2_{beobachtet,df}$ -Wert größer ist als der kritische  $\chi^2$ -Wert mit den entsprechenden Freiheitsgraden  $df$  auf einem gegebenen Signifikanzniveau  $\alpha$ ,  $\chi^2_{beobachtet,df} > \chi^2_{krit,df,\alpha}$  (zu Freiheitsgraden, siehe die angeführten Lehrbücher). Da der  $\chi^2$ -Test bei ausreichend großem Stichprobenumfang zunehmend streng wird und eine Ablehnung des Modells selbst bei vernachlässigbaren Abweichungen bedingen würde (Schmitt, 2011), werden ergänzend deskriptive Indizes und entsprechende Konventionen zur Bewertung der Modellpassung herangezogen (sog. *Anpassungskoeffizienten* bzw. *Fit-Indizes*). Die wohl bekannteste Konvention stammt von Hu und Bentler (1999), wonach ein Modell als hinreichend passend zu bewerten ist, wenn der *Comparative Fit Index* (CFI) den Wert .95 annimmt oder überschreitet ( $CFI \geq .95$ ), der *Root Mean Square Error of Approximation* (RMSEA) den Wert .05 annimmt oder unterschreitet ( $RMSEA \leq .05$ ) und das *Standardized Root Mean Residual* (SRMR) den Wert .08 annimmt oder unterschreitet ( $SRMR \leq .08$ ). Ein Modell kann und darf dadurch jedoch nicht als “wahr” oder “korrekt” angenommen werden, weil zu jedem spezifizierten Modell eine große Zahl alternativer Modelle existiert, welche die gleiche Varianz-Kovarianz-Matrix erzeugen (Bühner, 2021). Ferner sollten neben den genannten Fit-Statistiken auch die Ladungs- und Korrelationsmuster untersucht werden. So ist es etwa denkbar, dass selbst bei ausgezeichneter Anpassung (im Sinne der Fit-Indizes) nur unbefriedigende Faktorladungen vorliegen, ein Item entgegen der Annahme eine negative Ladung aufweist oder Faktoren deutlich stärker oder deutlich schwächer korreliert sind als angenommen (zu weiteren, detaillierteren Untersuchungsmöglichkeiten von gutem oder schlechtem Modellfit, siehe Bühner, 2021; Eid et al., 2017; Rosseel, 2012).

### 3.3 Prüfung

Nachfolgend schauen wir uns die theoretisch angenommene fünffaktorielle Struktur der bereits mehrfach betrachteten *Big Five*-Skala an. Zunächst prüfen wir ein Modell, in welchem die fünf Faktoren – übereinstimmend mit ursprünglichen theoretischen Annahmen – untereinander durchweg unkorreliert sind. Wie oben beschrieben, werden Kreuz- oder Nebenladungen explizit nicht zugelassen.

```

model_CFA_unkorreliert <- '
O =~ O1 + O2 + O3 + O4 + O5
C =~ C1 + C2 + C3 + C4 + C5
E =~ E1 + E2 + E3 + E4 + E5
A =~ A1 + A2 + A3 + A4 + A5
N =~ N1 + N2 + N3 + N4 + N5

O ~~ O * C + O * E + O * A + O * N
C ~~ O * E + O * A + O * N
E ~~ O * A + O * N
A ~~ O * N'

fit.model_CFA_unkorreliert <- cfa(model_CFA_unkorreliert, data_BFI)
fitMeasures(fit.model_CFA_unkorreliert,
             c("chisq", "df", "pvalue", "cfi", "rmsea", "srmr"))

```

```

##      chisq      df  pvalue      cfi   rmsea   srmr
## 5182.991 275.000   0.000   0.698   0.089   0.137

