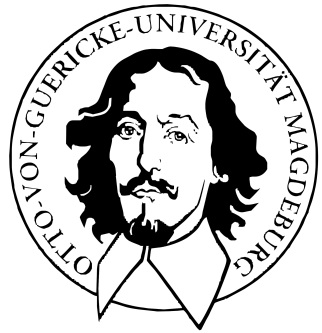


Grundlagen der Testtheorie

BSc Psychologie WiSe 2025/26

Prof. Dr. Dirk Ostwald

(3) Faktorenanalyse

Grundlagen

Modellformulierung

Modellschätzung

Modellevaluation

Modellinterpretation

Selbstkontrollfragen

Grundlagen

Modellformulierung

Modellschätzung

Modellevaluation

Modellinterpretation

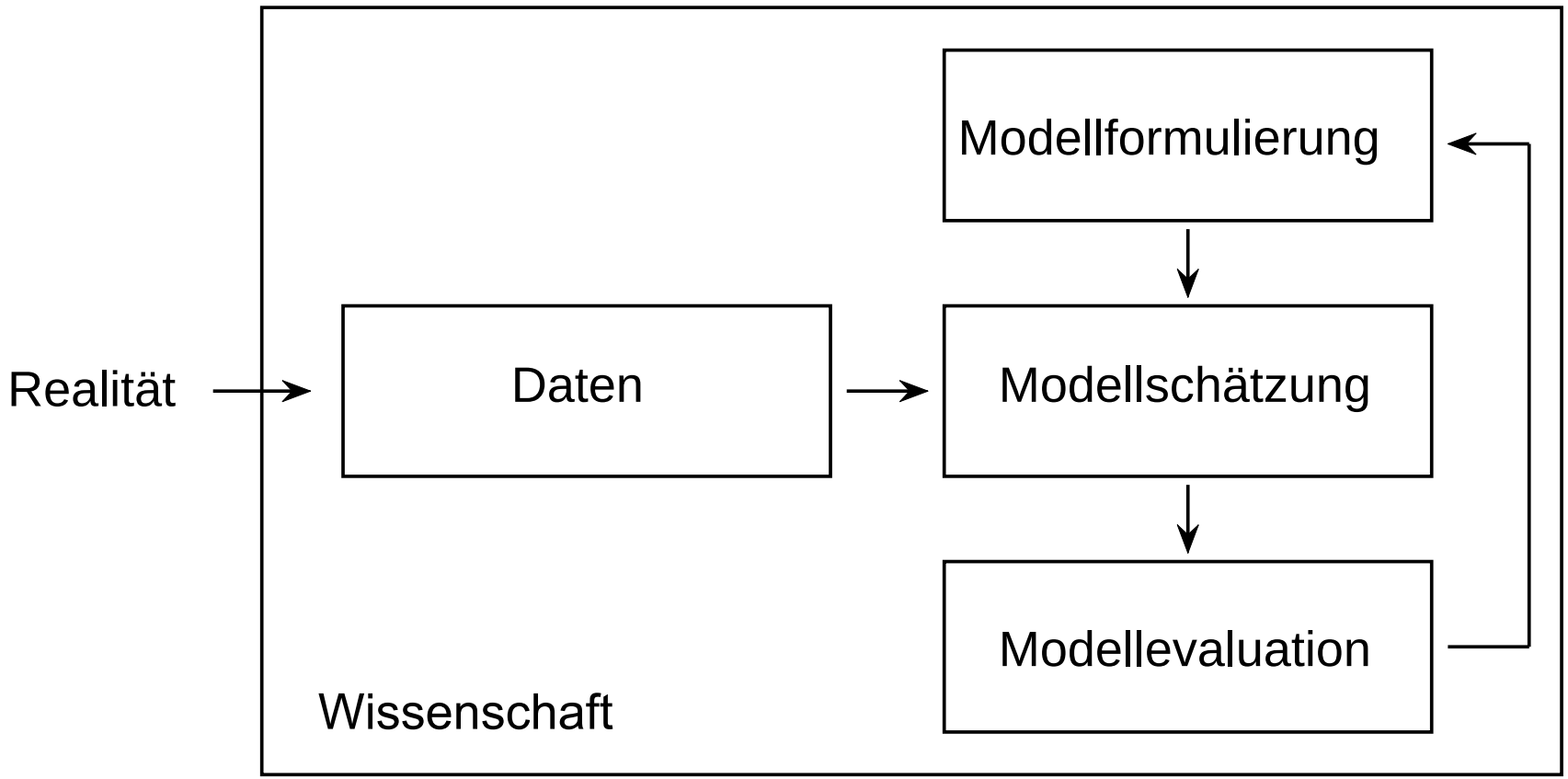
Selbstkontrollfragen

Motivation

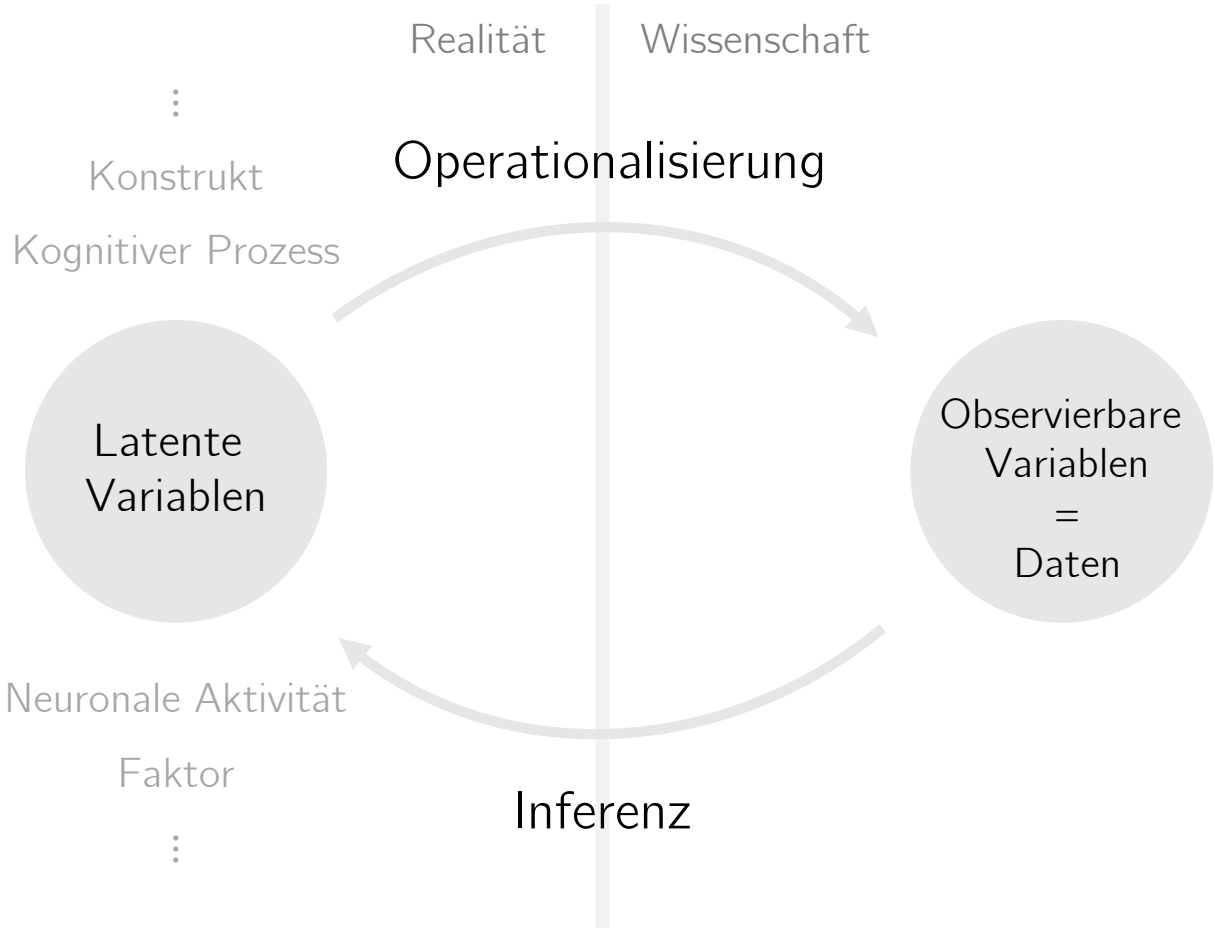
Faktorielle Validität des BDI-II nach Hautzinger et al. (2006)

"In der Untersuchung von Beck et al. (1996) legten *Scree-Tests* aus *explorativen Faktorenanalysen* sowohl für die Patientenstichprobe als auch für die Studentenstichprobe die *Extraktion* zweier Faktoren nahe. Nach parallel durchgeführten *Hauptachsenanalysen* mit anschließender *obliquer (Promax-)Rotation* resultierte für die Patientenstichprobe ein Faktor, den die Autoren als"somatisch-affektiven Faktor" bezeichneten, auf einem zweiten Faktor *luden* v.a. kognitive Items des BDI II, er wurde demnach als "kognitiver Faktor" bezeichnet. Die beiden obliquen Faktoren korrelierten miteinander zu r = 0.66. In der Studentenstichprobe repräsentierte dagegen der erste Faktor die "kognitiv-affektive", der zweite Faktor die "somatische" Dimension. Hier betrug die Korrelation der beiden Faktoren r = 0.62. Beck et al. (1996) führen dazu an, das das BDI-II zwei hoch korrelierende Faktoren repräsentiert und es daher möglich ist, dass einzelnen affektive Items (wie "Traurigkeit" oder "Weinen") in Abhängigkeit der untersuchten Stichprobe einmal auf dem einen, ein anderes Mal auf dem anderen Faktor substanzieller laden."

Psychologische Datenwissenschaft



Psychologische Datenwissenschaft



Grundlagen

Latente Variablenmodelle mit psychologischer Historie

- Faktorenanalyse und Strukturgleichungsmodelle
- Erklärung von Kovarianzen (vieler) beobachteter Variablen durch (wenige) latente Variablen

Klassische und aktuelle Anwendungsszenarien

- Analyse menschlicher Fähigkeiten (g-Faktor) und Persönlichkeitspsychologie (Big Five)
- Vielzahl von Phänomenen in der Soziologie, Politikwissenschaft, Biologie, Medizin, Sprachwissenschaft, ...

Varianten im psychologischen Sprachgebrauch

Explorative Faktorenanalyse (EFA)

- Datengetriebenes, eher prinzipienfreies und intuitiv geleitetes Inspirationsverfahren
- Fokus auf der numerische Behandlung von Stichprobenkovarianzmatrizen

Konfirmative Faktorenanalyse (CFA)

- Moderneres Verfahren mit expliziter probabilistischer Modellspezifikation
- Fokus auf probabilistischer Modellschätzung und Modellevaluation

Strukturgleichungsmodelle (SEM)

- Generalisierte konfirmative Faktorenanalyse mit Faktoreninteraktion
- Linearer Spezialfall genereller probabilistischer Modelle

Historie

Modellfreie explorative datenzentrische Periode (1900 - 1950)

- Pearson (1901) beschreibt erste Anklänge der Faktorenanalyse
- Spearman (1904) beginnt die Einfaktorenanalyse im Rahmen der Intelligenzforschung
- Hotelling (1933) entwickelt die eng verwandte Hauptkomponentenanalyse
- L. L. Thurstone (1947) beginnt die Mehrfaktorenanalyse im Bereich der Psychometrie

Modellbasierte konfirmative inferenzzentrische Periode (1950 - heute)

- Lawley (1940) schlägt die ML-Schätzung basierend auf Fisher (1922) und Wishart (1928) vor.
- Lawley & Maxwell (1962) machen den modellbasierten Charakter der Faktorenanalyse explizit.
- Jöreskog (1970) initiiert die Generalisierung zu Strukturgleichungsmodelle (vgl. Bollen (1989))
- Weitere Generalisierungen (vgl. z.B. Mulaik (2010) oder Bartholomew et al. (2011))

Software Periode (1970 - heute)

- lisrel (kommerziell, proprietär) nach Jöreskog (1970)
- mPlus (kommerziell, proprietär) nach Muthén & Muthén (1998--2017)
- lavaan (gratis, quelloffen) nach Rosseel (2012)
- $\Rightarrow$ Faktorenanalyse jeweils als Spezialfall von Strukturgleichungsmodellen

Anwendungsbeispiel

Simulierter Datensatz basierend auf Keller et al. (2008)

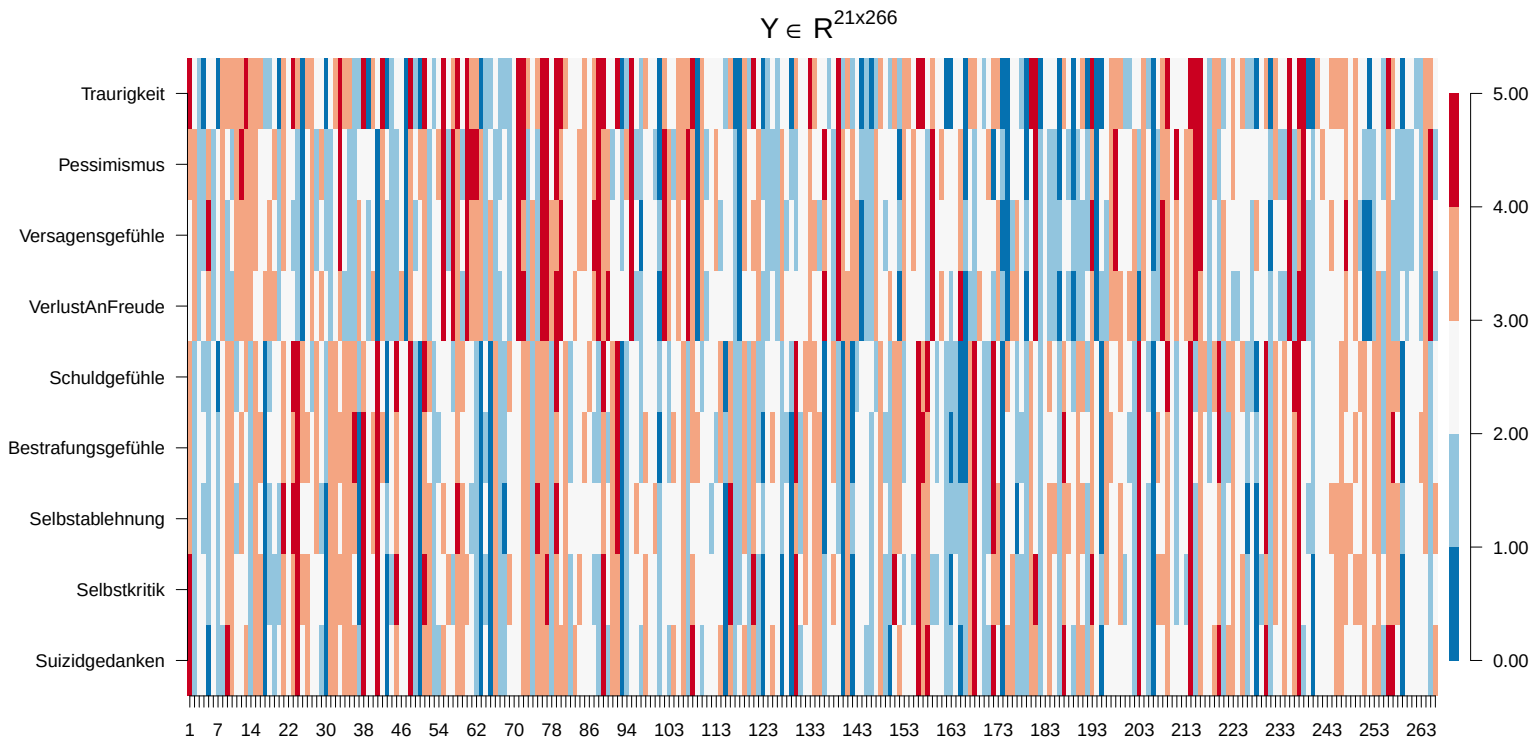
- Patient:innen 1 - 12 von n = 266 depressiven Patient:innen

	1	2	3	4	5	6	7	8	9	10	11	12
Traurigkeit	4	2	1	0	2	2	0	3	3	3	3	3
Pessimismus	3	3	1	1	3	1	2	3	2	1	3	4
Versagensgefühle	2	3	1	1	4	1	2	3	1	2	3	3
VerlustAnFreude	2	3	1	2	3	1	2	3	1	1	3	3
Schuldgefühle	3	1	2	1	1	2	0	2	3	3	1	2
Bestrafungsgefühle	3	1	2	2	1	2	1	2	3	3	2	3
Selbstablehnung	3	1	2	1	1	2	1	2	3	3	1	3
Selbstkritik	4	1	2	2	1	2	1	2	3	3	2	2
Suizidgedanken	4	1	2	2	0	2	1	1	4	3	2	2

Anwendungsbeispiel

Simulierter Datensatz basierend auf Keller et al. (2008)

- Gesamter Datensatz (n = 266)



Definition (Kovarianzmatrix eines Zufallsvektors)

ξ sei ein n-dimensionaler Zufallsvektor. Dann ist die *Kovarianzmatrix* von ξ definiert als die $n \times n$ Matrix

$$\mathbb{C}(\xi) := \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right). \tag{1}$$

Bemerkungen

- Die Kovarianzmatrix eines Zufallsvektors ist die symmetrische Matrix der Kovarianzen seiner Komponenten.
- Insbesondere gilt (vgl. Theorem Eigenschaften der Kovarianzmatrix)

$$\mathbb{C}(\xi) = \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1 \leq i,j \leq m}. \tag{2}$$

- Die Diagonalelemente von $\mathbb{C}(\xi)$ sind die Varianzen der Komponenten von ξ, da

$$\mathbb{V}(\xi_i) = \mathbb{C}(\xi_i, \xi_i) \text{ für } i = 1, \ldots m. \tag{3}$$

Definition (Korrelationsmatrix eines Zufallsvektors)

ξ sei ein n-dimensionaler Zufallsvektor. Dann ist die *Korrelationsmatrix* von ξ definiert als die $n \times n$ Matrix

$$\mathbb{R}(\xi) := \left(\rho_{ij}\right)_{1 \leq i,j \leq n} = \left(\frac{\mathbb{C}(\xi_i, \xi_j)}{\sqrt{\mathbb{V}(\xi_i)}\sqrt{\mathbb{V}(\xi_j)} \right)_{1 \leq i,j \leq n} . \tag{4}$$

Bemerkungen

- Die Korrelationsmatrix ist in der Kovarianzmatrix implizit.
- Es gilt, dass $\rho_{ij} \in [-1, 1]$ für $1 \leq i, j \in n$ und $\rho_{ii} = 1$ für $1 \leq i \leq n$.

