CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Multivariate Verfahren

Einheit 6: Gruppieren: Exploratorische Faktorenanalyse (EFA)

21.11.2024 | Prof. Dr. Stephan Goerigk

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Was ist eine Faktorenanalyse?

Die Faktorenanalyse ist ein statistisches Verfahren, das darauf abzielt:

- Zusammenhänge zwischen Variablen zu identifizieren und zu erklären.
- Eine große Anzahl von Items/Variablen auf eine kleinere Anzahl latenter Dimensionen (Faktoren) zu reduzieren.
- Latente Konstrukte wie Intelligenz, Persönlichkeit oder Stress messbar zu machen.

Zentrale Ziele der Faktorenanalyse:

1. Datenreduktion:
   - Komplexe Datensätze vereinfachen.
   - Finden von zugrundeliegenden (latenten) Strukturen.
2. Erforschung von Zusammenhängen:
   - Welche Variablen (latenten) hängen miteinander zusammen?
3. Hypothesenbildung:
   - Explorative Analyse unbekannter Strukturen in Daten.

Exploratorische Faktorenanalyse (EFA)

Exploratorische Faktorenanalyse (EFA) vs. Konfirmatorische Faktorenanalyse (CFA)

Zwei Varianten der Faktorenanalyse:

Kriterium	Exploratorische Faktorenanalyse (EFA)	Konfirmatorische Faktorenanalyse (CFA)
Ziel	Strukturen und Zusammenhänge explorieren	Vorab definierte Strukturen bestätigen
Hypothesen	Keine festen Hypothesen	Basierend auf theoretischen Annahmen
Vorgehen	Datengetrieben	Hypothesengetrieben
Faktorenstruktur	Unbekannt, wird aus den Daten extrahiert	Vorab spezifiziert
Anwendungsbeispiel	Erforschung neuer Skalen oder Konstrukte	Validierung bestehender Modelle/Skalen

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Exploratorische Faktorenanalyse (EFA) vs. Konfirmatorische Faktorenanalyse (CFA)

Wann EFA? Wann CFA?

- EFA:
  - Erste Schritte in der Skalenentwicklung
  - Untersuchung neuer latenter Strukturen
- CFA:
  - Testen von Modellen basierend auf Theorie
  - Validierung und Replikation von Befunden

Beispiel aus der Psychologie:

- EFA:
  - Entwicklung eines Fragebogens zu „Arbeitszufriedenheit".
  - Prüfen, wie in wie viele Dimensionen das Konstrukt Persönlichkeit unterteilt werden kann
- CFA:
  - Prüfung, ob „Arbeitszufriedenheit" durch die Dimensionen „Arbeitsbedingungen" und „Motivation" beschrieben werden kann.
  - Vergleich eines 3-Faktoren (z.B. Eysenck) vs. 5-Faktoren Modells (z.B. Big Five) der Persönlichkeit

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE UNIVERSITY OF PSYCHOLOGY

Beobachtbare (manifeste) Indikatoren und Latente Variablen

1. Beobachtbare (manifeste) Indikatoren
    - Messbare Variablen aus einem Datensatz (z. B. Antworten auf Fragebögen).
    - Werden direkt erhoben (z. B. durch Fragen zu Stimmung, Verhalten oder Meinungen).
    - Beispiel: „Wie zufrieden sind Sie mit Ihrer Arbeit?“ (Skala von 1 bis 5).
2. Latente Variablen (Konstrukte)
    - Nicht direkt messbare, theoretische Konzepte, die hinter den beobachtbaren Indikatoren stehen.
    - Latente Variablen können wir nicht direkt messen, sondern nur über die Verhältnisse von verschiedenen Variablen zueinander schätzen.
    - Beispiel:
        - Latente Variable: „Arbeitszufriedenheit“.
        - Beobachtbare Indikatoren: Antworten zu Gehalt, Arbeitsklima, Work-Life-Balance.

Exploratorische Faktorenanalyse (EFA)

Mathematische Grundidee der Exploratorischen Faktorenanalyse (EFA)

- Ziel:
  - Reduzierung einer großen Anzahl von Variablen (z. B. Items eines Fragebogens) auf wenige zugrundeliegende Dimensionen (Faktoren).
  - Vereinfachung der Datenstruktur, ohne wesentliche Informationen zu verlieren.
- Logik:
  - Items, die hoch miteinander korrelieren, könnten zum selben dahinter zugrundeliegenden Konstrukt gehören (man sagt auch "auf dieses laden")
  - z.B. Fragen zu „Geselligkeit“ und „Redseligkeit“ könnten beide stark mit dem Faktor „Extraversion“ zusammenhängen.

$\rightarrow$ mathematisch funktioniert die EFA also über Korrelationen zwischen den Indikatoren

Exploratorische Faktorenanalyse (EFA)

Das Faktorenmodell der EFA: X = LF + E

$$X = \mathbf{L}F + E$$

- X: Matrix der beobachteten Variablen (Datenmatrix).
- $\mathbf{L}$: Ladungsmatrix (gibt an, wie stark die Variablen auf die Faktoren laden; Wertebereich: -1 bis +1.).
- F: Faktorenmatrix (die latenten Variablen/Dimensionen).
- E: Fehlerterm (Anteil der Varianz, der nicht durch die Faktoren erklärt wird).