```

Der  $\chi^2$ -Index ist statistisch signifikant,  $\chi^2(275) = 5,182.99$ ,  $p < .001$ . Das deutet darauf hin, dass die beobachtete Varianz-Kovarianz-Matrix die modellimplizierte Varianz-Kovarianz-Matrix nicht adäquat abbildet. Aufgrund der angesprochenen Sensitivität des  $\chi^2$ -Indexes bei großen Stichproben sollten zusätzlich die deskriptiven Fit-Statistiken untersucht werden. Diese deuten ebenfalls durchweg auf inakzeptablen Modell-Fit hin. Um eine mögliche Quelle für schlechten Fit zu prüfen, geben wir nachfolgend die Annahme unkorrelierter Faktoren auf. Das hier verwendete R-Programmpaket (*lavaan*; Rosseel, 2012) schätzt automatisch sämtliche Faktor-Interkorrelationen, sodass diese nicht händisch angefordert werden müssen.

```

model_CFA_korreliert <- '
O =~ O1 + O2 + O3 + O4 + O5
C =~ C1 + C2 + C3 + C4 + C5
E =~ E1 + E2 + E3 + E4 + E5
A =~ A1 + A2 + A3 + A4 + A5
N =~ N1 + N2 + N3 + N4 + N5'

fit.model_CFA_korreliert <- cfa(model_CFA_korreliert, data_BFI)

```

Bevor wir das Modell an sich bewerten, prüfen wir zuerst, ob die Aufgabe der Annahme unkorrelierter

Faktoren die Modellpassung verbessert hat.

```
anova(fit.model_CFA_unkorreliert, fit.model_CFA_korreliert)

##
## Chi-Squared Difference Test
##
##           Df      AIC      BIC Chisq Chisq diff  RMSEA Df diff
## fit.model_CFA_korreliert   265 182699 183041 3843.3
## fit.model_CFA_unkorreliert 275 184018 184304 5183.0    1339.7 0.24386    10
##
##                   Pr(>Chisq)
## fit.model_CFA_korreliert
## fit.model_CFA_unkorreliert < 0.00000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Der Unterschied der  $\chi^2$ -Indizes (Chisq diff = 1,339.70 mit Df diff = 10 Freiheitsgraden) ist signifikant ( $\text{Pr}(>\text{Chisq}) < .001$ ). Wir gehen deshalb davon aus, dass das angepasste Modell mit korrelierten Faktoren die Daten besser widerspiegelt als das ursprüngliche Modell, im welchem alle fünf Faktoren als untereinander unkorreliert angenommen wurden. Nachfolgend betrachten wir die Fit-Statistiken des angepassten Modells.

```
##   chisq      df  pvalue      cfi   rmsea   srmr
## 3843.296 265.000   0.000   0.780   0.078   0.076
```

Übereinstimmend mit unseren Erwartungen legt der  $\chi^2$ -Index aufgrund der sehr großen Stichprobe wieder nahe, das Modell abzulehnen,  $\chi^2(265) = 3,843.30$ ,  $p < .001$ . Sowohl der CFI (.780 < .95) als auch der RMSEA (.078 > .05) nehmen Werte an, welche die konventionellen Trennwerte unter- beziehungsweise überschreiten und deshalb wieder auf einen inakzeptablen Modell-Fit hinweisen. Nur das SRMR ist gerade noch akzeptabel (.076 < .08). Diese Befunde sollen jedoch nicht ausdrücken, dass das Modell “falsch” sei. Die gängigen Konventionen für Fit-Indizes legen häufig unbefriedigende Modellpassung nahe, weil sie anhand von simulierten – also idealisierten – Daten gewonnen wurden. Angesichts überwiegend befriedigender Ladungen im vorliegenden Datensatz (s.u.), welche die Befunde für realistische Daten relativ gut widerspiegeln, sollten CFI, RMSEA und SRMR nicht überinterpretiert werden (Heene et al., 2011; Hopwood & Donnellan, 2010).

