

Explorative Faktorenanalyse

Jalynskij et al.

Einstieg

<https://openpsychometrics.org/tests/IPIP-BFFM/1.php>

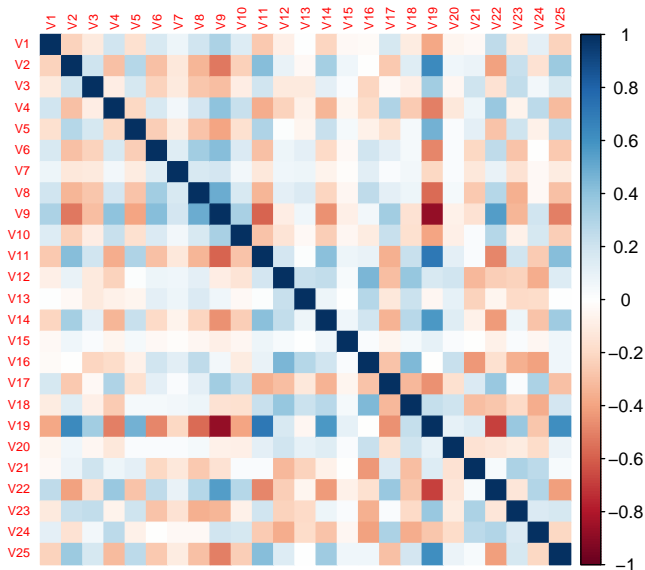
Die "Large Data Set Challenge"

Example

Stellen Sie sich vor, die von Ihnen soeben beantworteten Fragen ergäben die Korrelationsmatrix R auf der nächsten Folie. Die "Large Data Set Challenge" lautet: Erkennen Sie eine Struktur in den Daten? D.h., wenn ja weiter; welche Items könnten man Ihrer Meinung nach zu Itemgruppen zusammenfassen?

Anmerkung: Nein, das sind (wirklich) nicht ihre Antworten; $V1 - V30$ sind Zufallsvariablen!

Struktur erkennen & Itemgruppen finden: Übungsaufgabe 1



Die "Large Data Set Challenge"

Large Data Set Challenge

Mit zunehmender Itemzahl nimmt die Anzahl der Korrelationen, die für eine Analyse zu berücksichtigen sind schnell zu. Die "Challenge" ist eine mögliche Struktur zu erkennen!

In a Nutshell:

- Problem: Anzahl der Korrelationen
- z.B.: 25 Items $\hat{=}$ $2^{25} = 625$ Korrelationen
- Krux: Struktur erkennen
- \Leftrightarrow finde: hoch korrelierende Itemgruppen

Explorative Faktorenanalyse: in a Nutshell

- (ein) Hilfsmittel: .. (explorative) **Faktorenanalyse**

Faktorenanalyse

"The basic idea is to find latent variables (factors) based on the correlation structure of the manifest input variables (indicators)." (Mair 2018, S. 23)

- andere Helferlein zur *Datenreduktion* (eine Auswahl):
 - ▶ Hauptkomponentenanalyse
 - ▶ Clusteranalyse
 - ▶ Explorative Likertskalierung
 - ▶ (Non-) Metric Data Scaling (Voraus.: Distanzmatrizen)

Wichtig: "meaningful compression" vs. "full compression"

Strategie & Vorgehen: Simulieren & Evaluieren

- ① Man erschaffe ≥ 1 eine latente Variable (LV)
- ② ...lasse die LV Antwortmuster produzieren
- ③ ...wandel sie in eine Korrelationsmatrix um
- ④ ... und versucht die Struktur mit der Faktorenanalyse aufzufinden

Vom generativen Prozess zur Korrelationsmatrix

Der generative Prozess, d.h. wie genau ein Konstrukt die Antworten auf den Items erzeugt, bleibt meist verborgen. Wir untersuchen meistens lediglich Verhaltensspuren des Konstruktes, die sich in den Items ausdrückt, d.h. in der Struktur der Korrelationsmatrix niederschlägt. Strukturen zu simulieren ist hilfreich, weil wir dort "die Wahrheit" kennen und das Verfahren damit besser beurteilen können (\sim fake data analysis)