Exploratorische Faktorenanalyse (EFA)

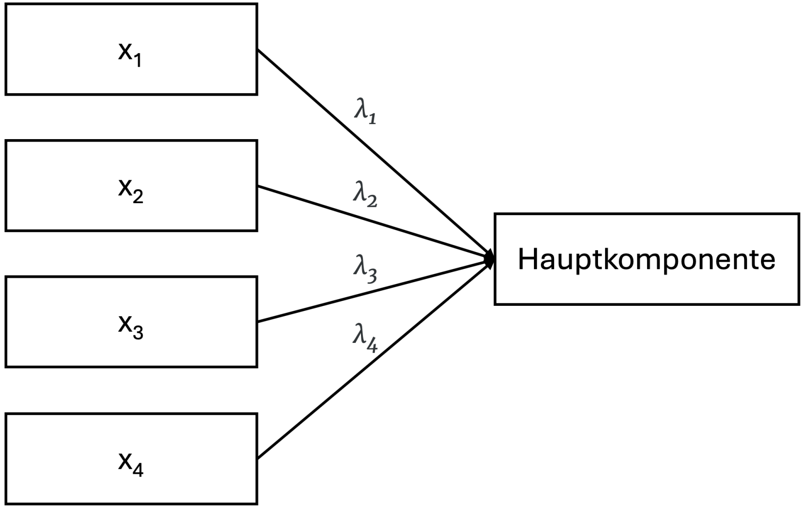
Hauptkomponentenanalyse (PCA) vs. Faktorenanalyse (EFA)

- Beides sind Methoden zur Dimensionalitätreduktion - was sind die Unterschiede?

Kriterium	**Hauptkomponentenanalyse (PCA)**	**Faktorenanalyse (FA)**
Ziel	Reduktion der Dimensionalität, Beibehaltung maximaler Varianz	Identifikation latenter Variablen, Erklärung von Zusammenhängen
Mathematische Grundlage	Eigenwertzerlegung der Kovarianz- oder Korrelationsmatrix	Faktorenmodell: $X = \mathbf{L}F + E$
Berücksichtigung von Fehlern	Kein Fehlerterm berücksichtigt	Fehlerterm wird explizit modelliert
Varianz	Erklärt die gesamte Varianz der Variablen	Erklärt nur die gemeinsame Varianz der Variablen
Rotation	Optional, da Hauptkomponenten unkorreliert	Häufig erforderlich, um Faktoren zu interpretieren
Anwendung	Datenreduktion, Visualisierung, Transformation	Theoriebildung, Identifikation latenter Konstrukte

Exploratorische Faktorenanalyse (EFA)

Hauptkomponentenanalyse (PCA)



$$x_1 \times \lambda_1 + x_2 \times \lambda_2 + x_3 \times \lambda_3 + x_4 \times \lambda_4 = Hauptkomponente$$

$\rightarrow$ Linearkombination

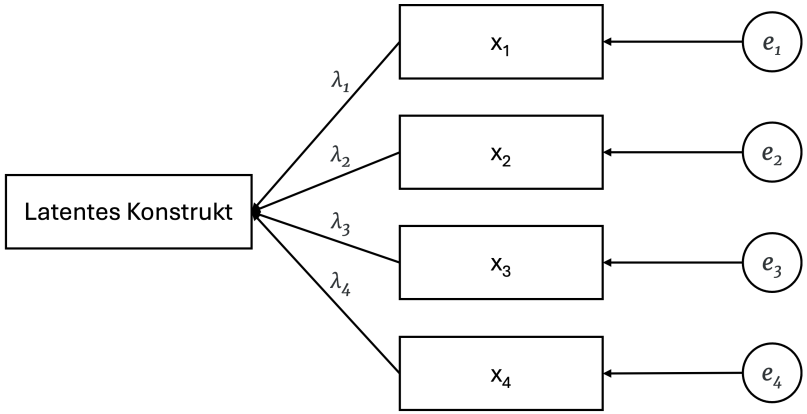
- Daten auf eine Weise reduzieren, bei der die geringste Menge an Informationen verloren geht
- Gleichzeitig unsere Modellgüte nicht senken

Vorgehen:

1. Hauptkomponentenanalyse erstellt k (Variablenzahl) Hauptkomponenten
2. Diese sind mit anderen Hauptkomponenten vollkommen unkorrelliert
3. Ermitteln, wie stark jede Variable auf jeder Hauptkomponente lädt.
4. Ziel: Optimale Anzahl an Hauptkomponenten und die optimalen Ladungen jeder Variable auf jede Hauptkomponente

Exploratorische Faktorenanalyse (EFA)

Faktorenanalyse (EFA)



- Ziel: Zugrunde liegende Faktorstruktur aus einer Reihe von Variablen zu identifizieren

Aus diesem Modell resultieren damit vier Gleichungen:

- $x_1 = \xi \times \lambda_1 + e_1$
- $x_2 = \xi \times \lambda_2 + e_2$
- $x_3 = \xi \times \lambda_3 + e_3$
- $x_4 = \xi \times \lambda_4 + e_4$

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Hauptkomponentenanalyse (PCA) vs. Faktorenanalyse (EFA)

Schlussfolgerungen:

- PCA reduziert die Variablen auf ihre Hauptkomponenten und versucht Informationsgehalt beizubehalten
- EFA wird genutzt, um latente (unbeobachtbare) Variablen zu messen.
- Wenn die Variablen keine Gemeinsamkeiten aufweisen, kann die EFA keine geeigneten Faktoren identifizieren, während die PCA trotzdem eine Hauptkomponente findet, die die maximale Varianz in den Daten erklärt.
- PCA um latente Variablen zu messen: Kann irreführend sein $\rightarrow$ behalten gesamte Variabilität bei, wollen aber nur Variabilität, die über alle Variablen hinweg gemeinsam ist.