```
##   lhs op rhs est.std
## 1   0 =~ 01   0.562
## 2   0 =~ 02   0.431
```

```

## 3  O =~ O3  0.722
## 4  O =~ O4  0.221
## 5  O =~ O5  0.476
## 6  C =~ C1  0.536
## 7  C =~ C2  0.578
## 8  C =~ C3  0.550
## 9  C =~ C4  0.697
## 10 C =~ C5  0.622
## 11 E =~ E1  0.560
## 12 E =~ E2  0.694
## 13 E =~ E3  0.645
## 14 E =~ E4  0.704
## 15 E =~ E5  0.561
## 16 A =~ A1  0.330
## 17 A =~ A2  0.634
## 18 A =~ A3  0.741
## 19 A =~ A4  0.503
## 20 A =~ A5  0.678
## 21 N =~ N1  0.821
## 22 N =~ N2  0.796
## 23 N =~ N3  0.722
## 24 N =~ N4  0.571
## 25 N =~ N5  0.509
## 56 O ~~ C   0.284
## 57 O ~~ E   0.448
## 58 O ~~ A   0.268
## 59 O ~~ N  -0.114
## 60 C ~~ E   0.360
## 61 C ~~ A   0.326
## 62 C ~~ N  -0.279
## 63 E ~~ A   0.693
## 64 E ~~ N  -0.234
## 65 A ~~ N  -0.207

```

Aufgrund der Verwendung eines anderen Programmpakets (`lavaan` [Rosseel, 2012] statt `psych` [Revelle, 2023]) ist die Darstellung der Ladungsmuster anders angelegt als zuvor. Im angepassten Modell finden sich vereinzelt nur schwache Ladungen (der Ausdruck `=~` bezeichnet die Ladung eines Items auf die den jeweiligen

Faktor in Höhe der Zahl in der Spalte `est.std`; so weist das Item 04 eine standardisierte Ladung des Faktors *Offenheit* in Höhe von .221 auf). Die Faktoren sind mal mehr, mal weniger stark korreliert (bspw. Ausdrücke  $E \sim A; \rho = .693$  beziehungsweise  $0 \sim N; \rho = -.114$ ).

## 4 Ausblick: Strukturgleichungsmodelle als Konfirmatorische Faktorenanalyse plus Regression

Latente Modelle können nicht nur genutzt werden, um die Struktur von einzelnen Variablen zu explorieren oder zu stützen (Exploratorische bzw. Konfirmatorische Faktorenanalyse), sondern auch, um (*kausale Beziehungen zwischen latenten Konstrukten*) zu untersuchen. Modelle, in denen die latenten Konstrukte als Prädiktoren, Moderatoren, Mediatoren oder Kriterien genutzt werden, heißen *Strukturgleichungsmodelle*. Vereinfacht gesagt, kombinieren Strukturgleichungsmodelle die Konfirmatorische Faktorenanalyse (d.h. eine Spezifikation der Beziehungen zwischen Items und latenten Faktoren) mit der Regressionsanalyse (also mit einer Festlegung, welche Konstrukte anhand welcher anderen Konstrukte vorhergesagt werden sollen). Anders als in Konfirmatorischen Faktorenanalysen werden in Strukturgleichungsmodellen (analog zu Regressionsmodellen) konkrete Annahmen dazu aufgestellt, welche Variable(n) die angenommene(n) Ursache(n) darstellt/darstellen und welche Variable die angenommene(n) Konsequenz(en) darstellen soll(en).

Die Vorteile liegen dabei auf der Hand: Erstens kann durch Strukturgleichungsmodelle direkt bestimmt werden, welche Items geeignet sind, das interessierende Konstrukt abzubilden. In Regressionsanalysen, die auf Summen- oder Mittelwert-Scores basieren, bleibt ohne vorgeschaltete Itemanalyse unklar, ob und wie stark einzelne Items zur Ausprägung des Konstrukts beitragen. Zweitens sind Strukturgleichungsmodelle im Gegensatz zu den meisten manifesten Regressionsmodellen *messfehlerfrei*. Die üblicherweise in der einfachen oder multiplen Regression eingehenden Summen- oder Mittelwert-Scores weisen fast immer nicht-perfekte Reliabilitätsschätzungen auf (z.B. Cronbachs  $\alpha < 1$ ). Bei Konfirmatorischen Faktorenanalysen (und entsprechend auch bei Strukturgleichungsmodellen) hingegen werden die Messfehler explizit modelliert, indem aufgegliedert wird, welcher Anteil der Itemvarianz durch den Faktor aufgeklärt wird (quantifiziert durch die Faktorladung) und welcher Anteil durch den Faktor unerklärt bleibt (sog. *Item-Residuum*). Wird eine konkrete Beziehung zwischen zwei Konstrukten also einmal mittels einfacher Regression und einmal mittels Strukturgleichungsmodell modelliert, wird die Effektstärke im Strukturgleichungsmodell stets betragsmäßig größer sein als im korrespondierenden Regressionsmodell. Dies lässt sich durch die *doppelte Attenuationskorrektur* (d.h. Korrektur für mangelnde Messgenauigkeit im Prädiktor und im Kriterium) erklären, die im Strukturgleichungsmodell automatisch erfolgt.

Dies soll nachfolgend kurz veranschaulicht werden.

Konkret berechnen wir die einfache Regression eines manifesten Mittelwert-Scores für *Verträglichkeit* auf den manifesten Mittelwert-Score der *Extraversion*, wobei hierbei natürlich keine Kausalität angenommen wird.