Definition (Stichprobenkovarianzmatrix und -korrelationsmatrix)

$y_1, \ldots, y_n$ sei eine Menge von m-dimensionalen Zufallsvektoren, genannt *Stichprobe*. und es sei

$$\bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i. \tag{5}$$

das Stichprobenmittel der $y_1, \ldots, y_n$. Dann gelten ist die *Stichprobenkovarianzmatrix* der $y_1, \ldots, y_n$ definiert als die $m \times m$ Matrix

$$C := \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})^T. \tag{6}$$

und die *Stichprobenkorrelationsmatrix* der $y_1, \ldots, y_n$ ist definiert als die $m \times m$ Matrix

$$R := \left(\frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \right)_{1 \leq i,j \leq m}. \tag{7}$$

Bemerkungen

- Haben unabhängige $y_1, \ldots, y_n$ die identische Kovarianzmatrix $\mathbb{C}(y)$, so dient C als Schätzer von $\mathbb{C}(y)$
- Haben unabhängige $y_1, \ldots, y_n$ die identische Korrelationsmatrix $\mathbb{R}(y)$, so dient R als Schätzer von $\mathbb{R}(y)$
- Für eine konzise Berechnung von C und R, siehe Theorem Datenmatrix und Stichprobenstatistiken

Anwendungsbeispiel

Simulierter Datensatz basierend auf Keller et al. (2008)

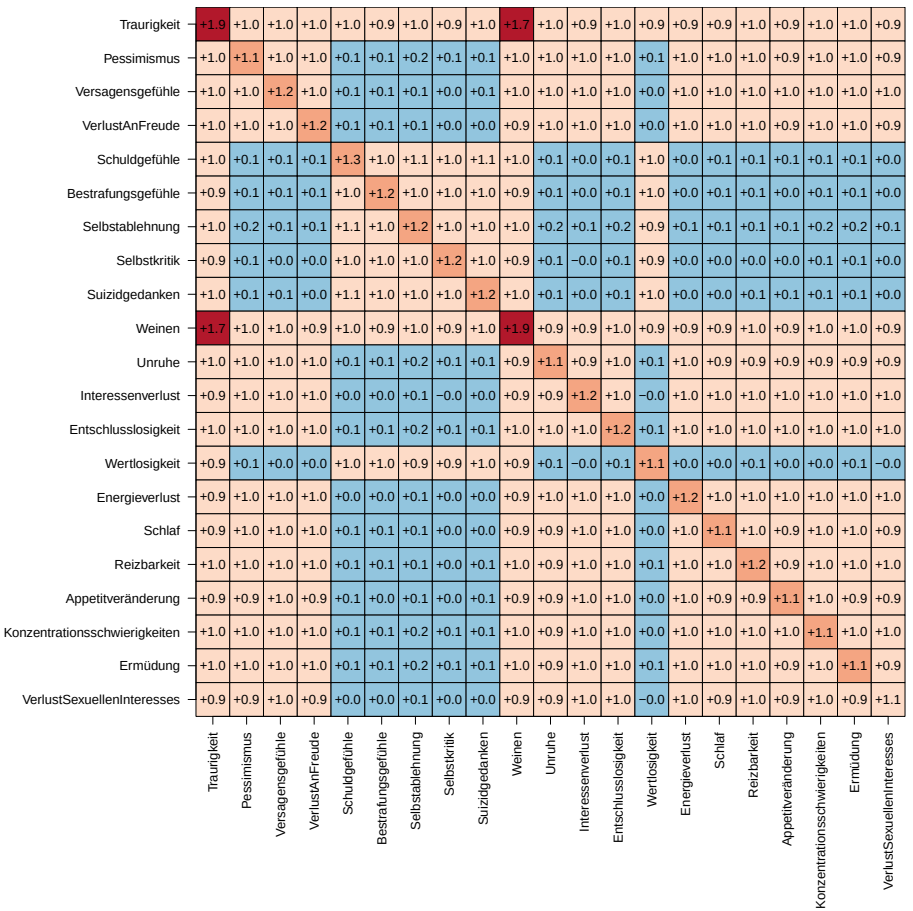
- Berechnung von Stichprobenkovarianzmatrix und Stichprobenkorrelationsmatrix

```
Y         = t(read.csv("3-Daten/3-fa-keller.csv"))          # Dateneinlesen
n         = ncol(Y)                                         # Anzahl Datenpunkte
I_n       = diag(n)                                         # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                    # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))      # Stichprobenkovarianzmatrix
D         = diag(1/sqrt(diag(C)))                           # Kov-Korr-Transformationsmatrix
R         = D %*% C %*% D                                   # Stichprobenkorrelationsmatrix
```

Anwendungsbeispiel

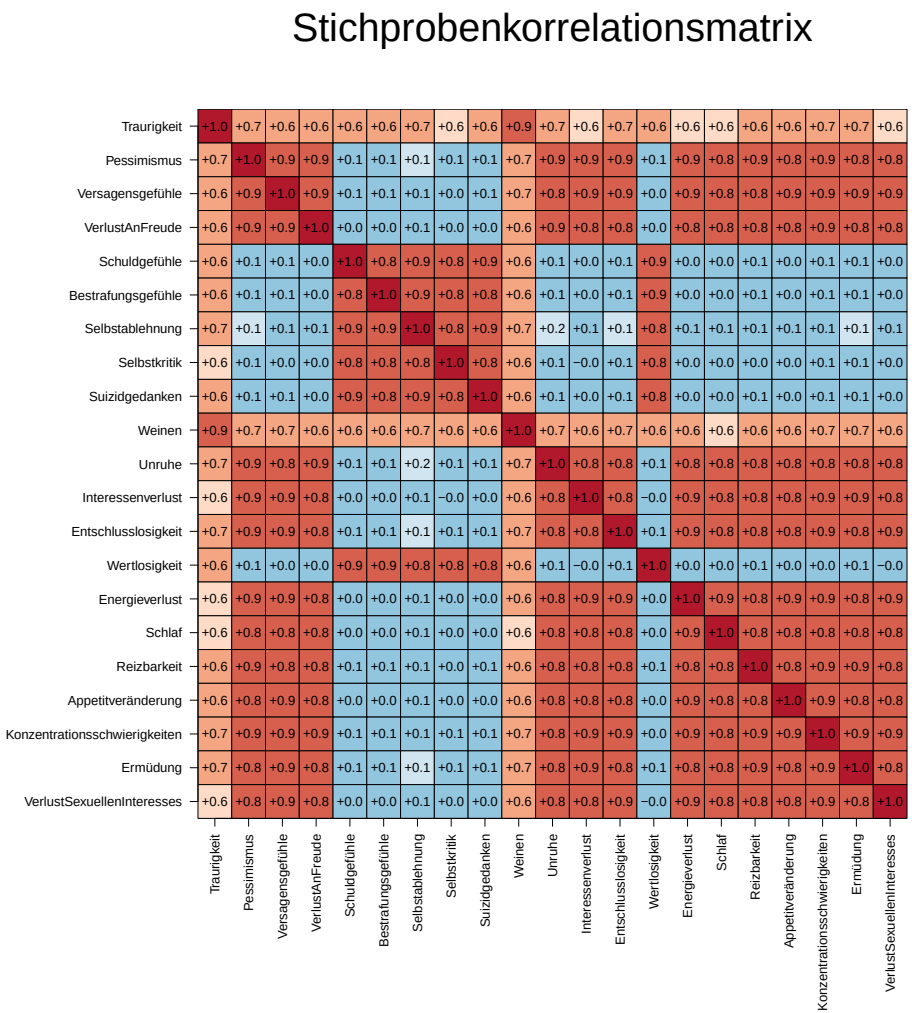
Simulierter Datensatz basierend auf Keller et al. (2008)

Stichprobenkovarianzmatrix



Anwendungsbeispiel

Simulierter Datensatz basierend auf Keller et al. (2008)



Grundlagen

Modellformulierung

Modellschätzung

Modellevaluation

Modellinterpretation

Selbstkontrollfragen

Definition (Faktorenanalysemodelle in struktureller Form)

Es sei

$$y = Lf + \varepsilon \tag{8}$$

wobei für $m > k$

- y ein m-dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird, ist,
- $L = (l_{ij}) \in \mathbb{R}^{m \times k}$ eine Matrix ist, die *Faktorladungsmatrix* genannt wird,
- f ein k-dimensionaler nicht-beobachtbarer Zufallsvektor ist, dessen Komponenten *Faktoren* genannt werden und für den gilt, dass

$$\mathbb{E}(f) = 0_k \text{ und } \mathbb{C}(f) = I_k, \tag{9}$$

- ε ein m-dimensionaler nicht-beobachtbarer und von f unabhängiger Zufallsvektor ist, der *Beobachtungsfehler* genannt wird und für den gilt, dass

$$\mathbb{E}(\varepsilon) = 0_m \text{ und } \mathbb{C}(\varepsilon) = \text{diag}\,(\psi_1, \ldots, \psi_m) =: \Psi \text{ mit } \psi_i > 0 \text{ für } i = 1, \ldots, m. \tag{10}$$

Dann heißt (4) *Faktorenanalysemodell in struktureller Form*. Gelten darüberhinaus, dass

- $f \sim N(0_k, \Phi)$ mit $\Phi \in \mathbb{R}^{k \times k}$ p.d. und
- $\varepsilon \sim N(0_m, \Psi)$ mit $\Psi := \text{diag}\,(\psi_1, \ldots, \psi_m)$ für $\psi_i > 0$ und $i = 1, \ldots, m$,

dann heißt (4) *Normalverteilungsmodell der Faktorenanalyse in struktureller Form*.

Modellformulierung

Bemerkungen

- Die Komponenten von y modellieren m *Itemscores* eines psychologischen Tests einer Testperson.
- Die Komponenten von f werden manchmal *gemeinsame Faktoren (common factors)* genannt.
- Die Komponenten von ε werden manchmal *spezifische Faktoren (unique factors)* genannt.
- l_{ij} wird die *Faktorladung des $j$ten Faktors auf die $i$te Datenkomponente* genannt.
- Der Eintrag l_{ij} von L entsprichtdem Wert y_i für $f = e_j$ ($j$ter kanonischer Basisvektor).
- Wir schreiben das Faktorenanalysemodell meist in der Kurzform

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_m, \Psi). \tag{11}$$

- $\zeta \sim (\mu, \Sigma)$ soll ausdrücken, dass $\mathbb{E}(\zeta) = \mu$ und $\mathbb{C}(\zeta) = \Sigma$.

Theorem (Faktorenanalysemodelle in Verteilungsform)

Gegeben seien die Faktorenanalysemodell in struktureller Form. Dann kann das Faktorenanalysemodell äquivalent geschrieben werden als

$$\mathbb{P}(f, y) = \mathbb{P}(f)\mathbb{P}(y|f) \text{ mit } f \sim (0_k, I_k) \text{ und } y \,|\, f \sim (Lf, \Psi) \tag{12}$$

und das Normalverteilungsmodell der Faktorenanalyse kann äquivalent in der Wahrscheinlichkeitsdichteform

$$p(f, y) = p(f)p(y|f) \text{ mit } p(f) = N(0_k, \Phi) \text{ und } p(y|f) = N(Lf, \Psi). \tag{13}$$

geschrieben werden.

Bemerkungen

- In Verteilungsform ist das Modell analog zum vereinfachten Modell der klassischen Testtheorie
- Die Bedingtheit der Verteilung von y auf f erschließt sich aus datengenerativer Sicht

(1) Es wird ein k-dimensionaler Wert von f anhand von $\mathbb{P}(f)$ realisiert
(2) f wird durch L transformiert und bildet als m-dimensionaler Wert Lf den Erwartungswert von $\mathbb{P}(y|f)$.
(3) Es wird ein m-dimensionaler Wert von ε realisiert
(4) Die m-dimensionalen Werte von Lf und ε werden zu einem Wert von y addiert.

Modellformulierung

Bemerkungen

- Im datenanalytischen Kontext betrachtet man die vorliegenden m Itemscores einer Testperson $i = 1, ..., n$, die einen psychologischen Fragebogen oder Test bearbeitet hat, als Realisierung eines Testperson-spezifischen und anhand der Verteilung des Faktorenanalysemodells verteilten Zufallsvektors y_i.
- Gleichzeitig unterstellt man (man analog zur Annahme eines wahren Werts der klassischen Testtheorie), dass dieser Zufallsvektorrealisierung eine Realisierung des nicht-beobachtbaren Zufallsvektors f_i, also der Testperson-spezifischen Faktorwerte, entspricht.
- Weiterhin nimmt man an, dass die Realisierungen von beobachtbaren Zufallsvektoren $y_1, ..., y_n$ und ihren entsprechenden nicht-beobachtbaren Zufallsvektoren $f_1, ..., f_n$ über Testperson unabhängig und identisch nach einem zugrundeliegenden Testpersonen-übergreifenden Faktorenanalysemodell verteilt sind.
- Entscheidend ist, dass man bei entsprechendem Vorliegen einer Datenmatrix $Y \in \mathbb{R}^{m \times n}$, unterstellt, dass (analog zum wahren Wert der klassischen Testtheorie) zu jedem Spalteneintrag ein "virtueller", nicht-beobachteter, Testpersonen-spezifischer, Faktorenvektorwert korrespondiert.

Modellformulierung

Bemerkungen

- Für $k := 2$ Faktoren, $m := 4$ Testitems und $n := 10$ Testpersonen z.B.

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
f_{i_1}	2	5	1	0	4	3	2	1	5	0
f_{i_2}	4	0	2	5	3	4	1	2	0	5
y_{i_1}	1	3	2	5	0	4	2	5	3	1
y_{i_2}	0	2	5	3	4	1	5	0	2	3
y_{i_3}	3	1	4	2	5	0	4	1	3	2
y_{i_4}	5	4	0	3	1	2	0	4	5	3

- Inferenz bezüglich Testpersonen-spezifischer Faktorwerte, also die Parameterschätzer-basierte Angabe der wahrscheinlichsten Werte des Testpersonen-spezifischen Faktorvektors, wird im Kontext der Faktorenanalyse als *Evaluation von Factorscores* bezeichnet,
- Die Evaluation von Factorscores steht in der Anwendung allerdings meist nicht im Vordergrund.
- Das Obige zusammenfassen ergeben sich die Datenzsatzmodelle der Faktorenanalyse

Definition (Datensatzmodelle der Faktorenanalyse)

Es sei $\mathbb{P}(f, y)$ das Faktorenanalyse in Verteilungsform. Dann hat das entsprechende Datensatzmodell für n Testpersonen die Form

$$\mathbb{P}(f_1, y_1, ..., f_n, y_n) = \prod_{i=1}^{n} \mathbb{P}(f_i, y_i) \text{ mit } \mathbb{P}(f_i, y_i) = \mathbb{P}(f_j, y_j) \text{ für } 1 \leq i, j \leq n \tag{14}$$

oder in Kurzform

$$(f_1, y_1), ..., (f_n, y_n) \sim \mathbb{P}(f, y) \tag{15}$$

Sei weiterhin $p(f, y)$ das Normalverteilungsmodell der Faktorenanalyse in Verteilungsform. Dann hat das entsprechende Datensatzmodell für n Testpersonen die Form

$$p(f_1, y_1, ..., f_n, y_n) = \prod_{i=1}^{n} p(f_i, y_i) \text{ mit } p(f_i, y_i) = p(f_j, y_j) \text{ für } 1 \leq i, j \leq n \tag{16}$$

Die entsprechende spaltenweise Konkatenation der Daten,

$$Y = (y_1, ..., y_n) \tag{17}$$

nennen wir einen *Datensatz*.