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Voraussetzungen

1. Stichprobengröße
   - Faustregel: Mindestens 5–10 Beobachtungen pro Variable.
   - Große Stichproben (z. B. > 300) führen zu stabileren Ergebnissen.
2. Korrelationen zwischen Variablen
   - Notwendigkeit signifikanter Korrelationen:
   - Kaiser-Meyer-Olkin (KMO) Kriterium: Wert > 0.6 (geeignet für EFA).
   - Bartlett-Test auf Sphärizität: Signifikant ($p < 0.05$).
3. Linearität
   - Die Beziehungen zwischen den Variablen sollten linear sein.
4. Normalverteilung
   - Besonders wichtig bei der Hauptachsenanalyse (PAF).
5. Keine oder geringe Multikollinearität
   - Variablen dürfen nicht perfekt miteinander korrelieren (Korrelationen < 0.9).
6. Messniveau der Variablen
   - Variablen sollten mindestens intervallskaliert sein.

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Beispiel Big 5 Skala

Item	Beschreibung	Item	Beschreibung
V1	Bin gleichgültig gegenüber den Gefühlen anderer.	V2	Erkundige mich nach dem Wohlergehen anderer.
V3	Weiß, wie man andere trösten kann.	V4	Liebe Kinder.
V5	Bringe Menschen dazu, sich wohlzufühlen.	G1	Bin genau in meiner Arbeit.
G2	Mache weiter, bis alles perfekt ist.	G3	Mache Dinge nach einem Plan.
G4	Mache Dinge halbherzig.	G5	Verschwende meine Zeit.
E1	Rede nicht viel.	E2	Finde es schwierig, auf andere zuzugehen.
E3	Weiß, wie man Menschen fesselt.	E4	Schließe leicht Freundschaften.
E5	Übernehme das Kommando.	N1	Werde leicht wütend.
N2	Werde leicht gereizt.	N3	Habe häufige Stimmungsschwankungen.
N4	Fühle mich oft niedergeschlagen.	N5	Gerate leicht in Panik.
O1	Bin voller Ideen.	O2	Vermeide schwierige Lesematerialien.
O3	Hebe die Konversation auf ein höheres Niveau.	O4	Verbringe Zeit damit, über Dinge nachzudenken.
O5	Gehe nicht tiefgründig in ein Thema ein.		

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

```
head(big5, 10) # erste 10 Zeilen des Datensatzes
```

```
##     V1 V2 V3 V4 V5 G1 G2 G3 G4 G5 E1 E2 E3 E4 E5 N1 N2 N3 N4 N5 O1 O2 O3 O4 O5
## 1    6  6  5  6  5  6  6  6  1  3  2  1  6  5  6  3  5  2  2  3  4  3  5  6  1
## 2    4  3  1  5  1  3  2  4  2  4  3  6  4  2  1  6  3  2  6  4  3  2  4  5  3
## 3    4  4  5  6  5  4  3  5  3  2  1  3  2  5  4  3  3  4  2  3  5  3  5  6  3
## 4    4  5  2  2  1  5  5  5  2  2  3  4  3  6  5  2  4  2  2  3  5  2  5  5  5
## 5    1  5  6  5  6  4  3  2  4  5  2  1  2  5  2  2  2  2  2  2  6  1  5  5  2
## 6    2  6  5  6  5  3  5  6  3  6  2  2  4  6  6  4  4  4  6  6  6  1  5  6  1
## 7    4  5  5  6  5  5  5  4  1  1  3  2  5  5  6  2  3  3  1  1  6  2  5  6  2
## 8    1  6  6  1  6  5  2  5  1  1  1  1  6  6  6  2  3  1  2  1  6  4  5  5  3
## 9    2  4  4  4  3  6  5  6  1  1  2  4  4  2  6  3  3  5  3  2  5  2  6  6  1
## 10   2  5  1  3  5  5  4  5  2  5  1  2  6  5  4  1  4  2  2  5  2  4  5  4  1
```