```
data_BFI$agree <- rowMeans(data_BFI[paste0("A", 1:5)])
data_BFI$extra <- rowMeans(data_BFI[paste0("E", 1:5)])

lm.beta(lm(agree ~ extra, data = data_BFI))
```

```
##
## Call:
## lm(formula = agree ~ extra, data = data_BFI)
##
## Standardized Coefficients::
## (Intercept)      extra
##           NA    0.4747705
```

In dieser einfachen Regressionsanalyse mit dem Prädiktor *Extraversion* und dem Kriterium *Verträglichkeit* beträgt das standardisierte Regressionsgewicht  $\beta = .47$ . Nachfolgend spezifizieren wir das korrespondierende Strukturgleichungsmodell.

```
m <- 'A =~ A1 + A2 + A3 + A4 + A5
      E =~ E1 + E2 + E3 + E4 + E5

      A ~ E'
fit <- sem(m, data_BFI)
standardizedsolution(fit)[11, 1:4]
```

```
##   lhs op rhs est.std
## 11  A  ~  E    0.692
```

Hierbei bedeutet der Ausdruck  $A \sim E$ , dass *Verträglichkeit* durch *Extraversion* vorhergesagt werden soll. Der sogenannte *Pfadkoeffizient* beträgt  $\beta = .69$ .

Der dritte Vorteil von Strukturgleichungsmodellen besteht in ihrer verhältnismäßig größeren Flexibilität. Im Gegensatz zu Regressionsmodellen lässt sich nämlich eine – im Rahmen des Verhältnisses von Modellparametern zum Stichprobenumfang – beliebige Anzahl von Kriterien, Moderatoren und Mediatoren aufnehmen, und es ist möglich, nur einzelne Pfade für bestimmte Variablen zu kontrollieren. Regressionsmodelle würden dazu einen größeren Programmieraufwand und/oder mehrschrittige Prüfverfahren erfordern. Diese Aspekte sollen anhand eines weiteren Beispiels veranschaulicht werden.

```

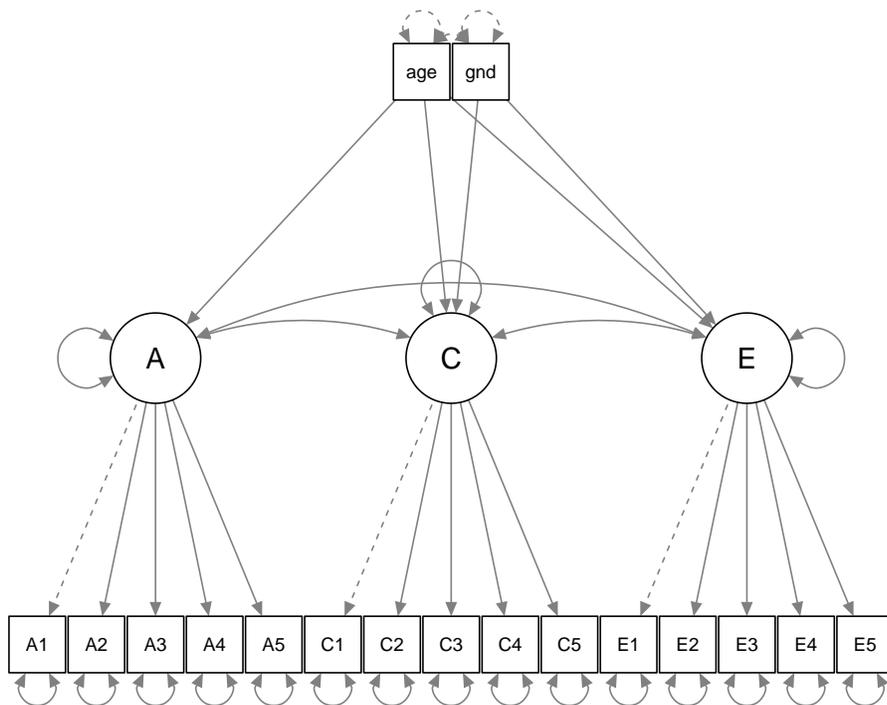
model_SEM <- '
A =~ A1 + A2 + A3 + A4 + A5
C =~ C1 + C2 + C3 + C4 + C5
E =~ E1 + E2 + E3 + E4 + E5
A ~ age
C ~ age + gender
E ~ age + gender'