Bemerkungen

- Die Formulierung ist analog zu einer vektorisierten Form des Modells multipler Testmessungen.
- $p(f_1, y_1, ..., f_n, y_n)$ ist die multivariate Normalverteilung eines $n(m + k)$-dimensionalen Zufallsvektors.

Beispiel (1) Spearmans g-Faktor Modell

Grundlage von Spearman's Intelligenzmodell ist die Korrelationsstruktur von Leistungswerten in $m = 6$ Themenbereichen (Classics, French, English, Mathematics, Pitch discrimination, Musical talent) in einer Stichprobe von ungefähr 30 Kindern im Alter von 9 und 13 Jahren (Yanai & Ichikawa (2007)).

Spearman (1904) erklärt die entsprechenden Leistungswerte y_i mithilfe eines Faktorenanalysemodells der Form

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{pmatrix} f + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \Leftrightarrow y_i = l_i f + \varepsilon_i \text{ für } i = 1, \ldots, 6. \tag{18}$$

Dabei bezeichnet die latente Zufallsvariable f den *Allgemeinen Faktor der Intelligenz* und wird traditionell als *g-Faktor* bezeichnet. Für die Faktorladungen $l_i, i = 1, \ldots, 6$ fordert Spearman (1904) $|l_i| < 1$. Die Komponenten ε_i werden im Kontext von Spearman (1904) als (Themenbereichs-)*spezifische Faktoren* bezeichnet.

Entsprechend wird auch oft von *Spearman's Zweifaktoren Modell* gesprochen, wobei es im Sinne der modernen Faktorenanalysea allerdings nur einen Faktor und einen Vektor von Zufallsfehlern in diesem Modell gibt.

Beispiel (2) Thurstones' Multifaktorenmodell

L. L. Thurstone (1947) schlägt mit dem Multifaktorenmodell ein generelles Modell von k Faktoren vor, um die Kovarianzstruktur eines m-dimensionalen Zufallskvetors, der m Leistungstestwerte einer Gruppe von Testpersonen modelliert, zu erklären. Dabei soll $k < m$ gelten.

In struktureller Form hat dieses Modell die Form

$$y_i = l_{i1} f_1 + l_{i2} f_2 + \cdots + l_{ik} f_k + \varepsilon_i \text{ für } i = 1, \ldots, m \tag{19}$$

und entspricht damit der allgemeinen Form eines Faktorenanalysemodell, wobei hier die Faktoren $f_1, \ldots, f_k$ als *gemeinsame Faktoren (common factors)* und die Zufallsfehler $\varepsilon_1, \ldots, \varepsilon_m$ als *spezifische Faktoren (unique factors)* bezeichnet werden.

In L. Thurstone (1936) beispielsweise wird versucht zu zeigen, dass zur Erklärung der beobachteten Kovarianzstruktur von Leistungstestdaten von 240 Testperson die sieben Faktoren *Memory*, *Word fluency*, *Verbal relation*, *Number*, *Perceptual Speed*, *Visualization* und *Reasoning*, benötigt werden.

Beispiel (3) Das Modell paralleler Testmessungen

$y_1, \ldots, y_m$ seien die beobachtbaren Itemwerte eines psychologischen Tests und sei das Modell paralleler Testmessungen für den wahren Wert τ in der Form

$$y_i = \tau + \varepsilon_i \text{ für } i = 1, \ldots, m \tag{20}$$

mit Messfehler ε_i.

Dann entspricht dieses Modell dem Faktorenanalysemodell

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tau + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{21}$$

Insbesondere entspricht der Faktor f hier also dem wahren Wert τ und die Faktorladungsmatrix hat die als bekannt vorausgesetzte Form $L := 1_m$.

Theorem (Marginale Datenkovarianzmatrix des Faktorenanalysemodells)

Gegeben sei ein Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_m, \Psi). \tag{22}$$

Dann gilt für die marginale Kovarianzmatrix des Datenvektors

$$\mathbb{C}(y) = LL^T + \Psi. \tag{23}$$

Beweis

Mit Theorem Eigenschaften der Kovarianzmatrix gilt aufgrund der Unabhängigkeit von f und ε

$$\mathbb{C}(y) = L\mathbb{C}(f)L^T + \mathbb{C}(\varepsilon) = LI_kL^T + \Psi = LL^T + \Psi. \tag{24}$$

□

Bemerkungen

- Das Theorem ist die zentrale Eigenschaft des Faktorenanalysemodells
- Im Normalverteilungsmodell gilt mit dem Theorem zu marginalen Normalverteilungen entsprechend

$$\mathbb{C}(y) = L\Phi L^T + \Psi. \tag{25}$$

Theorem (Varianzzerlegung der faktoranalytischen Datenkomponenten)

Gegeben sei ein Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_k, \Psi). \tag{26}$$

Dann ist für $i = 1, \ldots, m$ die Varianz der $i$ten Komponente von y gegeben durch

$$\mathbb{V}(y_i) = \sum_{j=1}^{k} l_{ij}^2 + \psi_i. \tag{27}$$

Bemerkungen

- Es gilt bekanntlich $\mathbb{V}(y_i) = \mathbb{C}(y_i, y_i)$.

Beweis

Mit Theorem Marginale Datenkovarianzmatrix des Faktorenanalysemodells gilt

$$
\begin{aligned}
\mathbb{C}(y) &= LL^T + \Psi \\
&= \begin{pmatrix} l_{11} & \cdots & l_{1k} \\ l_{21} & \cdots & l_{2k} \\ \vdots & \ddots & \vdots \\ l_{m1} & \cdots & l_{mk} \end{pmatrix} \begin{pmatrix} l_{11} & \cdots & l_{m1} \\ l_{12} & \cdots & l_{m2} \\ \vdots & \ddots & \vdots \\ l_{1k} & \cdots & l_{mk} \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \psi_m \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^k l_{1j}l_{1j} & \sum_{j=1}^k l_{1j}l_{2j} & \cdots & \sum_{j=1}^k l_{1j}l_{mj} \\ \sum_{j=1}^k l_{2j}l_{1j} & \sum_{j=1}^k l_{2j}l_{2j} & \cdots & \sum_{j=1}^k l_{2j}l_{mj} \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{j=1}^k l_{mj}l_{1j} & \sum_{j=1}^k l_{mj}l_{2j} & \cdots & \sum_{j=1}^k l_{mj}l_{mj} \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \psi_m \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^k l_{1j}^2 + \psi_1 & \sum_{j=1}^k l_{1j}l_{2j} & \cdots & \sum_{j=1}^k l_{1j}l_{mj} \\ \sum_{j=1}^k l_{2j}l_{1j} & \sum_{j=1}^k l_{2j}^2 + \psi_2 & \cdots & \sum_{j=1}^k l_{2j}l_{mj} \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{j=1}^k l_{mj}l_{1j} & \sum_{j=1}^k l_{mj}l_{2j} & \cdots & \sum_{j=1}^k l_{mj}^2 + \psi_m \end{pmatrix}.
\end{aligned}
\tag{28}
$$

Für den $i$ten Diagonaleintrag von $\mathbb{C}(y)$ aber gilt dann bekanntlich $\mathbb{C}(y_i, y_i) = \mathbb{V}(y_i)$.

Definition (Kommunalität und Spezifität)

Gegeben sei ein Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_k, \Psi). \tag{29}$$

Dann werden in

$$\mathbb{V}(y_i) = \sum_{j=1}^{k} l_{ij}^2 + \psi_i \tag{30}$$

- $h_i^2 := \sum_{j=1}^{k} l_{ij}^2$ die *Kommunalität von* y_i und
- ψ_i die *Spezifität von* y_i

genannt.

Bemerkungen

- Die Kommunalität von y_i ist der durch die Faktorladungen erklärte Varianzanteil von y_i
- Die Spezifität von y_i ist der nicht durch die Faktorladungen erklärte Varianzanteil von y_i
- Die Spezifität ist also für jede Datenkomponente *spezifisch*.
- Für jede Datenkomponente eines Faktorenanalysemodells gilt

$$\text{Varianz} \;=\; \text{Kommunalität} \;+\; \text{Spezifität}. \tag{31}$$

Definition (Gesamtvarianz)

Gegeben sei ein Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_k, \Psi). \tag{32}$$

Dann wird

$$\mathbb{G} := \sum_{i=1}^{m} \mathbb{V}(y_i) = \sum_{i=1}^{m} \sum_{j=1}^{k} l_{ij}^2 + \sum_{i=1}^{m} \psi_i \tag{33}$$

die *Gesamtvarianz von* y genannt.

Bemerkungen

- Die Gesamtvarianz des Datenvektors ist definiert als die Summe der Varianzen der Datenkomponenten
- Die Gesamtvarianz entspricht damit der *Spur (trace)* der marginalen Datenkovarianzmatrix.
- Für die Gesamtvarianz gilt mnemonisch also

$$\text{Gesamtvarianz } = \text{ Summe der Kommunalitäten } + \text{ Summe der Spezifitäten.} \tag{34}$$

Definition (Orthogonale Matrix)

Eine Matrix $Q \in \mathbb{R}^{m \times m}$ heißt *orthogonal*, wenn

$$Q^T Q = I_m. \tag{35}$$

Bemerkung

- Die Spalten einer orthogonalen Matrix sind also paarweise orthogonal, es gilt für

$$Q = \begin{pmatrix} q_1 & \cdots & q_n \end{pmatrix} \text{ mit } q_i \in \mathbb{R}^n \text{ für } 1 \leq i \leq m, \tag{36}$$

  dass

$$q_i^T q_j = 0 \text{ für } i \neq j \text{ und } q_i^T q_j = 1 \text{ für } i = j \text{ mit } 1 \leq i, j \leq m. \tag{37}$$

- Die Spalten einer orthogonalen $m \times m$ Matrix sind also m orthonormale Vektoren.

Theorem (Eigenschaften orthogonaler Matrizen)

$Q \in \mathbb{R}^{m \times m}$ sei eine orthogonale Matrix. Dann gelten:

(1) (Inverse) Die Inverse von Q ist Q^T, es gilt

$$Q^{-1} = Q^T. \tag{38}$$

(2) (Transposition) Die Zeilen von Q sind orthonormal, es gilt

$$QQ^T = I_m. \tag{39}$$

(3) (Determinante) Es gilt

$$|Q| = 1 \text{ oder } |Q| = -1. \tag{40}$$

Beweis

(1) Unter der Annahme, dass Q^{-1} existiert, gilt

$$Q^T Q = I_n \Leftrightarrow Q^T Q Q^{-1} = I_n Q^{-1} \Leftrightarrow Q^{-1} = Q^T. \tag{41}$$

(2) Mit Eigenschaft (1) gilt

$$QQ^T = QQ^{-1} = I_m. \tag{42}$$

(3) Mit dem Multiplikationssatz und der Transpositionseigenschaft von Determinanten gilt

$$|Q|^2 = |Q||Q| = |Q||Q^T| = |QQ^T| = |I_m| = 1. \tag{43}$$

Aus $|Q|^2 = 1$ folgt dann aber, dass $|Q| = 1$ oder $|Q| = -1$ sein muss. □

Definition (Orthogonale Transformation eines Faktorenanalysemodell)

Gegeben sei ein Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ mit } \varepsilon \sim (0_m, \Psi) \tag{44}$$

und es sei $Q \in \mathbb{R}^{k \times k}$ eine orthogonale Matrix. Dann nennen wir

$$\tilde{y} = \tilde{L}\tilde{f} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{f} := Q^T f \tag{45}$$

eine *orthogonale Transformation des Faktorenanalysemodell*.

Bemerkungen

- Die orthogonale Transformation eines Faktorenanalysemodells ist Grundlage seiner Nichtidentifizierbarkeit.
- Die orthogonale Transformation eines Faktorenanalysemodells ist Grundlage "explorativen Faktorenanalyse".
- Die explorative Faktorenanalyse nutzt orthogonale Transformationen zur Interpretationsmaximierung.

Theorem (Nichtidentifizierbarkeit und Kovarianzinvarianz)

Gegeben sei ein Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ mit } \varepsilon \sim (0_m, \Psi) \tag{46}$$

sowie eine seiner orthogonale Transformationen

$$\tilde{y} = \tilde{L}\tilde{f} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{f} := Q^T f \tag{47}$$

für eine orthogonale Matrix $Q \in \mathbb{R}^{k \times k}$. Dann gelten

$$y = \tilde{y} \text{ und } \mathbb{C}(\tilde{y}) = \mathbb{C}(y) \tag{48}$$

Beweis

Es gilt zum einen

$$\tilde{y} = \tilde{L}\tilde{f} + \varepsilon = LQQ^T f + \varepsilon = LI_k f + \varepsilon = Lf + \varepsilon = y. \tag{49}$$

Es gilt zum anderen

$$\mathbb{C}(\tilde{y}) = LQ(LQ)^T + \Psi = LQQ^T L^T + \Psi = LI_k L^T + \Psi = LL^T + \Psi = \mathbb{C}(y). \tag{50}$$

□

Bemerkungen

- Orthogonale Transformationen lassen Datenvektor und Kovarianzmatrix unberührt.
- Aus
$$y = Lf + \varepsilon = \tilde{L}\tilde{f} + \varepsilon \tag{51}$$
folgt unmittelbar, dass für einen festen Wert von y weder L noch f eindeutig bestimmt sind.
- Verschiedene Faktorladungsmatrizen und Faktorwerte können also die gleichen Daten erklären.
- Aus einer gegebenen Stichprobenkovarianzmatrix kann nicht eindeutig auf L und f geschlossen werden.
- Aus
$$\mathbb{C}(y) = LL^T + \Psi = \tilde{L}\tilde{L}^T + \Psi \tag{52}$$
folgt aber auch, dass sich Gesamtvarianz und Kommunalitäten bei orthogonaler Transformation nicht ändern.
- Fundamental ist die Nichtidentifizierbarkeit des Faktorenanalysemodell dadurch bedingt, dass nach Modellannahme weder die Faktorladungsmatrix, noch die Werte der Faktoren, noch die Beobachtungsfehler bekannt sind und Daten daher durch das Zusammenbeispiel dreier unbekannter Entitäten generiert werden.

Definition (Identifizierbares Normalverteilungsmodell der Faktorenanalyse)

Gegeben sei ein Normalverteilungsmodell der Faktorenanalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{53}$$

Weiterhin sei der *Parametervektor* dieses Modells definiert als die spaltenweise Konkatenation der Parametermatrizen L, Φ, Ψ, also

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{54}$$

so dass die marginale Datenkovarianz geschrieben werden kann als

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{55}$$

Dann heißen das Normalverteilungsmodell der Faktorenanalyse und der Parametervektor θ *identifizierbar*, wenn für alle $\theta_1, \theta_2 \in \Theta$ gilt, dass

$$\Sigma_{\theta_1} = \Sigma_{\theta_2} \Rightarrow \theta_1 = \theta_2. \tag{56}$$

Wenn für $\theta_1 \neq \theta_2$ gilt, dass $\Sigma_{\theta_1} = \Sigma_{\theta_2}$, so heißen das Modell und θ *nicht identifizierbar*.

Bemerkungen

- Normalverteilungsmodelle können durch zusätzliche Parameterannahmen identifizierbar gemacht werden.
- Entscheidend ist die Anzahl der Parameter und der beobachteten Statistiken (Kovarianzen).

Theorem (Anzahl unikaler skalarer Parameter und Statistiken)

Gegeben sei ein Normalverteilungsmodell der Faktorenanalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{57}$$

Dann gilt für die Anzahl an unikalen skalaren Parametern des Modells

$$n_\theta = mk + \frac{k(k+1)}{2} + m \tag{58}$$

Weiterhin sei C sei die Stichprobenkovarianzmatrix eines Datensatzes von n unabhängigen Beobachtungen von y. Dann gilt für die Anzahl an unikalen skalaren Stichprobenstatistiken des Normalverteilungsmodell der Faktorenanalyse

$$n_c = \frac{m(m+1)}{2}. \tag{59}$$

Bemerkungen

- n_θ ist die Dimension des unikalen Parametervektors $\theta \in \mathbb{R}^{n_\theta}$.
- n_c ist die Anzahl der unikalen skalaren Einträge der Stichprobenkovarianzmatrix C.
- Das Verhältnis von n_θ und n_c ist die Grundlage der *Ordnungsbedingung* zur Identifizierbarkeit.