- 5 Items pro Persönlichkeitsdimension
- Wertebereich der Items von 1 bis 6

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

```r
cor(big5)
```

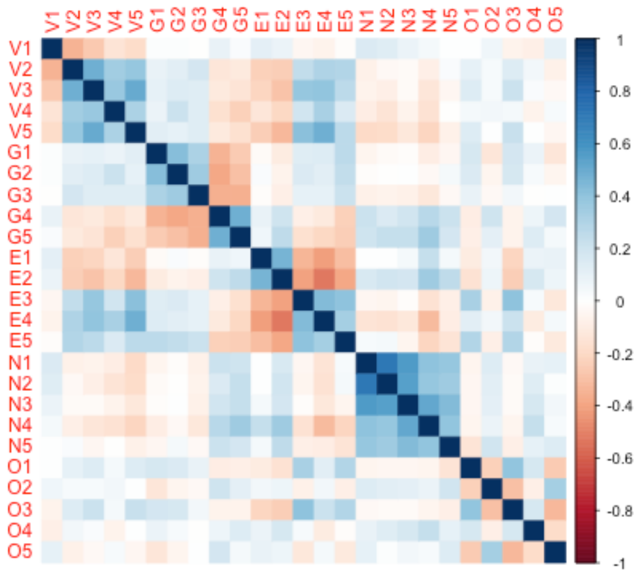
```
##        V1    V2    V3    V4    V5    G1    G2    G3    G4    G5    E1    E2    E3    E4    E5    N1    N2    N3    N4    N5    O1    O2    O3
## V1  1.00 -0.34 -0.26 -0.14 -0.19  0.02  0.01 -0.01  0.10  0.02  0.12  0.08 -0.04 -0.07 -0.02  0.16  0.13  0.09  0.04  0.01  0.00  0.07 -0.06
## V2 -0.34  1.00  0.48  0.34  0.38  0.09  0.13  0.19 -0.14 -0.11 -0.24 -0.24  0.25  0.30  0.30 -0.08 -0.04 -0.02 -0.09  0.02  0.11  0.03  0.15
## V3 -0.26  0.48  1.00  0.38  0.50  0.10  0.14  0.13 -0.12 -0.15 -0.22 -0.29  0.38  0.39  0.26 -0.07 -0.08 -0.03 -0.13 -0.04  0.14  0.03  0.22
## V4 -0.14  0.34  0.38  1.00  0.32  0.08  0.22  0.13 -0.16 -0.24 -0.14 -0.20  0.20  0.33  0.16 -0.09 -0.15 -0.07 -0.16  0.00  0.04  0.05  0.04
## V5 -0.19  0.38  0.50  0.32  1.00  0.12  0.11  0.13 -0.12 -0.16 -0.25 -0.33  0.41  0.48  0.27 -0.19 -0.19 -0.13 -0.21 -0.08  0.15  0.00  0.22
## G1  0.02  0.09  0.10  0.08  0.12  1.00  0.43  0.32 -0.35 -0.25 -0.03 -0.10  0.13  0.14  0.26 -0.06 -0.03 -0.01 -0.09 -0.05  0.18 -0.13  0.19
## G2  0.01  0.13  0.14  0.22  0.11  0.43  1.00  0.36 -0.38 -0.30  0.02 -0.07  0.15  0.12  0.25 -0.02  0.00  0.01 -0.04  0.05  0.16 -0.05  0.18
## G3 -0.01  0.19  0.13  0.13  0.13  0.32  0.36  1.00 -0.35 -0.35 -0.02 -0.09  0.10  0.10  0.22 -0.08 -0.06 -0.07 -0.13 -0.04  0.09 -0.03  0.06
## G4  0.10 -0.14 -0.12 -0.16 -0.12 -0.35 -0.38 -0.35  1.00  0.48  0.10  0.21 -0.09 -0.12 -0.23  0.21  0.15  0.20  0.28  0.21 -0.10  0.21 -0.07
## G5  0.02 -0.11 -0.15 -0.24 -0.16 -0.25 -0.30 -0.35  0.48  1.00  0.07  0.26 -0.17 -0.21 -0.24  0.21  0.24  0.23  0.35  0.18 -0.09  0.12 -0.07
## E1  0.12 -0.24 -0.22 -0.14 -0.25 -0.03  0.02 -0.02  0.10  0.07  1.00  0.47 -0.33 -0.42 -0.31  0.01  0.01  0.05  0.23  0.04 -0.10  0.06 -0.21
## E2  0.08 -0.24 -0.29 -0.20 -0.33 -0.10 -0.07 -0.09  0.21  0.26  0.47  1.00 -0.40 -0.52 -0.39  0.17  0.20  0.19  0.35  0.26 -0.16  0.08 -0.24
## E3 -0.04  0.25  0.38  0.20  0.41  0.13  0.15  0.10 -0.09 -0.17 -0.33 -0.40  1.00  0.43  0.40 -0.04 -0.06 -0.01 -0.15 -0.09  0.33 -0.07  0.41
## E4 -0.07  0.30  0.39  0.33  0.48  0.14  0.12  0.10 -0.12 -0.21 -0.42 -0.52  0.43  1.00  0.33 -0.14 -0.15 -0.13 -0.31 -0.09  0.12  0.05  0.21
## E5 -0.02  0.30  0.26  0.16  0.27  0.26  0.25  0.22 -0.23 -0.24 -0.31 -0.39  0.40  0.33  1.00  0.04  0.05 -0.06 -0.21 -0.14  0.29 -0.09  0.30
## N1  0.16 -0.08 -0.07 -0.09 -0.19 -0.06 -0.02 -0.08  0.21  0.21  0.01  0.17 -0.04 -0.14  0.04  1.00  0.71  0.57  0.41  0.38 -0.05  0.14 -0.03
## N2  0.13 -0.04 -0.08 -0.15 -0.19 -0.03  0.00 -0.06  0.15  0.24  0.01  0.20 -0.06 -0.15  0.05  0.71  1.00  0.55  0.39  0.35 -0.05  0.12 -0.02
## N3  0.09 -0.02 -0.03 -0.07 -0.13 -0.01  0.01 -0.07  0.20  0.23  0.05  0.19 -0.01 -0.13 -0.06  0.57  0.55  1.00  0.52  0.43 -0.05  0.11 -0.03
## N4  0.04 -0.09 -0.13 -0.16 -0.21 -0.09 -0.04 -0.13  0.28  0.35  0.23  0.35 -0.15 -0.31 -0.21  0.41  0.39  0.52  1.00  0.40 -0.06  0.08 -0.06
## N5  0.01  0.02 -0.04  0.00 -0.08 -0.05  0.05 -0.04  0.21  0.18  0.04  0.26 -0.09 -0.09 -0.14  0.38  0.35  0.43  0.40  1.00 -0.15  0.20 -0.08
## O1  0.00  0.11  0.14  0.04  0.15  0.18  0.16  0.09 -0.10 -0.09 -0.10 -0.16  0.33  0.12  0.29 -0.05 -0.05 -0.05 -0.06 -0.15  1.00 -0.23  0.39
## O2  0.07  0.03  0.03  0.05  0.00 -0.13 -0.05 -0.03  0.21  0.12  0.06  0.08 -0.07  0.05 -0.09  0.14  0.12  0.11  0.08  0.20 -0.23  1.00 -0.29
## O3 -0.06  0.15  0.22  0.04  0.22  0.19  0.18  0.06 -0.07 -0.07 -0.21 -0.24  0.41  0.21  0.30 -0.03 -0.02 -0.03 -0.06 -0.08  0.39 -0.29  1.00
## O4 -0.09  0.05  0.02 -0.06  0.00  0.08  0.03  0.00  0.07  0.14  0.08  0.17  0.04 -0.10 -0.02  0.09  0.13  0.17  0.23  0.11  0.17 -0.08  0.17
## O5  0.11 -0.08 -0.04  0.04 -0.04 -0.13 -0.06  0.00  0.18  0.05  0.09  0.08 -0.13  0.04 -0.11  0.10  0.02  0.05  0.03  0.14 -0.25  0.33 -0.32
##        O4    O5
## V1 -0.09  0.11
## V2  0.05 -0.08
## V3  0.02 -0.04
## V4 -0.06  0.04
## V5  0.00 -0.04
## G1  0.08 -0.13
## G2  0.03 -0.06
## G3  0.00  0.00
## G4  0.07  0.18
```