fit.model_SEM <- sem(model_SEM, data_BFI)

```

In diesem exemplarischen Modell wurden Verträglichkeit, Gewissenhaftigkeit und Extraversion mithilfe von Alter vorhergesagt, Gewissenhaftigkeit und Extraversion darüber hinausgehend aber auch durch das Geschlecht (bspw.  $A \sim \text{age}$ ,  $C \sim \text{age} + \text{gender}$  und  $E \sim \text{age} + \text{gender}$ ). Hieraus lässt sich auch ein weiterer flexibilitätsbezogener Vorteil ableiten, nämlich die Kombinierbarkeit latenter (Verträglichkeit, Gewissenhaftigkeit, Extraversion) und manifester Variablen (Alter, Geschlecht), welche beliebige Skalenniveaus annehmen können (Alter: kontinuierlich; Geschlecht: binär [im vorliegenden, von Revelle adaptierten Datensatz]). Die nachfolgende Grafik veranschaulicht das aufgestellte Gesamtmodell (Abbildung 7).

**Abbildung 7**



Die Bewertung von Strukturgleichungsmodellen erfolgt anhand derselben Maßstäbe wie die Bewertung von

Konfirmatorischen Faktorenanalysen.

```
fitMeasures(fit.model_SEM, c("chisq", "df", "pvalue", "cfi", "rmsea", "srmr"))
```

```
##      chisq      df  pvalue      cfi  rmsea  srmr
## 1545.290 112.000   0.000   0.840  0.076  0.063
```

Wir würden daher auf Basis des CFI ( $.840 < .95$ ) und des RMSEA ( $.076 > .06$ ) von unbefriedigender Modellpassung ausgehen, auf Basis des SRMR hingegen von ausreichender Modellpassung ( $.063 < .08$ ). Kritik an der strikten Verwendung der oben genannten Grenzwerte für Konfirmatorische Faktorenanalysen gilt in äquivalenter Weise für Strukturgleichungsmodelle (Heene et al., 2011; Hopwood & Donnellan, 2010).