Beweis

Die Anzahl der Einträge der Faktorladungsmatrix $L \in \mathbb{R}^{m \times k}$ ist mk.

Die Anzahl der Einträge einer symmetrischen Matrix $S \in \mathbb{R}^{k \times k}$ ist k^2. Aufgrund der Symmetrie von S sind dabei allerdings nur die k Einträge der Hauptdiagonale und die Einträge oberhalb (oder unterhalb) der Hauptdiagonalen unikal. Die Anzahl der Einträge oberhalb (oder unterhalb) der Hauptdiagonalen ist die Hälfte aller $k^2 - k$ nicht-diagonalen Einträge von S, also $(k^2 - k)/2$. Zusammen mit den Einträgen auf der Hauptdiagonalen ergibt sich damit für die Anzahl unikaler Einträge einer symmetrischen Matrix

$$\frac{k^2 - k}{2} + k = \frac{k^2 - k}{2} + \frac{2k}{2} = \frac{k^2 + k}{2} = \frac{k(k+1)}{2}. \tag{60}$$

Als Kovarianzmatrix ist $\Phi \in \mathbb{R}^{k \times k}$ symmetrisch und hat damit $\frac{k(k+1)}{2}$ unikale Einträge. Die Anzahl der von Null verschiedenden Einträge von $\Psi \in \mathbb{R}^{m \times m}$ ist m.

Für die Anzahl an unikalen skalaren Parametern des Normalverteilungsmodell der Faktorenanalyses e rgibt sich also zusammenfassend

$$n_\theta = mk + \frac{k(k+1)}{2} + m. \tag{61}$$

Die Stichprobenkovarianzmatrix $C \in \mathbb{R}^{m \times m}$ eines Datensatzes ist symmetrisch. Mit obigen Überlegungen zu den unikalen Einträgen einer symmetrischen Matrix ergibt sich also direkt

$$n_c = \frac{m(m+1)}{2}. \tag{62}$$

Definition (Ordnungsbedingung)

Gegeben sei ein Normalverteilungsmodell der Faktorenanalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{63}$$

Dann sagt man, dass das Modell der *Ordnungsbedingung* genügt, wenn für die Anzahl n_θ der unikalen skalaren Parameter und für die Anzahl n_c der unikalen skalaren Statistiken gilt, dass

$$n_\theta \leq n_c. \tag{64}$$

Bemerkungen

- Softwarelösungen wie Lavaan implementieren die Ordnungsbedingung meist per default (vgl. Rosseel (2021))).

Grundlagen

Faktorenanalysemodelle

Modellschätzung

Modellevaluation

Modellinterpretation

Selbstkontrollfragen

Modellschätzung

Überblick

- Die Parameter des Faktorenanalysemodells sind L und Ψ (und Φ)
- Die Schätzung dieser Parameter geht von einer vordefinierten Anzahl k von Faktoren aus
- In der psychologischen Literatur wird die Modellschätzung auch als *Faktorenextraktion* bezeichnet
- Es sind mindestens drei verschiedene Verfahren im Einsatz, da sie mit Toolboxen leicht zu verwenden sind
- Zur Popularität verschiedener Schätzverfahren siehe Fabrigar et al. (1999) und Goretzko et al. (2021)

Hauptkomponentenschätzung

- Basaler Ansatz zur Schätzung von L und Ψ nach Hotelling (1933) und L. L. Thurstone (1935)
- Die Zielsetzung einer *Hauptkomponentanalyse* ist i.d.R. nicht die Schätzung von L und Ψ
- `fa()`mit `fm = "pc"` im `psych` R Paket

Iterierte Hauptachsenfaktorisierung

- Iterativer, basaler Ansatz zur Schätzung von L und Ψ nach Horst et al. (1941) und Horst (1965)
- Weitere Bezeichnungen sind *Principal factor analysis* und *Principal axis analysis* (Hauptachsenanalyse)
- `fa()`mit `fm = "pa"` im `psych` R Paket

Maximum-Likelihood-Schätzung

- Probabilistischer Modellansatz nach Lawley (1940) und Lawley & Maxwell (1971)
- Numerische Maximizierung der Log-Likelihood-Funktion des Faktorenanalysen-Normalverteilungsmodells
- `fa()`mit `fm = "ml"` im `psych` R Paket

Theorem (Orthonormalzerlegung einer symmetrischen Matrix)

$S \in \mathbb{R}^{m \times m}$ sei eine symmetrische Matrix mit m verschiedenen Eigenwerten. Dann kannn S geschrieben werden als

$$S = Q \Lambda Q^T, \tag{65}$$

wobei $Q \in \mathbb{R}^{m \times m}$ eine orthogonale Matrix ist und $\Lambda \in \mathbb{R}^{m \times m}$ eine Diagonalmatrix ist.

Bemerkungen

- Die Orthonormalzerlegung der Stichprobenkovarianzmatrix ist für die Modellschätzung grundlegend.
- Die Spalten von Q heißen *Eigenvektoren* von S, die Diagonaleinträge von Λ sind die *Eigenwerte* von S.
- Die Matrizen Q und Λ können mithilfe einer *Eigenanalyse* von S bestimmt werden (`eig(S)` in R).
- Die Theorie der Eigenanalyse ist ein weites Feld, wir geben einige Erläuterungen im Anhang.
- Die Eigenwerte und ihre zugehörigen Eigenvektoren sind absteigend sortiert.
- Für einen Beweis, siehe Theorem Orthonormalzerlegung einer symmetrischen Matrix

Beispiel einer Orthonormalzerlegung

Es sei

$$S := \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}. \tag{66}$$

Dann gilt mit

$$Q := \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \text{ und } \Lambda := \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \tag{67}$$

dass

$$\begin{aligned} S &= Q\Lambda Q^T \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T \\ &= \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right) \\ &= \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 3 & -1 \\ 3 & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right) \\ &= \frac{1}{2} \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \end{aligned}$$

Motivation der Hauptkomponentenschätzung

- Ziel der Hauptkomponentenschätzung ist die Approximation der Stichprobenkovarianzmatrix in der Form

$$C \approx \hat{L}\hat{L}^T + \hat{\Psi}. \tag{68}$$

- Die Hauptkomponentenschätzung vernachlässigt dabei zunächst $\hat{\Psi}$ und nutzt die Orthonormalzerlegung

$$C = Q\Lambda Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = \left(Q\Lambda^{1/2}\right)\left(Q\Lambda^{1/2}\right)^T \tag{69}$$

- Die Hauptkomponentenschätzung vernachlässigt dann die $k+1, \ldots, m$ Spalten von Q und die $k+1, \ldots, m$ Zeilen und Spalten von Λ und setzt

$$\hat{L}\hat{L}^T = Q_k\Lambda_k^{1/2}\left(Q_k\Lambda_k^{1/2}\right)^T \text{ mit } \hat{L} \in \mathbb{R}^{m\times k}. \tag{70}$$

- Für die Diagonalelemente c_{ii}, $\hat{h}_i^2$ und $\hat{\psi}_i$ von $C, \hat{L}\hat{L}^T$ bzw. $\hat{\Psi}$ folgt dann, dass

$$c_{ii} = \sum_{j=1}^{k}\hat{l}_{ij}^2 + \hat{\psi}_i \Leftrightarrow \hat{\psi}_i = c_{ii} - \sum_{j=1}^{k}\hat{l}_{ij}^2, \tag{71}$$

  worauf sich die entsprechenden Spezifitätsschätzer bei Hauptkomponentenschätzung gründen.

Definition (Hauptkomponentenschätzer $k$ter Ordnung von L und Ψ)

Gegeben sei ein Datensatz $Y \in \mathbb{R}^{m \times n}$ von n unabhängigen Beobachtungen eines Faktorenanalysemodell

$$y = Lf + \varepsilon \text{ mit } f \sim (0_k, I_k) \text{ und } \varepsilon \sim (0_m, \Psi). \tag{72}$$

$C \in \mathbb{R}^{m \times m}$ sei die Stichprobenkovarianzmatrix von Y und

$$C = Q \Lambda Q^T \tag{73}$$

sei ihre Orthonormalzerlegung mit spaltenweise der Größe nach sortierten Eigenwerten und zugehörigen Eigenvektoren. Dann sind die *Hauptkomponentenschätzer $k$ter Ordnung von L und Ψ* definiert als

$$\hat{L} := Q_k \Lambda_k^{1/2} \in \mathbb{R}^{m \times k} \text{ und } \hat{\Psi} := \text{diag}\left(\hat{\psi}_1, ..., \hat{\psi}_m\right), \tag{74}$$

wobei $Q_k \in \mathbb{R}^{m \times k}$ die Matrix bestehend aus den ersten k Spalten von $Q \in \mathbb{R}^{m \times m}$, $\Lambda_k \in \mathbb{R}^{k \times k}$ die Matrix bestehend aus den ersten k Zeilen und k Spalten von $\Lambda \in \mathbb{R}^{m \times m}$ und für $i = 1, ..., m$

$$\hat{\psi}_i := c_{ii} - \sum_{j=1}^{k} \hat{l}_{ij}^2 \tag{75}$$

mit den Diagonaleinträgen c_{ii} von C bezeichnen.

Bemerkungen

- Die Wahl der ersten k Spalten von Q und der ersten k Zeilen und k Spalten von Λ impliziert k Faktoren.

Definition (Kommunalitäts-, Spezifitäts-, und Gesamtvarianzschätzer)

Für einen Datensatz $Y \in \mathbb{R}^{m \times n}$ von n unabhängigen Beobachtungen und ein festes $k < m$ seien

- $\hat{L} = (\hat{l}_{ij})_{1 \leq i \leq m, 1 \leq j \leq k} \in \mathbb{R}^{m \times k}$ und
- $\hat{\Psi} = \text{diag}(\hat{\psi}_1, \ldots, \hat{\psi}_m) \in \mathbb{R}^{m \times m}$

die auf der Stichprobenkovarianzmatrix $C = (c_{ij})_{1 \leq i,j \leq m}$ basierenden Hauptkomponentenschätzer $k$ter Ordnung. Dann ergeben sich für $i = 1, \ldots, m$

- $\hat{h}_i^2 := \sum_{j=1}^{k} \hat{l}_{ij}^2$ als Schätzer von h_i^2 (*Kommunalitätsschätzer*),
- $\hat{\psi}_i$ als Schätzer von ψ_i (*Spezifitätsschätzer*),
- c_{ii} als Schätzer von $\mathbb{V}(y_i)$ (*Itemvarianzschätzer*),
- $G := \text{tr}(C)$ als Schätzer von $\mathbb{G}$ (*Gesamtvarianzschätzer*).

Bemerkungen

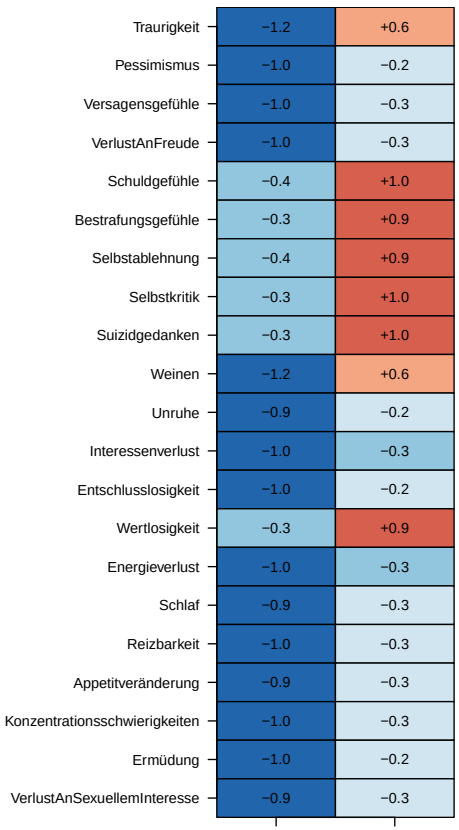
- Genauer handelt es sich um *Hauptkomponentenschätzer-$k$ter-Ordnung-basierte* Schätzer von h_i^2 und ψ_i.

Anwendungsbeispiel

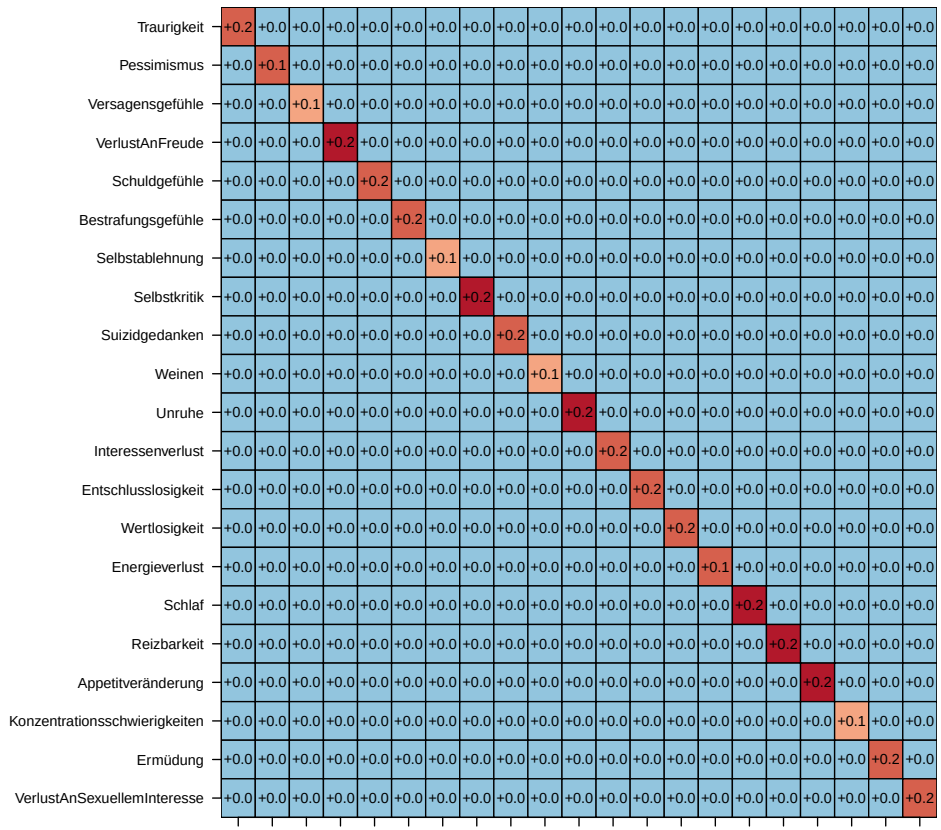
```
Y         = as.matrix(t(read.csv("3-Daten/3-fa-keller.csv"))) # Y \in \mathbb{R}^{m x n}
m         = nrow(Y)                                          # Datendimension
n         = ncol(Y)                                          # Datenpunktanzahl
k         = 2                                                # Faktoranzahl
I_n       = diag(n)                                          # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                     # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))       # Stichprobenkovarianzmatrix
EA        = eigen(C)                                         # Eigenanalyse von R
lambda_k  = EA$values[1:k]                                   # k größte Eigenwerte von R
Q_k       = EA$vectors[,1:k]                                 # k zugehörige Eigenvektoren von R
L_hat     = Q_k %*% diag(sqrt(lambda_k))                     # Faktorladungsmatrixschätzer
Psi_hat   = diag(diag(C) - diag(L_hat %*% t(L_hat)))         # Fehlerkovarianzmatrixschätzer
V_i_hat   = diag(C)                                          # Varianzschätzer
h2_i_hat  = rowSums(L_hat^2)                                 # Kommunalitätsschätzer
psi_i_hat = diag(Psi_hat)                                    # Spezifitätsschätzer
```

Anwendungsbeispiel

$\hat{L}$



$\hat{\Psi}$



Motivation der iterierten Hauptachsenfaktorisierung

- Ziel der iterierten Hauptachsenfaktorisierung ist wie bei der Hauptkomponentenschätzung die Approximation

$$C \approx \hat{L}\hat{L}^T + \hat{\Psi} \tag{76}$$

- Die Hauptkomponentenschätzung ignoriert $\hat{\Psi}$ zunächst vollständig
- Die Hauptkomponentenschätzung von L basiert sowohl auf Kommunalitäten und Spezifitäten
- Für die Diagonalelemente c_{ii} von C gilt nach dem Faktoranalysemodell aber

$$c_{ii} = \hat{\mathbb{V}}(y_i) = \sum_{j=1}^{k} \hat{l}_{ij}^2 + \hat{\psi}_i \text{ für } i = 1, \ldots, m \tag{77}$$

- Es wäre also wünschenswert, die Schätzung von L nur auf Basis der Kommunalitäten vorzunehmen.
- Die iterierte Hauptachsenfaktorisierung nutzt eine initiale Schätzung von $\hat{\Psi}$ und approximiert dann iterativ

$$C - \hat{\Psi} \approx \hat{L}\hat{L}^T \tag{78}$$

- Die iterierte Hauptachsenfaktorisierung favorisiert die Modellschätzung mit der Korrelationsmatrix

Definition (Iterierte Hauptachsenfaktorisierung)

Gegeben sei eine Stichprobenkorrelationsmatrix $R \in \mathbb{R}^{m \times m}$. Dann hat der Algorithmus der iterierten Hauptachsenfaktorisierung nach die folgende Form:

Initialisierung

(S0) Definiere ε_{min}, i_{max}, und setze $R^{(0)} := R$, $h^{(0)} := \sum_{j=1}^{m} R_{jj}^{(0)}$, $\varepsilon^{(1)} := h^{(0)}$ und $i := 1$

Iterationen

Solange $\varepsilon^{(i)} > \varepsilon_{\text{min}}$ und $i \leq i_{\text{max}}$

(S1) Zerlege $R^{(i-1)} = Q^{(i)} \Lambda^{(i)} Q^{(i)^T}$ und setze $\hat{L}^{(i)} := Q_k^{(i)} \Lambda_k^{(i)^{1/2}}$

(S2) Setze $R^{(i)} := R^{(i-1)}$ und $R_{jj}^{(i)} := \left(\hat{L}^{(i)} \hat{L}^{(i)^T} \right)_{jj}$ für $j = 1, \ldots, m$

(S3) Setze $h^{(i)} := \sum_{j=1}^{m} R_{jj}^{(i)}$, $\varepsilon^{(i+1)} := |h^{(i)} - h^{(i-1)}|$ und $i := i + 1$

Finalisierung

(S4) Setze $\hat{L} := \hat{L}^{(i)} D$ mit $D := \text{diag}\left(\text{sgn}\left(\sum_{j=1}^{m} \hat{l}_{j1}^{(i)} \right), \ldots, \text{sgn}\left(\sum_{j=1}^{m} \hat{l}_{jk}^{(i)} \right) \right)$

Bemerkungen

- Der Algorithmus beschreibt die Implementation im `psych` Paket nach Revelle (2024).