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Beispiel Big 5 Skala

```
corrplot(cor(big5), method = 'color')
```



- Darstellung der Korrelationen mittels Heatmap
- Korrelationskoeffizient wird farblich visualisiert
- Einige Items innerhalb der Subskalen sind negativ korreliert

→ Diese müssen zunächst umgepolt werden!

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

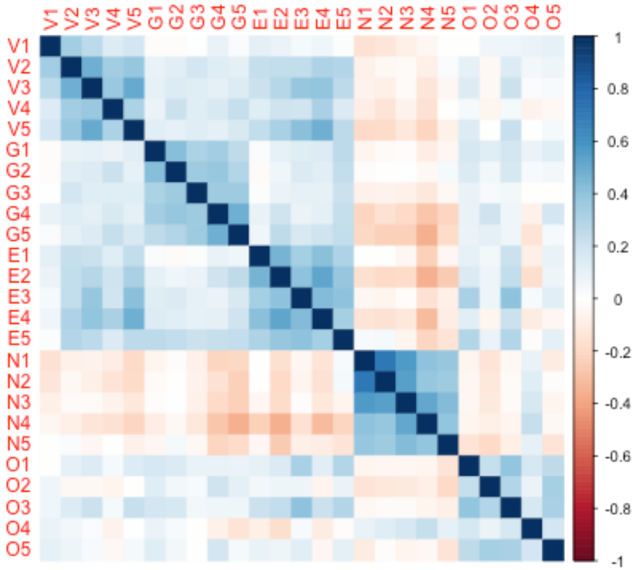
Umkodieren am Beispiel des 1. Verträglichkeitsitems:

```
library(dplyr)

big5$V1 = recode(big5$V1,
                 "1" = 6,
                 "2" = 5,
                 "3" = 4,
                 "4" = 3,
                 "5" = 2,
                 "6" = 1)
```

Jetzt stimmen die Korrelationen →

```
corrplot(cor(big5), method = 'color')
```



Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Faktorenanalyse (EFA) - Voraussetzungen

Kaiser-Meyer-Olkin Kriterium (KMO)

- KMO gibt an, ob ein Datensatz für eine Faktorenanalyse geeignet ist.
- KMO-Wert wird basierend auf den Korrelationen zwischen den Indikatoren berechnet.
- Wertebereich von 0 bis +1. Werte näher an 1 $\rightarrow$ Variablen korrelieren und Daten sind für eine Faktorenanalyse geeignet

KMO-Wert	Interpretation
Über 0.90	Hervorragend (Marvelous)
0.80 bis 0.90	Gut (Meritorious)
0.70 bis 0.80	Durchschnittlich (Average)
0.60 bis 0.70	Mittelmäßig (Mediocre)
0.50 bis 0.60	Schlecht (Terrible)
Unter 0.50	Unakzeptabel (Unacceptable)

Kaiser, Henry F. 1974. “An Index of Factorial Simplicity.” Psychometrika 39 (1): 31–36.