```
standardizedsolution(fit.model_SEM)[c(1:20, 39:41), c(1:4, 7)]
```

```
##      lhs op      rhs est.std pvalue
## 1      A =~      A1  0.332  0.000
## 2      A =~      A2  0.637  0.000
## 3      A =~      A3  0.739  0.000
## 4      A =~      A4  0.506  0.000
## 5      A =~      A5  0.677  0.000
## 6      C =~      C1  0.541  0.000
## 7      C =~      C2  0.590  0.000
## 8      C =~      C3  0.559  0.000
## 9      C =~      C4  0.685  0.000
## 10     C =~      C5  0.610  0.000
## 11     E =~      E1  0.565  0.000
## 12     E =~      E2  0.695  0.000
## 13     E =~      E3  0.621  0.000
## 14     E =~      E4  0.727  0.000
## 15     E =~      E5  0.549  0.000
## 16     A ~      age  0.132  0.000
## 17     C ~      age  0.112  0.000
## 18     C ~ gender  0.036  0.127
## 19     E ~      age  0.055  0.022
## 20     E ~ gender  0.015  0.450
## 39     A ~~      C   0.310  0.000
## 40     A ~~      E   0.691  0.000
## 41     C ~~      E   0.347  0.000
```

Das Modell legt unter anderem nahe, dass Verträglichkeit mit dem Alter ansteigt (Ausdruck  $A \sim \text{age}$ ;  $\beta = .132$ ). Verträglichkeit ist positiv mit Gewissenhaftigkeit (Ausdruck  $A \sim\sim C$ ;  $\beta = .310$ ) und Extraversion (Ausdruck  $A \sim\sim E$ ;  $\beta = .691$ ) korreliert. Gewissenhaftigkeit und Extraversion sind ebenfalls positiv korreliert (Ausdruck  $C \sim\sim E$ ;  $\beta = .347$ , alle genannten  $ps < .001$ ). Frauen und Männer unterscheiden sich im Mittel nicht hinsichtlich Gewissenhaftigkeit ( $C \sim\sim \text{gender}$ ;  $\beta = .036$ ) oder Extraversion ( $E \sim\sim \text{gender}$ ;  $\beta = .015$ , beide  $ps \geq .13$ ).

#### 4.1 Fazit zu Konfirmatorischen Faktorenanalysen und Strukturgleichungsmodellen

Konfirmatorische Faktorenanalysen und Strukturgleichungsmodelle lassen sich nicht mittels eines einzigen isolierten Kennwerts bewerten. Vielmehr sollten Fit-Indizes, Faktorladungen und Faktorinterkorrelationen gegeneinander abgewogen werden. Lai und Green (2016) wiesen beispielsweise darauf hin, dass insbesondere der CFI und der RMSEA in den nahegelegten Schlussfolgerungen zur Modellpassung divergieren können. Generell sollten die konventionellen Fit-Indizes also nicht als bindend oder als unumstößliche Regeln angesehen werden (Schmitt, 2011). Gerade im Fall eines angeblich schlechten Fits bei Konfirmatorischen Faktorenanalysen und Strukturgleichungsmodellen neigen viele Forschende dazu, in explorativer Weise Restriktionen auf Basis eines zu erwartenden verbesserten Fits zu lockern (Schmitt, 2011). Dadurch geben sie den konfirmatorischen Gedanken auf und laufen Gefahr, aufgrund von stichprobenabhängigen Charakteristika ein angeblich gutes, aber nicht notwendigerweise replizierbares Modell zu etablieren. Obwohl Exploratorische und Konfirmatorische Faktorenanalysen nicht im selben Datensatz verwendet werden sollten, um eine exploratorisch ermittelte Struktur zu “beweisen” (Fokkema & Greiff, 2017), kann es dennoch ratsam sein, Gründe für schlechten Fit einer Konfirmatorischen Faktorenanalyse anhand einer Exploratorischen Faktorenanalysen zu suchen (Schmitt, 2011). Entsprechende Unterschiede sollten dabei jedoch transparent berichtet werden (für eine Anwendung exploratorischer Verfahren bei unzureichendem Modellfit der *Short Dark Tetrad*, siehe etwa Blötner et al., 2023). Im Umkehrschluss kann es vorkommen, dass Konfirmatorische Faktorenanalysen und Strukturgleichungsmodelle durch die gängigen Modellgüteindizes gestützt werden. Das bedeutet aber nicht, dass das aufgestellte Modell das “wahre” Modell darstellt (s.o.). Ebenso existiert in der EFA nie die *eine* richtige Lösung, weil jede Faktorenlösung mitsamt des angewandten Rotationsverfahrens willkürlich und austauschbar ist. Das Ziel bei der Anwendung von Faktorenanalysen in der Testentwicklung, -evaluation und -anwendung besteht deshalb darin, das eigene Vorgehen gut zu rechtfertigen und eine praktikable und interpretierbare Lösung zu eruieren.

Bezogen auf einen anderen Inhalt dieses Moduls sei auf die Stichprobenabhängigkeit der faktorenanalytischen Befunde an sich verwiesen. So ließen sich die Unterschiede in den Intelligenzmodellen nach Thurstone und Spearman insbesondere dadurch erklären, dass in den von Thurstone verwendeten, sehr homogenen Stichproben die Varianzen der Items eingeschränkt waren. Eine sogenannte *Varianzeinschränkung* führt jedoch auch dazu, dass die zu beobachtende Korrelation schrumpft. Infolgedessen werden Item-