Anwendungsbeispiel

```r
Y             = as.matrix(read.csv("3-Daten/3-fa-keller.csv"))          # Y \in \mathbb{R}^{m x n}
R             = cor(Y)                                                  # m x m Korrelationsmatrix
k             = 2                                                       # Faktorenzahl
eps_min       = 0.01                                                    # Konvergenzkriterium
i_max         = 25                                                      # Maximale Anzahl an IPAF Iterationen
i             = 0                                                       # Iterationszähler
R_i           = R                                                       # Initiale Korrelationsmatrix R^{(0)}
h_i           = sum(diag(R_i))                                          # Summe der Kommunalitätsschätzer
eps_i         = h_i                                                     # Fehlerinitialisierung
while(eps_i > eps_min && i <= i_max){                                   # Iterationen
    QLQ_i       = eigen(R_i)                                            # Eigenanalyse
    L_hat_i     = QLQ_i$vectors[,1:k] %*% diag(sqrt(QLQ_i$values[1:k])) # Hauptkomponentenschätzer kter Ordnung
    diag(R_i)   = diag(L_hat_i %*% t(L_hat_i) )                         # Kommunalitätsschätzer Update
    h_ii        = sum(diag(R_i))                                        # Kommunalitätschätzersummen Update
    eps_i       = abs(h_ii - h_i)                                       # Fehlerkonvergenkriterium
    h_i         = h_ii                                                  # Kommunalitätsschätzersummen Update
    i           = i + 1}                                                # Iterationszähler Update
D             = diag(sign(colSums(L_hat_i)))                            # Vorzeichenfinalisierung
L_hat         = L_hat_i %*% D                                           # IHAF Faktorladungsmatrixschätzer
```

Motivation der Maximum-Likelihood-Schätzung

- Hauptkomponenten- und Hauptachsenfaktorisierungschätzer sind *adhoc* motiviert.
- Wünschenswert wären Schätzer mit guten und bekannten frequentistischen Eigenschaften.
- Als Normalverteilungsmodell besitzt das Faktorenanalysemodell eine Likelihood-Funktion.
- Maximum-Likelihood-Parameterschätzer ergeben sich durch Maximierung dieser Likelihood-Funktion.
- Die höhere Modellkomplexität gegenüber dem ALM bedingt numerische Herangehensweisen.
- Klassischerweise basiert die Maximum-Likelihood-Schätzung auf der *Diskrepanzfunktion*.
- Rosseel (2021) gibt dazu einen Überblick, insbesondere bezüglich von Schätzverfahren in `lavaan`.
- Wir formulieren die Log-Likelihood-Funktion und die Diskrepanzfunktion des Faktorenanalysemodells.
- Wir zeigen die Äquivalenz von Log-Likelihood- und Diskrepanzfunktion-basierten Schätzern.

Theorem (Maximum-Likelihood-Schätzung)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{79}$$

mit Parametervektor

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{80}$$

und marginaler Datenkovarianzmatrix

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{81}$$

Weiterhin sei $Y := (y_1, ..., y_n) \in \mathbb{R}^{m \times n}$ ein zentrierter Datensatz von n unabhängigen Beobachtungen von y, C seine Stichprobenkovarianzmatrix und

$$S := \frac{n-1}{n} C \tag{82}$$

seine verzerrte Stichprobenkovarianzmatrix. Dann kann die Log-Likelihood-Funktion von Y geschrieben werden als

$$\ell_Y : \Theta \to \mathbb{R}, \theta \mapsto \ell_Y(\theta) := -\frac{n}{2} \ln |\Sigma_\theta| - \frac{n}{2} \text{tr}\left(S\Sigma_\theta^{-1}\right) - \frac{nm}{2} \ln(2\pi) \tag{83}$$

und eine Maximumstelle von ℓ_Y, also ein Wert

$$\hat{\theta}_{\text{ML}} = \text{argmax}_{\theta \in \Theta} \ell_Y(\theta), \tag{84}$$

heißt *Maximum-Likelihood-Schätzer* für θ.

Beweis

Mit der Definition der Log-Likelihood-Funktion als und der marginalen Datenverteilung des Normalverteilungsmodells der Faktorenanalyse ergibt sich

$$\begin{aligned}
\ell_Y(\theta) &= \ln \prod_{i=1}^{n} p_\theta(y_i) \\
&= \prod_{i=1}^{n} \ln N\left(y_i; 0_m, L\Phi L^T + \Psi\right) \\
&= \ln \left(\prod_{i=1}^{n} (2\pi)^{-m/2} |\Sigma_\theta|^{-1/2} \exp\left(-\frac{1}{2}(y_i - 0_m)^T \Sigma_\theta^{-1}(y_i - 0_m)\right)\right) \\
&= -\frac{mn}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma_\theta| - \frac{1}{2} \sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i.
\end{aligned} \tag{85}$$

Um die Gleicheit des letzten Terms auf der rechten Seite mit dem letzten Term in der postulierten Funktionsform zu zeigen, halten wir zunächst fest, dass mit elementaren Eigenschaften der Matrixspur gilt, dass

$$\mathsf{tr}\left(\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i\right) = \sum_{i=1}^{n} \mathsf{tr}\left(y_i^T \Sigma_\theta^{-1} y_i\right) = \sum_{i=1}^{n} \mathsf{tr}\left(\Sigma_\theta^{-1} y_i y_i^T\right) = \mathsf{tr}\left(\Sigma_\theta^{-1} \sum_{i=1}^{n} y_i y_i^T\right). \tag{86}$$

Beweis (fortgeführt)

Weiterhin gilt mit dem Binomischen Lehrsatz

$$
\begin{aligned}
\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i &= \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y} + \bar{y}) \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \left(\sum_{i=1}^{n} \left(y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \right)^T \right) \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \left(\sum_{i=1}^{n} y_i^T - \frac{n}{n} \sum_{i=1}^{n} y_i^T \right) \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} + 2 \left(0_m^T \Sigma_\theta^{-1} \bar{y} \right) \\
&= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y}
\end{aligned}
\tag{87}
$$

Beweis (fortgeführt)

Also folgt mit obiger Eigenschaft der Matrixspur, der Zentrierung des Datensatzes $\bar{y} = 0_m$ und der Definition von S

$$\begin{aligned}
\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i &= \sum_{i=1}^{n} (y_i - \bar{y})^T \Sigma_\theta^{-1} (y_i - \bar{y}) + \sum_{i=1}^{n} \bar{y}^T \Sigma_\theta^{-1} \bar{y} \\
&= \sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i \\
&= \mathsf{tr}\left(\sum_{i=1}^{n} y_i^T \Sigma_\theta^{-1} y_i \right) \\
&= \mathsf{tr}\left(\Sigma_\theta^{-1} \sum_{i=1}^{n} y_i y_i^T \right) \\
&= \mathsf{tr}\left(\Sigma_\theta^{-1} n S \right) \\
&= n\, \mathsf{tr}\left(S \Sigma_\theta^{-1} \right)
\end{aligned} \tag{88}$$

Substitution in $\ell_Y(\theta)$ von oben ergibt dann die postulierte funktionale Form der Log-Likelihood Funktion.

□

Definition (Diskrepanzfunktion der Faktorenanalyse)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse

$$y = Lf + \varepsilon \text{ mit } f \sim N(0_k, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{89}$$

mit Parametervektor

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{90}$$

und marginaler Datenkovarianzmatrix

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{91}$$

Weiterhin sei für einen Datensatz $Y \in \mathbb{R}^{m \times n}$ von n unabhängigen Beobachtungen von y S die verzerrte Stichprobenkovarianzmatrix von Y. Dann heißt die Funktion

$$F_Y : \Theta \to \mathbb{R}, \theta \mapsto F_Y(\theta) := \ln|\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln|S| - m \tag{92}$$

die *Diskrepanzfunktion der Faktorenanalyse*. Weiterhin heißt ein Wert $\hat{\theta} \in \Theta$ mit

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} F_Y(\theta), \tag{93}$$

also eine Minimumstelle von F_Y, ein *Diskrepanzfunktionsbasierter Faktorenanalyseparameterschätzer*.

Bemerkungen

- Rosseel & Vidal (2025) geben folgende intuitive Intpretation der Diskrepanzfunktion:
- Wenn $\Sigma_\theta = S$, dann gilt

$$\begin{aligned} F_Y(\theta) &= \ln|\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln|S| - m \\ &= \ln|\Sigma_\theta| + \text{tr}\left(\Sigma_\theta\Sigma_\theta^{-1}\right) - \ln|\Sigma_\theta| - m \\ &= \text{tr}\left(I_m\right) - m \\ &= m - m \\ &= 0 \end{aligned} \tag{94}$$

- Ein θ für das $F_Y(\theta)$ minimal wird approximiert via Σ_θ also S bestmöglich.
- Die Minimierung von $F_Y(\theta)$ ist intuitiv analog zur OLS Minimierung im ALM.
- $F_Y(\theta)$ und OLS Minimierung sind äquivalent zur Likelihood-Maximierung.

Theorem (Diskrepanzfunktionsbasierte und Maximum-Likelihood-Schätzer)

Jede Minimumstelle der Diskrepanzfunktion des Normalverteilungsmodells der Faktorenanalyse maximiert die Log-Likelihood-Funktion des Normalverteilungsmodells der Faktorenanalyse.

Bemerkungen

- Minimierung der Diskrepanzfunktion und Maximierung der Likelihoodfunktion liefern die gleichen Schätzer.
- Diskrepanzfunktionsminimierung ist in der Faktoranalyse etwas populärer als Likelihood-Maximizierung.
- Letztlich geht dies auf die Verwendung des χ^2-Kriteriums in Lawley & Maxwell (1962) zurück
- Ziel war aber immer die Maximum-Likelihood-Schätzung (vgl. Lawley (1940), Jöreskog (1967)).

Beweis

Wir halten zunächst fest, dass mit den von θ unabhängigen Konstanten

$$a := \frac{n}{2} \text{ und } b := -\frac{n}{2} \ln |S| - \frac{n}{2} m - \frac{nm}{2} \ln(2\pi) \tag{95}$$

gilt, dass

$$\begin{aligned} a\left(-F_Y(\theta)\right) + b &= -\frac{n}{2}\left(\ln|\Sigma_\theta| + \mathsf{tr}\left(S\Sigma_\theta^{-1}\right) - \ln|S| - m\right) - \frac{n}{2}\ln|S| - \frac{n}{2}m - \frac{nm}{2}\ln(2\pi) \\ &= -\frac{n}{2}\ln|\Sigma_\theta| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_\theta^{-1}\right) + \frac{n}{2}\ln|S| + \frac{n}{2}m - \frac{n}{2}\ln|S| - \frac{n}{2}m - \frac{nm}{2}\ln(2\pi) \\ &= -\frac{n}{2}\ln|\Sigma_\theta| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_\theta^{-1}\right) - \frac{nm}{2}\ln(2\pi) \\ &= \ell_Y(\theta). \end{aligned} \tag{96}$$

Die Log-Likelihood-Funktion ist also eine linear-affine und damit insbesondere monotone Transformation von F_Y. Da monotone Transformationen Extremstellen unverändert lassen und ein negatives Vorzeichen eine Minimumstelle in eine Maximumstelle transformiert, ergibt sich das Theorem direkt.

□

Vergleich der Faktorladungsmatrixschätzer im Anwendungsbeispiel

Hauptkomponentenschätzung

	1	2
1	+1.18	+0.61
2	+0.97	−0.24
3	+0.98	−0.28
4	+0.95	−0.29
5	+0.35	+1.00
6	+0.33	+0.95
7	+0.40	+0.93
8	+0.32	+0.96
9	+0.33	+0.96
10	+1.17	+0.60
11	+0.94	−0.23
12	+0.95	−0.32
13	+0.98	−0.24
14	+0.30	+0.93
15	+0.97	−0.31
16	+0.94	−0.29
17	+0.97	−0.27
18	+0.93	−0.28
19	+0.97	−0.25
20	+0.95	−0.23
21	+0.92	−0.29

Hauptachsenfaktorisierung

	1	2
1	+0.81	+0.50
2	+0.92	−0.15
3	+0.92	−0.18
4	+0.89	−0.19
5	+0.24	+0.89
6	+0.23	+0.89
7	+0.30	+0.88
8	+0.22	+0.88
9	+0.23	+0.89
10	+0.82	+0.50
11	+0.90	−0.14
12	+0.89	−0.22
13	+0.91	−0.15
14	+0.21	+0.89
15	+0.91	−0.21
16	+0.89	−0.19
17	+0.90	−0.17
18	+0.89	−0.19
19	+0.92	−0.16
20	+0.91	−0.15
21	+0.90	−0.20

Maximum−Likelihood−Schätzung

	1	2
1	+0.84	+0.44
2	+0.90	−0.22
3	+0.90	−0.25
4	+0.88	−0.26
5	+0.31	+0.87
6	+0.30	+0.87
7	+0.37	+0.86
8	+0.28	+0.86
9	+0.30	+0.87
10	+0.85	+0.44
11	+0.89	−0.21
12	+0.87	−0.29
13	+0.90	−0.22
14	+0.28	+0.87
15	+0.89	−0.28
16	+0.87	−0.26
17	+0.88	−0.24
18	+0.87	−0.26
19	+0.91	−0.23
20	+0.90	−0.21
21	+0.88	−0.27

Die Durchführung einer Maximum-Likelihood Schätzung betrachten wir im Kontext des χ^2-Tests genauer.

Datenanalyseansätze in der angewandten Faktorenanalyseliteratur

Explorativer Ansatz ("Exploratorische Faktorenanalyse")

- Annahme eines normalverteilungsfreien Faktorenanalysemodells
- Hauptkomponentenschätzung bzw. Hauptachsenfaktorisierung
- Verteilungsfreie Faktorenanzahlwahl durch z.B. Scree-Test oder Kaiser-Guttman Kriterium
- Faktorenrotation zur besseren Zuordnung von Items zu Faktoren (siehe Modellinterpretation)
- Verbale Interpretation der faktorbasierten Itemcluster (vgl. Groh & Ostwald (2025))

Konfirmatorischer Ansatz ("Konfirmatorische Faktorenanalyse")

- Annahme des Normalverteilungsmodell der Faktorenanalyse
- Nach Möglichkeit Sicherstellung der Modellidentifizierbarkeit
- Maximum-Likelihood-Schätzung bzw. Diskrepanzfunktionsminimierung
- Parameterinferenz und Modellvergleiche zur Identifikation einer plausiblen Datenerklärung
- Enge Verbindung zur Strukturgleichungsmodellierung (vgl. Rosseel & Vidal (2025), Bollen (1989))

Bemerkungen

- In der Anwendung wird allerdings auch oft bunt gemischt und recht wild möglichst viele Maße berichtet.
- Es ist dann meist nicht ganz klar, warum dieses oder jenes nun genau gemacht wurde.
- Typisches Beispiele in diesem Sinne sind Keller et al. (2008) oder Stefana & Hill (2025).