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

Berechnung KMO für unser Beispiel:

```
KMO(big5)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = big5)
## Overall MSA =  0.85
## MSA for each item =
##   V1   V2   V3   V4   V5   G1   G2   G3   G4   G5   E1   E2   E3   E4   E5   N1   N2   N3   N4   N5   O1   O2   O3   O4
## 0.74 0.83 0.87 0.87 0.90 0.84 0.79 0.85 0.82 0.86 0.84 0.88 0.89 0.88 0.89 0.78 0.78 0.86 0.89 0.86 0.86 0.78 0.83 0.78 0
```

- Zieht man Cutoffs nach Kaiser (1974) heran ist $r = .85$ als gut zu bewerten

Exploratorische Faktorenanalyse (EFA)

Voraussetzungen

Bartlett-Test auf Sphärizität

- Test, der die Hypothese überprüft, dass die Korrelationsmatrix eine Einheitsmatrix ist.
- Mit anderen Worten: Er prüft, ob die Korrelationen zwischen den Variablen gleich null sind.
- Ein statistischer Test mit einem p-Wert kleiner als .05 zeigt, dass die Korrelationsmatrix keine Einheitsmatrix ist und die Durchführung einer Faktorenanalyse gerechtfertigt ist.

Bartlett, Maurice S. 1951. "The Effect of Standardization on a χ^2 Approximation in Factor Analysis." Biometrika 38 (3/4): 337–44.

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

Berechnung Bartlett-Test für unser Beispiel:

```
cortest.bartlett(cor(big5))
```

```
## $chisq
## [1] 665.3162
##
## $p.value
## [1] 9.581928e-30
##
## $df
## [1] 300
```

- $p < .05 \rightarrow$ Test ist signifikant $\rightarrow$ Korrelationen bedeutsam > 0 $\rightarrow$ Faktoranalyse sinnvoll durchführbar

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Bestimmung der Anzahl der Faktoren

Eigenwerte und das Kaiser-Kriterium (Eigenwert > 1)

Eigenwerte

- Ein Eigenwert gibt an, wie viel Varianz ein Faktor in den Daten erklärt.
- Basieren auf der Korrelationsmatrix

Kaiser-Kriterium

- Nur Faktoren mit einem Eigenwert > 1 werden extrahiert.
- Diese Faktoren erklären mehr Varianz als eine einzelne Variable.
- Faktoren mit einem Eigenwert ≤ 1 werden als nicht signifikant betrachtet.
- Kritik: Das Kaiser-Kriterium neigt dazu, zu viele Faktoren zu extrahieren.

Exploratorische Faktorenanalyse (EFA)

Bestimmung der Anzahl der Faktoren

Scree-Plot

- Definition:
  - Ein Scree-Plot ist eine grafische Darstellung der Eigenwerte in absteigender Reihenfolge.
  - Ziel: Bestimmung der optimalen Anzahl an Faktoren, die extrahiert werden sollen.
- Darstellung:
  - X-Achse: Faktoren in der Reihenfolge ihrer Extraktion.
  - Y-Achse: Eigenwerte der Faktoren.
- Interpretation:
  - Suche den Knickpunkt ("Elbow"), an dem die Eigenwerte stark abfallen.
  - Die Faktoren vor dem Knickpunkt werden extrahiert, da sie die meiste Varianz erklären.
  - Faktoren nach dem Knickpunkt haben geringe Eigenwerte und erklären wenig zusätzliche Varianz.
- Vorteile:
  - Einfach und intuitiv zu interpretieren.
  - Ergänzt andere Methoden, wie das Kaiser-Kriterium.
- Kritik:
  - Der Knickpunkt ist nicht immer eindeutig erkennbar.
  - Kann subjektiv interpretiert werden, insbesondere bei flachen Kurven.

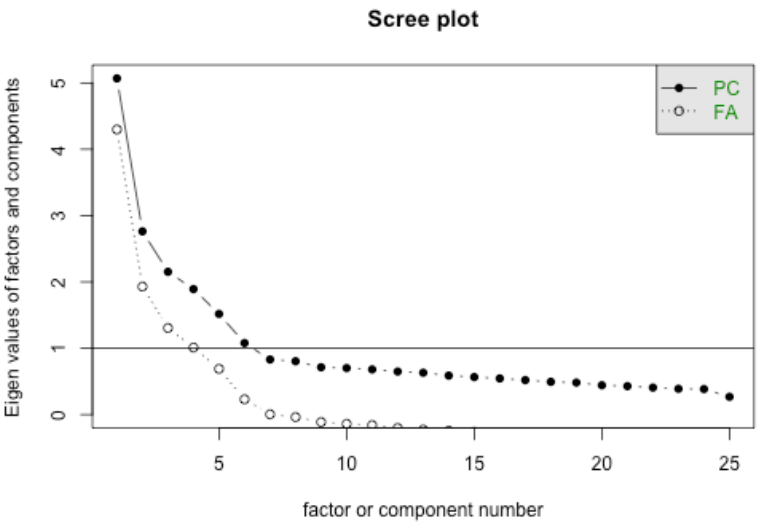
Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

```
eigen =  eigen(cor(big5))$values
eigen
```

```
psych::scree(cor(big5))
```

```
##  [1] 5.0685162 2.7624793 2.1526230 1.8923330 1.5175329 1.078 .6808421 0.6489
## [14] 0.5880320 0.5659652 0.5448396 0.5199335 0.4938686 0.482 .3847626 0.2681
```