“Playing Creator”: Zwei latente Variable erschaffen

```
# Anzahl der Items
N <- 8

# Faktorladungen
load_F1 <- c(0.6, -0.3, 0.5, 0.7, 0.1, 0.2, 0.2, 0.3)
load_F2 <- c(-0.1, 0.1, 0.1, 0.1, -0.7, 0.5, -0.6, 0.7)
fx <- cbind(load_F1, load_F2)

# Zwischenfaktorkorrelation
phi <- diag(rep(1, 2)) ; phi[1, 2] <- phi[2, 1] <- 0.6

# Korrelationsmatrix
R <- psych::sim.structure(fx, phi)$model
```


Vom generativen Prozess zur Korrelationsmatrix

Prognose & Selbstexperiment: Übungsaufgabe 2

Example

Versuchen Sie es selbst! Verändern sie systematisch $F1$; $F2$ und ϕ_i . 1. Wie verändert sich die Korrelationsmatrix, in Abhängigkeiten Ihrer Veränderungen? 2. Können Sie eine eindeutige Struktur konstruieren? 3. Wenn ja, mit welchen Werten von $F1$; $F2$ und ϕ_i haben Sie ihr Ziel erreicht?

Example

Haben Sie ein Strukturmodell gefunden, dass ihnen gefällt? Ja, dann überlegen Sie sich für welche Konstrukte diese Struktur Sinn macht (z.B.: extraversion \sim openness to experience)

Logik latenter Variablen (..reversed)

Von der Korrelationsmatrix zur LV

Die Faktoranalyse ist ein strukturentdeckendes Verfahren. D.h. den generativen Prozess, d.h. wie ein Konstrukt die Antworten auf den Items verursacht hat anhand der Struktur die sich in der vorgegebenen Korrelationsmatrix zu *modellieren*.

- Modell: *Common Factor Model* (CFM)

Eingangsgleichung

$$x = \Lambda\xi + \epsilon \quad (1)$$

“In other words EFA tries to find p latent variables on the basis of the correlation structure of the m manifest variables.” (ebd.)

Das Common Factor Model (CFM)

Anmerkung: Für die Reformulierung von Gleichung (1) zu (2): siehe McCallum (2009)

Fudnamentaltheorem

$$P = \Lambda\Phi\Lambda^t + \Psi \quad (2)$$

- P : Modell-implizierte Korrelationsmatrix
- Λ : Ladungsmatrix
- Φ : Matrix der Zwischenfaktorkorrelationen
- Ψ : Uniqueness

Zusammenhang: Modell & Struktur

Die von ihnen konstruierte Struktur versuchen wir nun mit der Faktorenanalyse unter Einsatz des CFM zu rekonstruieren. Das CFM ist also Ihr Tool im bevorstehenden Rekonstruktionsprozess!

Faktorenanalyse: “A hurdle race”

“Unfortunately, factor analysis is frequently misunderstood and often misused. Some researchers appear to use factor analysis as a kind of divining rod, hoping to find gold hidden underneath tons of dirt. But there is nothing magical about the technique. [...] Factor analysis will yield meaningful results only when the research was meaningful to begin with.” Gregory (2014, p. 165)

- ① Extraktionsproblem
- ② Rotationsproblem
- ③ Problem der Anzahl zu extrahierender Faktoren

Extraktionsproblem

Rotationsproblem

Orthogonale Rotation

Oblique Rotation

Problem der Anzahl zu extrahierender Faktoren

Findet die Faktorenanalyse ihre Struktur?

Vanishing effekt – Simulation? Korrelationsmatrix (grafisch: Vanishing effect) (1) Faktorenanalyse mit zunehmender Zwischenfaktorkorrelation (2) Faktorenanalyse mit abnehmenden Ladungen

Faktorscores

Man könnte versuchen mit den Faktorscores die True Scores der Probanden vorherzusagen

Faktorenanalyse als Hürdenlauf