Interkorrelationen bei “varianzeingeschränkten” Stichproben im Vergleich zu heterogeneren Stichproben verhältnismäßig kleiner, was im Falle des Intelligenzmodells nach Thurstone zur Extraktion mehrerer weitestgehend unabhängiger Faktoren führte. Bei den Untersuchungen von Spearman hingegen waren die Stichproben jeweils heterogener, sodass mehr Variabilität in den Itemantworten vorlag (d.h. mehr Varianz führte zu einer potentiell größeren Kovarianz und somit zu einer größeren Korrelation). Dadurch konnten stärkere Korrelationen zwischen den Items gefunden werden, was die Extraktion eines homogenen  $g$ -Faktor begünstigte (siehe Hagemann et al., 2022, für einen Überblick).

## 5 Literatur

- Blötner, C., & Mokros, A. (2023). The next distinction without a difference: Do psychopathy and sadism scales assess the same construct? *Personality and Individual Differences, 205*, Artikel 112102. <https://doi.org/10.1016/j.paid.2023.112102>
- Blötner, C., Webster, G. D., & Wongsomboon, V. (2023). Measurement invariance of the Short Dark Tetrad across cultures and genders. *European Journal of Psychological Assessment, 39*(5), 331–336. <https://doi.org/10.1027/1015-5759/a000715>
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-12770-0>
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4. Aufl.). Pearson.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, Artikel 7. <https://doi.org/10.7275/jyj1-4868>
- Crawford, C. B., & Koopman, P. (1979). Note: Inter-rater reliability of scree test and mean square ratio test of number of factors. *Perceptual and Motor Skills, 49*(1), 223–226. <https://doi.org/10.2466/pms.1979.49.1.223>
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33*(4), 533–552. [https://doi.org/10.1016/s0191-8869\(01\)00171-4](https://doi.org/10.1016/s0191-8869(01)00171-4)
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*(6), 1246–1256. <https://doi.org/10.1037/0022-3514.73.6.1246>
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Beltz.
- Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble: Overfitting in the assessment of internal structure and some editorial thoughts on it. *European Journal of Psychological Assessment, 33*(6), 399–402. <https://doi.org/10.1027/1015-5759/a000460>
- Hagemann, D., & Spinath, F. M., & Mueller, E. M. (2022). *Differentielle Psychologie und Persönlichkeitsforschung* (9. Aufl.). Kohlhammer.
- Harbin, S. M., Kelley, M. L., Piscitello, J., & Walker, S. J. (2018). Multidimensional Bullying Victimization Scale: Development and validation. *Journal of School Violence, 18*(1), 146–161. <https://doi.org/10.1080/15388220.2017.1423491>
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note in the usefulness of cutoff values of fit indices.

*Psychological Method*, 16(3), 319–336. <https://doi.org/10.1037/a0024917>

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14(3), 332–346. <https://doi.org/10.1177/1088868310361240>

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, 51(2-3), 220–239. <https://doi.org/10.1080/00273171.2015.1134306>

Luhmann, M. (2020). *R für Einsteiger. Einführung in die Statistiksoftware für die Sozialwissenschaften* (5. Aufl.). Beltz.

Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion* (3. Aufl.). Springer. <https://doi.org/10.1007/978-3-662-61532-4>

Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung*. Hogrefe Testzentrale.

Revelle, W. (2023). *psych: Proceures for psychological, psychometric, and personality research (R package version 2.3.3)*. CRAN. <https://cran.r-project.org/package=psych>

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>

Roy, S., Vize, C., Uzieblo, K., van Dongen, J. D. M., Miller, J., Lynam, D., Brazil, I., Yoon, D., Mokros, A., Gray, N. S., Snowden, R., & Neumann, C. S. (2021). Triarchic or septarchic?-Uncovering the Triarchic Psychopathy Measure's (TriPM) structure. *Personality Disorders*, 12(1), 1–15. <https://doi.org/10.1037/per0000392>

Salgado, J. F. (2017). Bandwidth-Fidelity Dilemma. In V. Zeigler-Hill, & T. Shackelford (Hrsg.), *Encyclopedia of Personality and Individual Differences* (S. 1–4). Springer. [https://doi.org/10.1007/978-3-319-28099-8\\_1280-1](https://doi.org/10.1007/978-3-319-28099-8_1280-1)

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4), 304–321. <https://doi.org/10.1177/0734282911406653>

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432–442.