Grundlagen

Faktorenanalysemodelle

Modellschätzung

Modellevaluation

Modellinterpretation

Selbstkontrollfragen

Überblick

Die zentrale Frage in der Datenmodellierung mit einem Faktorenanalysemodell ist, wie viele Faktoren zur Modellierung eines Datensatzes herangezogen werden sollen, welches k also gewählt werden soll.

Prinzipiell sind hier wie immer in der Datenmodellierung zwei Herangehensweisen möglich.

- Man hat keine theoretischen Ideen und hofft, die Daten selbst geben einem die Antwort und liefern ihre Theorie induktiv mit. Man hofft auf annahmenfreie Kriterien (die es nicht gibt), die einem das sinnvollste Modell verraten und lässt sich bestenfalls von den so gewonnenen Datenmustern zu einer Theorie inspirieren. Dies entspricht dem, was Psycholog:innen unter dem Begriff der "exploratorischen Faktorenanalyse" fassen.
- Eine theoretische Betrachtung des Forschungsgegenstandes motiviert deduktiv ein bestimmtes Modell, also eine bestimmte Faktoranzahl. Man nutzt den Modellierungsrahmen dann, um dieses Modell vor dem Hintergrund observierter Daten zu evaluieren und mit anderen möglichen theoriebasierten Datenerklärungen zu vergleichen. Ist man in seinen Annahmen dabei sehr explizit, ist dies ist ein sinnvoller intersubjektiver Zugang. Dies entspricht dem, was Psycholog:innen unter dem Begriff der "konfirmatorische Faktorenanalyse" fassen.

In der Praxis der psychologischen Faktorenanalyse sind eine Reihe von Kriterien zur Wahl der Faktorenanzahl populär, weil sie in gängigen Black-Box-Softwarelösungen implementiert sind.

Populäre Kriterien zur Wahl der Faktorenanzahl in der Anwendungsliteratur

Factor retention criterion	*N*	%
Kaiser-Guttman*	169	55.6
Scree-test*	141	46.4
Parallel Analysis*	128	42.1
Theory/Interpretability*	108	35.5
AIC	1	0.3
BIC	8	2.6
χ^2-test	1	0.3
Comparison Data	1	0.3
Eigenvalue > .70	1	0.3
Variance accounted for	24	7.9
Variables per factor ≥ 3	16	5.3
MAP test	59	19.4
RMSEA	3	1.0
SRMR	1	0.3
Standard Error Scree	16	5.3
Very Simple Structure	1	0.3
Not Reported	13	4.3

Percentages do not add up to 1, because multiple criteria were used among the majority of EFAs. * Criteria, that were used stand-alone to determine the number of factors in at least one study

Aus Goretzko et al. (2021), siehe auch Fabrigar et al. (1999)

Wir betrachten in der Folge davon

- den *Scree-Test* vor dem Hintegrund einer Stichprobenvarianzzerlegung,
- das sehr heuristische *Kaiser-Guttman-Kriterium*,
- den χ^2-*Test* im Sinne eines Log-Likelihood-Ratio-basierten Modellvergleichs.

Definition (Stichprobenvarianzzerlegung der Faktorenanalyse)

$Y \in \mathbb{R}^{m \times n}$ sei ein Datensatz von n unabhängigen Beobachtungen eines Faktorenanalysemodells, $C \in \mathbb{R}^{m \times m}$ sei seine Stichprobenkovarianzmatrix und $\hat{L} \in \mathbb{R}^{m \times k}$ und $\hat{\Psi} \in \mathbb{R}^{m \times m}$ seien die Hauptkomponentenschätzer $k$ter Ordnung. Dann werden die Summe der Diagonalemente von C als *Gesamtstichprobenvarianz*, die Summe der Diagonalelemente von $\hat{L}\hat{L}^T$ als *faktorbasierte Stichprobenvarianz* und die Summe der Diagonalelemente von $\hat{\Psi}$ als *fehlerbasierte Stichprobenvarianz* bezeichnet.

Bemerkungen

- Der Scree-Test basiert auf der Varianzzerlegung

$$\text{Gesamtstichprobenvarianz} = \text{Faktorbasierte Stichprobenvarianz} + \text{Fehlerbasierte Stichprobenvarianz} \quad (97)$$

- Der Scree-"Test" ist ein von Cattell (1966) vorgeschlagenes grafisches Verfahren.
- Im Kern geht es dem Scree-Test um Modellakkuratheits- und Modellkomplexitätsabwägung.
- Ziel ist es, möglichst viel Gesamtstichprobenvarianz mit möglichst wenigen Faktoren zu erklären.
- Für die weitere Ausarbeitung zu nicht-grafischen Verfahren vgl. Raîche et al. (2013).
- *Scree* heißt auf Deutsch Schutthalde, die Wahl des Begriffes wird im Folgenden klar.

Scree (Skriða, Talus, Schutthalde)

© Kevin Lenz

Theorem (Stichprobenvarianzzerlegung der Faktorenanalyse)

Basierend auf n unabhängigen Beobachtungen eines Faktorenanalysemodells seien

- $C = (c_{ij})_{1 \leq i,j \leq m} \in \mathbb{R}^{m \times m}$ die Stichprobenkovarianzmatrix,
- $\hat{L} = (\hat{l}_{ij})_{1 \leq i \leq m, 1 \leq j \leq k} \in \mathbb{R}^{m \times k}$ der Hauptkomponentenschätzer $k$ter Ordnung von L,
- $\hat{\Psi} = \text{diag}(\hat{\psi}_i, \ldots, \hat{\psi}_m) \in \mathbb{R}^{m \times m}$ der Hauptkomponentenschätzer $k$ter Ordnung von Ψ,

sowie

- $G := \sum_{i=1}^{m} c_{ii}$ die Gesamtstichprobenvarianz,
- $F := \sum_{i=1}^{m} \sum_{j=1}^{k} \hat{l}_{ij}^2$ die faktorbasierte Stichprobenvarianz,
- $R := \sum_{i=1}^{m} \hat{\psi}_i$ die fehlerbasierte Stichprobenvarianz.

Dann gilt

$$G = F + R. \tag{98}$$

Außerdem gilt mit den Eigenwerten $\lambda_1, \ldots, \lambda_k$ von C, dass

$$F = \sum_{j=1}^{k} \lambda_j, \text{ wobei } \lambda_j = \sum_{i=1}^{m} \hat{l}_{ij}^2 \tag{99}$$

für $j = 1, \ldots, k$ der Anteil des $j$ten Faktors an F ist.

Beweis

Wir erinnern zunächst daran, dass die Diagonalelemente von $\hat{L}\hat{L}^T$ durch

$$\sum_{j=1}^{k} \hat{l}_{ij}^2 \tag{100}$$

gegeben sind, wovon man sich durch Betrachtung der Einträge von $\hat{L}\hat{L}^T$ überzeugt:

$$\begin{aligned} \hat{L}\hat{L}^T &= \begin{pmatrix} \hat{l}_{11} & \cdots & \hat{l}_{1k} \\ \hat{l}_{21} & \cdots & \hat{l}_{2k} \\ \vdots & \ddots & \vdots \\ \hat{l}_{m1} & \cdots & \hat{l}_{mk} \end{pmatrix} \begin{pmatrix} \hat{l}_{11} & \cdots & \hat{l}_{m1} \\ \hat{l}_{12} & \cdots & \hat{l}_{m2} \\ \vdots & \ddots & \vdots \\ \hat{l}_{1k} & \cdots & \hat{l}_{mk} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^{k} \hat{l}_{1j}^2 & \sum_{j=1}^{k} \hat{l}_{1j}\hat{l}_{2j} & \cdots & \sum_{j=1}^{k} \hat{l}_{1j}\hat{l}_{mj} \\ \sum_{j=1}^{k} \hat{l}_{2j}\hat{l}_{1j} & \sum_{j=1}^{k} \hat{l}_{2j}^2 & \cdots & \sum_{j=1}^{k} \hat{l}_{2j}\hat{l}_{mj} \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{j=1}^{k} \hat{l}_{mj}\hat{l}_{1j} & \sum_{j=1}^{k} \hat{l}_{mj}\hat{l}_{2j} & \cdots & \sum_{j=1}^{k} \hat{l}_{mj}^2 \end{pmatrix} \end{aligned} \tag{101}$$

Beweis (fortgeführt)

Die Identität von G und $F + R$ folgt dann direkt aus der Identität der Diagonalelemente von C, $\hat{L}\hat{L}^T$ und $\hat{\Psi}$, die im Rahmen der Hauptkomponentenschätzung mithilfe von

$$\hat{\psi}_i := c_{ii} - \sum_{j=1}^{k} \hat{l}_{ij}^2 \text{ für } i = 1, \ldots, m \tag{102}$$

konstruiert wird. Um als nächstes

$$F = \sum_{j=1}^{k} \lambda_j \tag{103}$$

zu zeigen halten zunächst fest, dass mit der Definition des Hauptkomponentenschätzer $\hat{L}$ die Summe der quadrierten Einträge in der $j$ten Spalte von $\hat{L}$ gleich der Summe der quadrierten Einträge in der $j$ten Spalte von $Q_k \Lambda_k^{1/2}$ ist.

Beweis (fortgeführt)

Dies mag man sich zum Beispiel für $m = 5$ und $k = 2$ verdeutlichen:

$$\hat{L} = Q_k \Lambda_k^{1/2} \Leftrightarrow \begin{pmatrix} \hat{l}_{11} & \hat{l}_{12} \\ \hat{l}_{21} & \hat{l}_{22} \\ \hat{l}_{31} & \hat{l}_{32} \\ \hat{l}_{41} & \hat{l}_{42} \\ \hat{l}_{51} & \hat{l}_{52} \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \\ q_{31} & q_{32} \\ q_{41} & q_{42} \\ q_{51} & q_{52} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} q_{11} & \sqrt{\lambda_2} q_{12} \\ \sqrt{\lambda_1} q_{21} & \sqrt{\lambda_2} q_{22} \\ \sqrt{\lambda_1} q_{31} & \sqrt{\lambda_2} q_{32} \\ \sqrt{\lambda_1} q_{41} & \sqrt{\lambda_2} q_{42} \\ \sqrt{\lambda_1} q_{51} & \sqrt{\lambda_2} q_{52} \end{pmatrix} \tag{104}$$

Weiterhin halten wir fest, dass, wenn q_j für $j = 1, .., k$ die $j$te Spalte von Q_k bezeichnet aufgrund der Orthonormalität von Q folgt, dass

$$q_j^T q_j = \sum_{i=1}^{m} q_{ij}^2 = 1. \tag{105}$$

Dann ergibt sich für die Summe der Diagonalelemente von $\hat{L}\hat{L}^T$ aber

$$F = \sum_{i=1}^{m} \sum_{j=1}^{k} \hat{l}_{ij}^2 = \sum_{j=1}^{k} \sum_{i=1}^{m} \hat{l}_{ij}^2 = \sum_{j=1}^{k} \sum_{i=1}^{m} \left(\sqrt{\lambda}_j q_{ij} \right)^2 = \sum_{j=1}^{k} \lambda_j \sum_{i=1}^{m} q_{ij}^2 = \sum_{j=1}^{k} \lambda_j \tag{106}$$

Die Tatsache, dass der $j$te Eigenwert λ_j von C dabei der Anteil der durch den $j$ten Faktor erklärten Gesamtstichprobenvarianz ist ergibt sich dabei durch die Einsicht, dass der Beitrag des $j$ten Faktors in der $j$ten Spalte von $\hat{L}$ enkodiert ist und obige Gleichungskette impliziert, dass

$$\sum_{i=1}^{m} \hat{l}_{ij}^2 = \lambda_j \text{ für } j = 1, \ldots, k. \tag{107}$$

□

Anwendungsbeispiel

```
YT        = read.csv("3-Daten/3-fa-keller.csv")                 # Y^T \in \mathbb{R}^{n x m}
Y         = as.matrix(t(YT))                                    # Y \in \mathbb{R}^{m x n}
m         = nrow(Y)                                             # Datendimension
n         = ncol(Y)                                             # Datenpunktanzahl
k         = 2                                                   # Faktoranzahl
I_n       = diag(n)                                             # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                        # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))          # Stichprobenkovarianzmatrix
EA        = eigen(C)                                            # Eigenanalyse von R
lambda_k  = EA$values[1:k]                                      # k größte Eigenwerte von R
Q_k       = EA$vectors[,1:k]                                    # k zugehörige Eigenvektoren von R
L_hat     = Q_k %*% diag(sqrt(lambda_k))                        # Faktorladungsmatrixschätzer
Psi_hat   = diag(diag(C) - diag(L_hat %*% t(L_hat)))            # Fehlerkovarianzmatrixschätzer
GG        = sum(diag(C))                                        # Gesamtstichprobenvarianz
FF        = sum(diag(L_hat %*% t(L_hat)))                       # Faktorenbasierte Stichprobenvarianz
RR        = sum(diag(Psi_hat))                                  # Fehlerbasierte  Stichprobenvarianz
FF_lambda = sum(lambda_k)                                       # Summe der Eigenwerte \lambda_1,...,\lambda_k
```

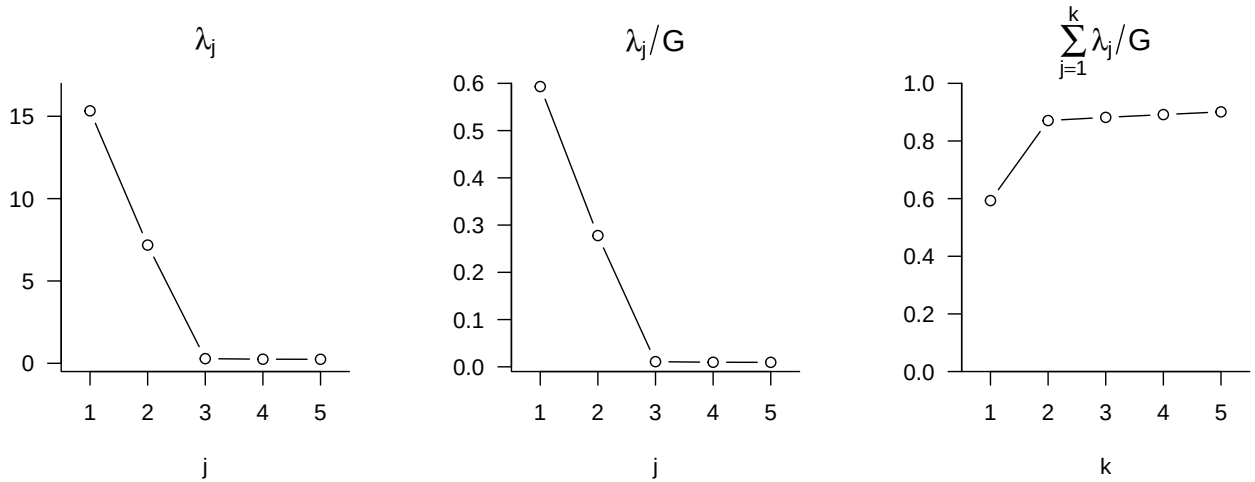
```
G   = 25.84
F   = 22.51
R   = 3.33
F+B = 25.84

Faktorbasierte Stichprobenvarianz F        = 22.51
Summe der Eigenwerte lambda_1,...,lambda_k = 22.51
```

Anwendungsbeispiel

```
# FA mit Hauptkomponentenschätzung für k = 5
YT        = read.csv("3-Daten/3-fa-keller.csv")                  # Y^T \in \mathbb{R}^{n x m}
Y         = as.matrix(t(YT))                                   # Y   \in \mathbb{R}^{m x n}
m         = nrow(Y)                                            # Datendimension
n         = ncol(Y)                                            # Datenpunktanzahl
k         = 5                                                  # Faktoranzahl
I_n       = diag(n)                                            # Einheitsmatrix I_n
J_n       = matrix(rep(1,n^2), nrow = n)                       # 1_{nn}
C         = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))         # Stichprobenkovarianzmatrix
EA        = eigen(C)                                           # Eigenanalyse von C
lambda_k  = EA$values[1:k]                                     # k größte Eigenwerte von C
Q_k       = EA$vectors[,1:k]                                   # k zugehörige Eigenvektoren von C
L_hat     = Q_k %*% diag(sqrt(lambda_k))                       # Faktorladungsmatrixschätzer
Psi_hat   = diag(diag(C) - diag(L_hat %*% t(L_hat)))           # Fehlerkovarianzmatrixschätzer
G         = sum(diag(C))                                       # Gesamtstichprobenvarianz
```

Anwendungsbeispiel

- Man visualisiert $\lambda_1, ...\lambda_k$ als eigentlichen *Scree-Plot*.
- Man betrachtet $\lambda_1/G, ..., \lambda_k/G$ als relative Maße mit Bezug zur Gesamtstichprobenvarianz.
- Man betrachtet $\sum_{i=1}^{j} \lambda_i/G$ für $j = 1, ..., k$ als Maß der kumulativ erklärten Gesamtstichprobenvarianz.