Exploratorische Faktorenanalyse (EFA)

Bestimmung der Anzahl der Faktoren

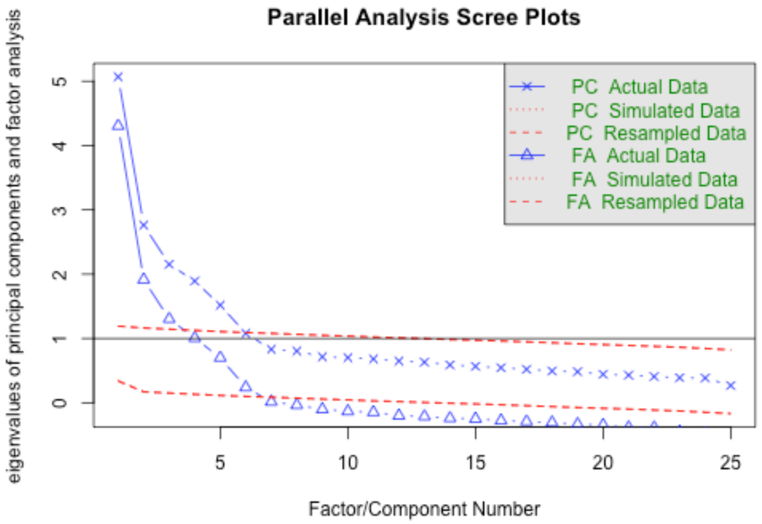
Parallelanalyse

- Vergleicht die Eigenwerte aus den tatsächlichen Daten mit den Eigenwerten, die aus zufälligen (simulierten) Daten stammen.
- Vorgehen:
  1. Generierung von Zufallsdaten mit derselben Stichprobengröße und Anzahl an Variablen.
  2. Berechnung der Eigenwerte aus den zufälligen Daten.
  3. Vergleich der Eigenwerte: Nur Faktoren, deren Eigenwerte größer sind als die entsprechenden Eigenwerte der Zufallsdaten, werden beibehalten.
- Darstellung:
  - Die Eigenwerte der tatsächlichen Daten und der Zufallsdaten werden grafisch gegenübergestellt.
  - Der Schnittpunkt der beiden Linien markiert die Anzahl der zu extrahierenden Faktoren.
- Vorteile:
  - Objektivere Methode im Vergleich zum Scree-Plot.
  - Berücksichtigt die Möglichkeit, dass hohe Eigenwerte zufällig auftreten können.

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

```r
fa.parallel(big5, fm = "ml")
```




```
## Parallel analysis suggests that the number of factors =  6  and the number of components =  5
```

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE UNIVERSITY OF PSYCHOLOGY

Rotation

- Nun haben wir zumindest eine Vorstellung davon haben, wie viele Faktoren extrahiert werden sollen.
- Nächste Frage: Welche Rotation?

Rotation:

- Faktorenrotation können Sie sich ähnlich vorstellen wie das Scharfstellen eines Teleskops.
- Ziel: klarere Zuordnung zwischen den einzelnen Faktoren und den verschiedenen Variablen zu erreichen.
- Wahl der Rotationsmethode hängt davon ab, ob Sie davon ausgehen, dass die Faktoren miteinander korreliert sein sollten oder nicht.
  - Oblique = korriliert
  - Orthogonal = nicht korriliert

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Rotation

Oblique vs. Orthogonal

- Unter idealen Umständen würden wir erwarten, dass die verschiedenen Faktoren unabhängig sind und daher nicht miteinander korrelieren.
- Das ist jedoch nicht immer realistisch. Daher wird oft empfohlen, zunächst davon auszugehen, dass die Faktoren nicht unabhängig sind und oblique Rotationsmethode zu verwenden (Costello & Osborne, 2005).

Oblique Rotation:

- Eine oblique Rotation liefert eine Korrelationsmatrix, die die Beziehungen zwischen den Faktoren beschreibt.
- Mit dieser Korrelationsmatrix können Sie bewerten, ob die Korrelationen beachtenswert sind (z. B. $r > |0.30|$).
- Falls die Korrelationen jedoch relativ gering sind (z. B. r < |0,30|), könnten Sie ggf. zur Annahme übergehen, dass die Faktoren unabhängig (orthogonal) sind (keine Korrelation aufweisen sollten).

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Rotation

Rotationstyp	**Option**	**Beschreibung**
Oblique	Promax	Die Promax-Rotation ist bekannt für ihre Fähigkeit, große Datensätze effizient zu verarbeiten. Zudem führt sie häufig zu höheren Korrelationen zwischen den Faktoren.
Oblique	Oblimin	Die direkte Oblimin-Rotation ist bei großen Datensätzen etwas weniger effizient, kann jedoch eine einfachere Faktorstruktur erzeugen.
Orthogonal	Varimax	Die Varimax-Rotation ist darauf ausgelegt, Kreuzladungen zu reduzieren und kleinere Ladungswerte zu minimieren, wodurch die Faktormodelle klarer werden.
Orthogonal	Quartimax	Die Quartimax-Rotation zielt darauf ab, die Anzahl der Variablen zu verringern, die benötigt werden, um einen Faktor zu erklären, was die Interpretation erleichtert.
Orthogonal	Equamax	Die Equamax-Option bietet einen Kompromiss zwischen Varimax und Quartimax.

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

```r
fa(r = cor(big5),
   nfactors = 5,
   fm = "ml",
   rotate = "promax")$loadings
```

```
##
## Loadings:
##    ML2    ML1    ML3    ML5    ML4
## V1 -0.145 -0.110         0.406
## V2         0.122         0.556
## V3         0.212         0.615
## V4                0.192  0.433 -0.192
## V5         0.278         0.539
## G1                0.550         0.169
## G2  0.137 -0.129  0.664
## G3                0.595
## G4 -0.141         0.665
## G5 -0.176         0.567
## E1         0.644 -0.161
## E2 -0.114  0.709
## E3         0.506         0.221  0.263
## E4         0.608         0.288 -0.104
## E5  0.164  0.497  0.226         0.168
## N1  0.879  0.231        -0.260
## N2  0.840  0.174        -0.250
## N3  0.723
## N4  0.490 -0.339
## N5  0.499 -0.157         0.147 -0.165
## O1         0.157                0.511
## O2 -0.177               -0.153  0.481
## O3         0.269                0.603
## O4  0.147 -0.250         0.162  0.365
## O5                              0.532
##
##                  ML2   ML1   ML3   ML5   ML4
## SS loadings    2.714 2.372 2.019 1.695 1.513
## Proportion Var 0.109 0.095 0.081 0.068 0.061
## Cumulative Var 0.109 0.203 0.284 0.352 0.413
```

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Kommunalitäten (Communalities)

- Kommunalitäten beschreiben den Anteil der Varianz einer Variable, der durch die extrahierten Faktoren erklärt wird.
- Wertebereich: Reicht von 0 (keine erklärte Varianz) bis 1 (komplette erklärte Varianz).
- Interpretation:
  - Hohe Kommunalität (z. B. > 0.6): Variable wird gut durch die Faktoren repräsentiert.
  - Niedrige Kommunalität (z. B. < 0.3): Variable passt möglicherweise nicht gut ins Modell.
- Unterschiede zwischen “Initial” und “Extraction”:
  - Initial Communality: Vor der Faktorenextraktion, basierend auf der Gesamtvarianz.
  - Extraction Communality: Nach der Extraktion, zeigt die durch die Faktoren erklärte Varianz.
- Ziel: Sicherstellen, dass die meisten Variablen durch die extrahierten Faktoren gut erklärt werden.
- Praktischer Hinweis: Bei niedrigen Kommunalitäten prüfen, ob Variablen aus der Analyse entfernt oder zusätzliche Faktoren extrahiert werden sollten.

Exploratorische Faktorenanalyse (EFA)

Beispiel Big 5 Skala

```
fa(r = cor(big5),
   nfactors = 5,
   fm = "ml",
   rotate = "varimax")$communalities
```

```
##        V1        V2        V3        V4        V5        G1        G2        G3        G4        G5        E1        E2
## 0.1569689 0.3975994 0.5153194 0.3058056 0.4754910 0.3307140 0.4206538 0.3250907 0.4841843 0.4386480 0.3603605 0.5458318 0
##        E5        N1        N2        N3        N4        N5        O1        O2        O3        O4        O5
## 0.4154737 0.7226180 0.6586281 0.5260834 0.4979814 0.3434101 0.3235170 0.2751398 0.4838962 0.2423501 0.2856415
```

Exploratorische Faktorenanalyse (EFA)

CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Beispiel Big 5 Skala

```
fa(r = cor(big5),
   nfactors = 5,
   fm = "ml",
   rotate = "varimax")$Vaccounted
```

```
##                              ML2        ML1        ML3        ML5        ML4
## SS loadings           2.6712948 2.46686713 1.98666795 1.85504025 1.54735040
## Proportion Var        0.1068518 0.09867469 0.07946672 0.07420161 0.06189402
## Cumulative Var        0.1068518 0.20552648 0.28499320 0.35919481 0.42108882
## Proportion Explained  0.2537512 0.23433224 0.18871723 0.17621368 0.14698565
## Cumulative Proportion 0.2537512 0.48808343 0.67680067 0.85301435 1.00000000
```

Take-aways

CHARLOTTE FRESENIUS HOCHSCHULE UNIVERSITY OF PSYCHOLOGY

- Die EFA dient dazu, zugrunde liegende Strukturen (Faktoren) in einem Datensatz zu entdecken und Variablenmuster zu identifizieren.
- Zur sparsamen Modellierung der Varianz wird oft die PCA, für die Erfassung einer latenten Variable die EFA verwendet.
- Die Anzahl der Faktoren/Komponenten basiert auf analystischen oder graphischen Methoden.
- Die Interpretation und Benennung der Faktoren ergibt sich aus der inhatlichen Betrachtung der Variablen, die auf den Faktore laden (inhaltlicher "gemeinsamer Nenner").
- Ausreichende Stichprobengröße und Korrelationen zwischen den Variablen (Bartlett-Test, KMO-Kriterium) sind notwendig.