Im Sinne des Scree-Tests liest man hier ab:

- Mit $k = 1$ können 56 % der Gesamtstichprobenvarianz erklärt werden.
- Mit $k = 2$ können 87 % der Gesamtstichprobenvarianz erklärt werden.
- Mit $k = 3$ können 88 % der Gesamtstichprobenvarianz erklärt werden können.

Im Sinne des Scree-Test schließt man hier:

- Der relative Mehrwert durch Hinzunahme des dritten Faktors ist mit 1% gering.
- Im Sinne der Modellparsimonität scheint die Wahl von $k = 2$ Faktoren sinnvoll.

Kaiser-Guttman-Kriterium

Das heuristische Kaiser-Guttman-Kriterium besagt:

"Wähle auf Grundlage Z-transfomierter Daten alle Faktoren mit Eigenwert > 1".

Dieses Vorgehen kann wie folgt begründet werden

- Z-Transformation der Daten resultiert in $\hat{\mathbb{V}}(y_i) = 1 = c_{ii}$ für $i = 1, ..., m$.
- Für die Gesamtstichprobenvarianz gilt dann $G = \sum_{i=1}^{m} c_{ii} = \sum_{i=1}^{m} 1 = m$.
- Eigenvektoren mit $\lambda_i > 1$ erklären dann mehr als die durchschnittliche Varianz pro Item

Kaiser (1960) hat mit obigem Kriterium die Arbeit von Guttman (1954) zusammengefasst.

Die Invalidität des Kriteriums wird z.B. in Zwick & Velicer (1986) und Cliff (1988) diskutiert

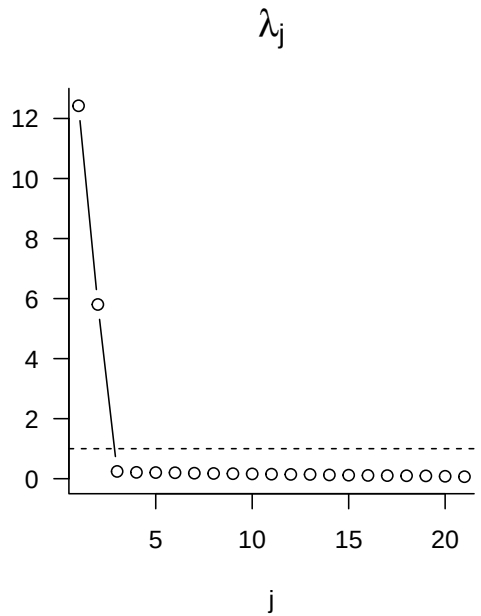
Braeken & Van Assen (2017) haben die Theorie zu einem empirischen Kaiser-Kriterium erweitert.

Anwendungsbeispiel

```
YT       = read.csv("3-Daten/3-fa-keller.csv")          # Y^T \in \mathbb{R}^{n x m}
Y        = as.matrix(t(scale(YT)))                      # Z-transfomierte Y in \mathbb{R}^{m x n}
m        = nrow(Y)                                      # Datendimension
n        = ncol(Y)                                      # Datenpunktanzahl
k        = 5                                            # Faktoranzahl
I_n      = diag(n)                                      # Einheitsmatrix I_n
J_n      = matrix(rep(1,n^2), nrow = n)                 # 1_{nn}
C        = (1/(n-1))*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))   # Stichprobenkovarianzmatrix
EA       = eigen(C)                                     # Eigenanalyse von C
Lambda   = EA$values                                    # Eigenwerte von C
```

Anwendungsbeispiel

- Man visualisiert $\lambda_1, ...\lambda_m$.



Im Sinne des Kaiser-Guttmann-Kriteriums liest man hier ab:

- Es gilt $\lambda_1 > 1, \lambda_2 > 1$ und $\lambda_j < 1$ für $j = 3, ..., m$.

Im Sinne des Kaiser-Guttmann-Kriteriums schließt man hier also:

- Man wählt $k = 2$ Faktoren.
- Im vorliegenden Fall entspricht dies dem Ergebnis des Scree-Tests.

Motivation des χ^2-Tests

- $Y := (y_1, ..., y_n)$ sei ein Datensatz von n unabhängigen Beobachtungen von m-dimensionalen Zufallsvektoren mit $\bar{y} = 0_m$ und vec$(A, B, ...)$ sei die konkatenierte Vektorisierung der unikalen Werte der Matrizen $A, B, ...$ sowie vec$^{-1}(A, B, ...)$ ihre Umkehrung. Grundlage des sogenannten χ^2-Tests der Faktorenanalyse ist dann der Vergleich folgender zwei Modelle M1 und M2.
- Ein Normalverteilungsmodell der Faktorenanalyse, das die Ordnungsrelation erfüllt und identifizierbar ist,

$$y \sim N(0, \Sigma_\theta) \text{ mit } \Sigma_\theta = L\Phi L^T + \Psi, \theta = \text{vec}(L, \Phi, \Psi) \in \Theta, \Theta \subset \mathbb{R}^p \text{ und } p \leq m(m+1)/2. \quad \text{(M1)}$$

- Ein multivariates Normalverteilungsmodell mit beliebigem Kovarianzmatrixparameter (M2),

$$y \sim N(0, \Sigma_\gamma) \text{ mit } \Sigma_\gamma = \text{vec}^{-1}(\gamma), \gamma \in \Gamma \text{ und } \Gamma \subset \mathbb{R}^{m(m+1)/2}. \quad \text{(M2)}$$

- Das dem χ^2-Test zugrundeliegende Kriterium für diesen Modellvergleich ist das Log-Likelihood-Ratio-Kriterium

$$\Lambda_Y := \ln\left(\frac{\max_{\theta\in\Theta}\prod_{i=1}^n N(y_i; 0_m, \Sigma_\theta)}{\max_{\gamma\in\Gamma}\prod_{i=1}^n N(y_i; 0_m, \Sigma_\gamma)}\right) \quad (108)$$

  $\Lambda_Y \in \mathbb{R}$ setzt also die maximierten Wahrscheinlichkeitsdichten von Y unter (M1) und (M2) ins Verhältnis. Große Werte von Λ_Y bedeuten, dass Y unter (M1) eine größere Wahrscheinlichkeitsdichte besitzt als unter (M2). Dies wird allgemein als Evidenz dafür verstanden, dass Y eher von (M1) als von (M2) generiert wurden. Dabei ist Λ_Y letztlich die zentrale Motivation für die funktionale Form der Diskrepanzfunktion (vgl. Lawley (1940)). Das Vorgehen ist hier also ähnlich dem eines varianzanalytischen F-Tests.

Theorem (Log-Likelihood-Ratio-Kriterium und Diskrepanzfunktion)

Für einen Datensatz $Y := (y_1, ..., y_n)$ mit $\bar{y} = 0_m$ von n unabhängigen Beobachtungen des Zufallsvektors y des Normalverteilungsmodells der Faktorenanalyse sei das *Log-Likelihood-Ratio-Kriterium* gegeben durch

$$\Lambda_Y := \ln\left(\frac{\max_{\theta\in\Theta}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta)}{\max_{\gamma\in\Gamma}\prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma)}\right), \tag{109}$$

wobei

$$\Sigma_\theta = L\Phi L^T + \Psi, \theta = \text{uvec}(L, \Phi, \Psi) \in \Theta, \Theta \subset \mathbb{R}^p \text{ und } p \leq m(m+1)/2 \tag{110}$$

und

$$\Sigma_\gamma = \text{uvec}^{-1}(\gamma), \gamma \in \Gamma \text{ und } \Gamma \subset \mathbb{R}^{m(m+1)/2} \tag{111}$$

seien. Weiterhin sei für die verzerrte Stichprobenkovarianzmatrix S von Y

$$F_Y(\theta) := \ln|\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln|S| - m \tag{112}$$

die Diskrepanzfunktion der CFA und $\hat{\theta}$ eine Minimumstelle von F_Y. Dann gilt

$$-2\Lambda_Y = nF_Y(\hat{\theta}) \tag{113}$$

Bemerkungen

- Das Theorem geht u.a. zurück auf D. N. Lawley (1943), Rippe (1953), Jöreskog (1967)
- Das Log-Likelihood-Kriterium ist umgekehrt proportional zur Diskrepanzfunktion an einer Minimumstelle $\hat{\theta}$.
- Verwendet man die unverzerrte Stichprobenkovarianzmatrix für S, so ergibt sich $-2\Lambda_Y = (n-1)F_Y(\hat{\theta})$.

Beweis

Wir halten zunächst fest, dass

$$\max_{\theta \in \Theta} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta) = \prod_{i=1}^{n} N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) \tag{114}$$

weil eine Minimumstelle $\hat{\theta}$ von F_Y wie oben gesehen die Log-Likelihood-Funktion und damit auch die Likelihood-Funktion des Normalverteilungsmodells der Faktorenanalyse maximiert. Weiterhin halten wir fest, dass

$$\max_{\gamma \in \Gamma} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma) = \prod_{i=1}^{n} N(y_i; 0_m, S) \tag{115}$$

weil die verzerrte Stichprobenkovarianz der Maximum-Likelihood-Schätzer des Kovarianzmatrixparameters einer multivariaten Normalverteilung ist. Die Logarithmuseigenschaften ergeben dann

$$\Lambda_Y = \ln\left(\frac{\prod_{i=1}^{n} N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right)}{\prod_{i=1}^{n} N(y_i; 0_m, S)}\right) = \sum_{i=1}^{n} \ln N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) - \sum_{i=1}^{n} \ln N(y_i; 0_m, S) \tag{116}$$

Beweis (fortgeführt)

Substitution der funktionalen Form der Log-Likelihood-Funktion des Normalverteilungsmodells der Faktorenanalyse und der funktionalen Form der WDF der multivariaten Normalverteilung ergibt dann

$$\begin{aligned}
\Lambda_Y =& \sum_{i=1}^{n} \ln N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) - \sum_{i=1}^{n} \ln N(y_i; 0_m, S) \\
=& -\frac{mn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma_{\hat{\theta}}| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) - \frac{n}{2}\bar{y}^T\Sigma_{\hat{\theta}}^{-1}\bar{y} \\
&+ \frac{mn}{2}\ln(2\pi) + \frac{n}{2}\ln|S| + \frac{n}{2}\mathsf{tr}\left(SS^{-1}\right) + \frac{n}{2}\bar{y}^T S^{-1}\bar{y} \\
=& -\frac{n}{2}\ln|\Sigma_{\hat{\theta}}| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) - \frac{n}{2}0_m^T\Sigma_{\hat{\theta}}^{-1}0_m \\
&+ \frac{n}{2}\ln|S| + \frac{n}{2}\mathsf{tr}\left(SS^{-1}\right) + \frac{n}{2}0_m^T S^{-1}0_m \\
=& -\frac{n}{2}\ln|\Sigma_{\hat{\theta}}| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) + \frac{n}{2}\ln|S| + \frac{mn}{2}
\end{aligned} \tag{117}$$

Multiplikation mit -2 ergibt schließlich

$$-2\Lambda_Y = n\ln|\Sigma_{\hat{\theta}}| + n\,\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) - n\ln|S| - mn = nF_Y(\hat{\theta}). \tag{118}$$

□

Theorem (X^2-Statistik)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse. Dann gilt für die X^2-Statistik

$$\mathrm{X}^2 := nF_Y(\hat{\theta}) \tag{119}$$

bei Gültigkeit des restringierten Modells

$$y \sim N(0, \Sigma_\theta) \text{ mit } \Sigma_\theta = L\Phi L^T + \Psi \text{ und } \theta \in \mathbb{R}^{n_\theta} \tag{120}$$

dass

$$\mathrm{X}^2 \sim \chi^2 \left(\frac{m(m+1)}{2} - n_\theta \right) \text{ für } n \to \infty \tag{121}$$

wobei n_θ die Anzahl an unikalen skalaren Parametern des Modells bezeichnet.

Bemerkungen

- Wir verzichten auf einen Beweis dieses auf Jöreskog (1967) zurückgehenden Theorems.
- Jöreskog (1967) begründet die asymptotische VErteilung von X^2 mit Wilks (1938).
- Wilks (1938) zeigt, dass die asympotische χ^2-Verteilung allen Log-Likelihood-Ratio-Kriterien gemein ist.
- Wir validieren die asymptotische Verteilung für $n = 266$ untenstehend in einem Anwendungsbeispiel.

Modellevaluation | χ^2-Test

Interpretation von $\mathrm{X}^2 = nF_Y(\hat{\theta})$

- Hohe Werte von $F_Y(\hat{\theta})$ bzw. X^2 sprechen für eine hohe *Diskrepanz* von Σ_θ und S
- Geringe Werte von $F_Y(\hat{\theta})$ bzw. X^2 sprechen für eine hohe *Ähnlichkeit* von Σ_θ und S

Es gilt für den χ^2-Test der Faktorenanalyse also:

$\Rightarrow$ Geringer X^2-Wert mit hohem assoziierten P-Wert (p $>$ 0.05)

- Der beobachtete X^2-Wert ist unter dem Modell nicht ungewöhnlich groß.
- Die Modell Datenkovarianzmatrix ist mit der Stichprobenkovarianzmatrix vereinbar.
- Das Modell weist nach dem χ^2-Test eine *akzeptable Passung* auf.
- Das Modell wird *nicht* verworfen.

$\Rightarrow$ Hoher X^2 Wert mit niedrigem assoziierten P-Wert (p $<$ 0.05)

- Der beobachtete X^2-Wert ist zu groß, um allein durch Stichprobenfehler erklärbar zu sein.
- Das Modell reproduziert die Stichprobenkovarianzstruktur nicht ausreichend.
- Das Modell weist nach dem χ^2-Test *keine* akzeptable Passung auf.
- Das Modell wird verworfen.

Anwendungsbeispiel

Als Beispiel für die Durchführung eines χ^2-Test folgen wir Rosseel & Vidal (2025). Wir fokussieren dabei auf die ersten 9 Items des BDI-II, da für mehr als 9 Items `lavaan` Methoden nutzt, die über das hier Gesagte hinausgehen. Es gelte also $m = 9$ und $C \in \mathbb{R}^{9\times 9}$. Die obigen Scree-Test- und Kaiser-Guttman-Kriterium-Analysen legen für den (gesamten) BDI-II zwei Faktoren, also $k = 2$, nahe. Ein naives Faktorenanalyse-Normalverteilungsmodell für # diesen Datensatz hat also die Form

$$y = Lf + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \\ l_{41} & l_{42} & l_{43} \\ l_{51} & l_{52} & l_{53} \\ l_{61} & l_{62} & l_{63} \\ l_{71} & l_{72} & l_{73} \\ l_{81} & l_{82} & l_{83} \\ l_{91} & l_{92} & l_{93} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix} \tag{122}$$

mit

$$f \sim N(0_2, \Phi) \text{ und } \varepsilon \sim N(0_9, \Psi) \tag{123}$$

und

$$\Phi := \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \text{ und } \Psi := \text{diag}\left(\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8, \psi_9\right). \tag{124}$$

Anwendungsbeispiel

Für die Anzahl unikaler skalarer Statistiken bei $m = 9$, gilt dass $n_c = 9(9+1)/2 = 45$.

Ein mögliches restringiertes Modell für die ersten 9 Items des BDI-II ist

$$y = Lf + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ l_{21} & 0 \\ l_{31} & 0 \\ l_{41} & 0 \\ 0 & l_{52} \\ 0 & l_{62} \\ 0 & l_{72} \\ 0 & l_{82} \\ 0 & l_{92} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix} \tag{125}$$

mit

$$f \sim N(0_2, \Phi) \text{ und } \varepsilon \sim N(0_9, \Psi) \tag{126}$$

und

$$\Phi := \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \text{ und } \Psi := \text{diag}\left(\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8, \psi_9\right). \tag{127}$$

Für die Anzahl der als unbekannt vorausgesetzten Parameter in diesem Modell ergibt sich also

$$n_\theta = 8 + 3 + 9 = 20 < 45 = n_c \tag{128}$$

und die Ordnungsbedingung ist erfüllt.

Anwendungsbeispiel

Um die Parameter des betrachteten Modells nach dem ML-Ansatzes durch Minimierung der Diskrepanzfunktion

$$F_Y : \Theta \to \mathbb{R}, \theta \mapsto F_Y(\theta) := n \ln |\Sigma_\theta| + n\text{tr}\left(S\Sigma_\theta^{-1}\right) - n \ln |S| - nm \tag{129}$$

zu schätzen, bestimmen wir zunächst n, m und den ML-Schätzer S seiner Kovarianzmatrix.

```
Y               = read.csv("3-Daten/3-fa-keller-small.csv") # Gesamter BDI-II Datensatz
colnames(Y)     = paste('I', 1:9, sep = "")                 # Vereinfachte Itemvariablennamen
m               = ncol(Y)                                   # Datendimension (9)
n               = nrow(Y)                                   # Stichprobenumfang (266)
S               = (n-1)/n*cov(Y)                            # ML (verzerrte) Stichprobenkovarianzmatrix
round(S,3)                                                  # Ausgabe der Stichprobenkovarianzmatrix
```

```
      I1     I2     I3     I4     I5     I6     I7     I8     I9
I1 1.754  0.897  0.846  0.813  0.784  0.783  0.800  0.860  0.830
I2 0.897  1.204  0.998  1.013 -0.073 -0.041 -0.036  0.007 -0.046
I3 0.846  0.998  1.126  0.976 -0.106 -0.071 -0.043 -0.030 -0.071
I4 0.813  1.013  0.976  1.161 -0.147 -0.112 -0.110 -0.068 -0.120
I5 0.784 -0.073 -0.106 -0.147  1.094  0.930  0.917  0.925  0.963
I6 0.783 -0.041 -0.071 -0.112  0.930  1.077  0.884  0.911  0.949
I7 0.800 -0.036 -0.043 -0.110  0.917  0.884  1.024  0.879  0.898
I8 0.860  0.007 -0.030 -0.068  0.925  0.911  0.879  1.098  0.948
I9 0.830 -0.046 -0.071 -0.120  0.963  0.949  0.898  0.948  1.133
```

Anwendungsbeispiel

Um Σ_θ als Funktion von θ darzustellen, definieren wir zwei **R** Funktionen

```
mat = function(theta){

  # Diese Funktion wandelt einen Faktorenanalysemodellparameter \theta in
  # entsprechende Modellmatrizen L, \Phi und \Psi um.
  # ------------------------------------------------------------------------
  # Faktorladungsmatrix
  L   = matrix(c(1          , 1,
                 theta[1]   , 0,
                 theta[2]   , 0,
                 theta[3]   , 0,
                 0          , theta[4],
                 0          , theta[5],
                 0          , theta[6],
                 0          , theta[7],
                 0          , theta[8]), nrow = 9, ncol = 2, byrow = TRUE)

  # Faktorkovarianzmatrix
  Phi = matrix(c(theta[9] , theta[10],
                 theta[10], theta[11]), nrow = 2, ncol = 2, byrow = TRUE)

  # Fehlerkovarianzmatrix
  Psi =  diag(theta[12:20])

  # Listenausggabe
  return(list(L= L, Phi = Phi, Psi = Psi))}

Sigma = function(theta){

  # Funktion zur Konstruktion der marginalen Kovarianzmatrix der Parameter für
  # das betrachtete Anwendungsbeispiel.
  # ------------------------------------------------------------------------
  # Matrizenform des Parametevektors
  M       = mat(theta)

  # Marginale Datenkovarianzmatrix
  Sigma   = M$L %*% M$Phi %*% t(M$L) + M$Psi}
```

Anwendungsbeispiel

Die Funktion `Sigma` liefert nun für beliebiges θ die entsprechende Matrix Σ_θ

```
theta_0     = c(1,1,1,1,1,1,1,1,                    # Startwert für \theta (L-Anteil)
                1,0,1,                              # Startwert für \theta (Phi-Anteil)
                0,.2,.2,.2,.2,.2,.2,.2,.2)          # Startwert für \theta (Psi-Anteil)
Sigma_0     = Sigma(theta_0)                        # Startwert für \Sigma_theta
round(Sigma_0,2)                                    # Ausgabe
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,]    2  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0
 [2,]    1  1.2  1.0  1.0  0.0  0.0  0.0  0.0  0.0
 [3,]    1  1.0  1.2  1.0  0.0  0.0  0.0  0.0  0.0
 [4,]    1  1.0  1.0  1.2  0.0  0.0  0.0  0.0  0.0
 [5,]    1  0.0  0.0  0.0  1.2  1.0  1.0  1.0  1.0
 [6,]    1  0.0  0.0  0.0  1.0  1.2  1.0  1.0  1.0
 [7,]    1  0.0  0.0  0.0  1.0  1.0  1.2  1.0  1.0
 [8,]    1  0.0  0.0  0.0  1.0  1.0  1.0  1.2  1.0
 [9,]    1  0.0  0.0  0.0  1.0  1.0  1.0  1.0  1.2
```

Für das gewählte θ sind S und Σ_θ noch sehr unähnlich.

Anwendungsbeispiel

Um F_Y auszuwerten, definieren wir eine weitere **R** Funktion.

```
F_Y = function(theta){
  # Funktion zur Auswertung der Diskrepanzfunktion für das Faktorenanalyse-
  # Normalverteilungsmodell basierend auf einem gegebenen Wert des Parameter-
  # vektors und mit der ML-Stichprobenkovarianz als verfügbare globale Variable.
  # ------------------------------------------------------------------------
  Sigma_theta = Sigma(theta)                          # \Sigma_\theta
  EA          = eigen(Sigma_theta)                    # PD Check
  if (any(EA$values) < 1e-4){                         # \lambda \approx 0
    return(1e9)}                                      # invalider F_Y Wert
  F_Y         = (log(det(Sigma_theta))                # ln |\Sigma_\theta|
                 + sum(diag(S%*% solve(Sigma_theta))) # tr(S\Sigma_theta^{-1})
                 - log(det(S))                        # ln |S|
                 - m)                                 # m
}
```

Anwendungsbeispiel

Zur Minimierung von F_Y bezüglich θ nutzen wir `nlminb` wie `lavaan`

```
min   = nlminb(start = theta_0, objective = F_Y)          # nichtlineare numerische Optimierung
min                                                       # Ausgabe
```

```
$par
 [1]  1.09946668  1.06317081  1.07808539  1.06795044  1.04244804  1.01439980
 [7]  1.04635850  1.07948968  0.85343226 -0.06462686  0.83492900  0.19533026
[13]  0.17203421  0.16181542  0.16891752  0.14187456  0.16992165  0.16527449
[19]  0.18418787  0.15962719

$objective
[1] 0.09684989

$convergence
[1] 0

$iterations
[1] 44

$evaluations
function gradient
      55     1169

$message
[1] "relative convergence (4)"
```

Anwendungsbeispiel

Umwandlung der Minimalstelle $\hat{\theta}$ in $\hat{L}, \hat{\Phi}$ und $\hat{\Psi}$ ergibt dann

```
L_hat =

       [,1]  [,2]
 [1,] 1.000 1.000
 [2,] 1.099 0.000
 [3,] 1.063 0.000
 [4,] 1.078 0.000
 [5,] 0.000 1.068
 [6,] 0.000 1.042
 [7,] 0.000 1.014
 [8,] 0.000 1.046
 [9,] 0.000 1.079

Phi_hat =

       [,1]   [,2]
[1,]  0.853 -0.065
[2,] -0.065  0.835

Psi_hat =

       [,1]  [,2]  [,3]  [,4]  [,5] [,6]  [,7]  [,8] [,9]
 [1,] 0.195 0.000 0.000 0.000 0.000 0.00 0.000 0.000 0.00
 [2,] 0.000 0.172 0.000 0.000 0.000 0.00 0.000 0.000 0.00
 [3,] 0.000 0.000 0.162 0.000 0.000 0.00 0.000 0.000 0.00
 [4,] 0.000 0.000 0.000 0.169 0.000 0.00 0.000 0.000 0.00
 [5,] 0.000 0.000 0.000 0.000 0.142 0.00 0.000 0.000 0.00
 [6,] 0.000 0.000 0.000 0.000 0.000 0.17 0.000 0.000 0.00
 [7,] 0.000 0.000 0.000 0.000 0.000 0.00 0.165 0.000 0.00
 [8,] 0.000 0.000 0.000 0.000 0.000 0.00 0.000 0.184 0.00
 [9,] 0.000 0.000 0.000 0.000 0.000 0.00 0.000 0.000 0.16
```

Anwendungsbeispiel

Äquivalent ist folgender Black-Box-Ansatz mit dem **R** Paket `lavaan` möglich, aber intransparent

```
library(lavaan)                                          # Lavaan Paket
Y              = read.csv("3-Daten/3-fa-keller-small.csv") # Erste 9 BDi-II Items Datensatz
colnames(Y)    = paste('I', 1:9, sep = "")               # Vereinfachte Itemvariablennamen
m              = ncol(Y)                                 # Datendimension
n              = nrow(Y)                                 # Stichprobenumfang
fam            = 'f_1 =~  I1 + I2 + I3 + I4
                  f_2 =~  I1 + I5 + I6 + I7 + I8 + I9'   # Spezifikation des restringierten Modells
fam            = cfa(fam, data = Y)                      # CFA Modellformulierung und -schätzung
theta_hat_l    = lavInspect(fam, what = "est")           # Inspektion der geschätzten Parameter
L_hat_l        = theta_hat_l$lambda                      # Geschätzte Faktorladungsmatrix
Phi_hat_l      = theta_hat_l$psi                         # Geschätzte Fehlerkovarianzmatrix
Psi_hat_l      = theta_hat_l$theta                       # Geschätzte Faktorkovarianzmatrix
```

Anwendungsbeispiel

Umwandlung der Minimalstelle $\hat{\theta}_l$ in $\hat{L}_l, \hat{\Phi}_l$ und $\hat{\Psi}_l$ ergibt dann

```
L_hat_l =

     f_1   f_2
I1 1.000 1.000
I2 1.099 0.000
I3 1.063 0.000
I4 1.078 0.000
I5 0.000 1.068
I6 0.000 1.042
I7 0.000 1.014
I8 0.000 1.046
I9 0.000 1.079

Phi_hat_l =

       f_1    f_2
f_1  0.853
f_2 -0.065  0.835

Psi_hat_l =

      I1    I2    I3    I4    I5    I6    I7    I8    I9
I1 0.195
I2 0.000 0.172
I3 0.000 0.000 0.162
I4 0.000 0.000 0.000 0.169
I5 0.000 0.000 0.000 0.000 0.142
I6 0.000 0.000 0.000 0.000 0.000 0.170
I7 0.000 0.000 0.000 0.000 0.000 0.000 0.165
I8 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.184
I9 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.160
```

Anwendungsbeispiel

Schließlich bestimmen wir basierend auf den erhaltenen Schätzern die X^2-Statistik

```
# Eigene Schätzung
Sgm_hat      = Sigma(theta_hat)                  # Sigma_\hat{\theta}
Sgm_hat_i    = solve(Sgm_hat)                    # Sigma_\hat{\theta}^{-1}
F_Y_hat      = (log(det(Sgm_hat))                # ln |Sigma_\hat{\theta}|
                + sum(diag(S%*%Sgm_hat_i))       # tr(S Sigma_\hat{\theta}^{-1} )
                - log(det(S))                    # ln |S|
                - m)                             # m
Chi_square   = n*F_Y_hat                         # Chi_square TestStatistik
df           = (m*(m+1))/2 - length(theta_0)     # Freiheitsgradparameter
p            = 1 - pchisq(Chi_square, df = df)   # p-Wert
```

Wir erhalten

```
Test statistic         25.762

Degrees of freedom     25

P-value (Chi-square)   0.42
```

Anwendungsbeispiel

Schließlich bestimmen wir basierend auf den erhaltenen Schätzern die X^2-Statistik

```
show(fam) # Ausgabe der Lavaan Ergebnisse
```

```
lavaan 0.6-21 ended normally after 38 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        20

  Number of observations                           266

Model Test User Model:

  Test statistic                                25.762
  Degrees of freedom                                25
  P-value (Chi-square)                           0.420
```

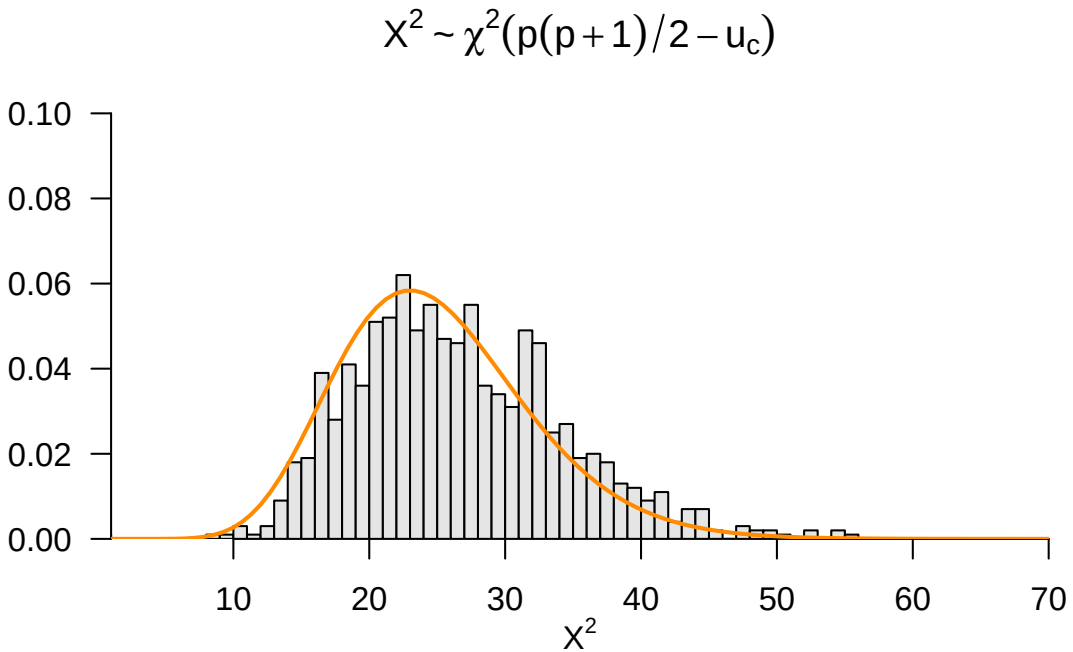
Anwendungsbeispiel

Abschließend validieren wir im Anwendungsbeispiel die Verteilung der X^2-Statistik

```
library(MASS)                                  # Multivariates Normalverteilungspaket
nsim           = 1000                          # Anzahl an Datensatzrealisierungen
k              = 2                             # Dimension des Faktorvektors
m              = 9                             # Dimension des Datenvektors
n              = 266                           # Beobachtungsanzahl
L              = L[1:9,]                       # W.a.u. Faktorladungsmatrix
Phi            = diag(k)                       # W.a.u. Faktorkovarianzmatrix
Psi            = .1*diag(m)                    # W.a.u. Fehlerkovarianmatrix
mu             = 2                             # Offset
df             = (m*(m+1))/2 - length(theta_0) # Freiheitsgradparameter
Chisq          = rep(NaN, nsim)                # Chi-Square Array
for(s in 1:nsim){                              # Simulationsiterationen
  Y            = matrix(rep(NaN,m*n), nrow = m) # Simulierte beobachtete Datenmatrix
  for(i in 1:n){                               # Simulationsiterationen
    f          = mvrnorm(1,rep(0,k), Phi)      # Faktorvektorrealisierung
    eps        = mvrnorm(1,rep(0,m), Psi)      # Fehlervektorrealisierung
    Y[,i]      = mu + L %*% f + eps}           # Realisierung des beobachtbaren Datenzufallsvektors
  Y[Y<0]       = 0                             # Zensur
  Y[Y>5]       = 5                             # Zensur
  Y            = round(Y, digits = 0)          # Rundung
  Y            = t(Y)                          # Standardformat n x m
  colnames(Y)  = paste('I', 1:9, sep = "")     # Vereinfachte Itemvariablennamen
  S            = (n-1)/n*cov(Y)                # ML (verzerrte) Stichprobenkovarianzmatrix
  min          = nlminb(start      = theta_0,  # Minimierungs Startwert
                        objective  = F_Y)      # Zielfunktion
  theta_hat    = min$par                       # Parameterschätzer
  Sgm_hat      = Sigma(theta_hat)              # Sigma_\hat{\theta}
  Sgm_hat_i    = solve(Sgm_hat)                # Sigma_\hat{\theta}^{-1}
  F_Y_hat      = (log(det(Sgm_hat))            # ln |Sigma_\hat{\theta}|
                  + sum(diag(S%*%Sgm_hat_i))   # tr(S Sigma_\hat{\theta}^{-1} )
                  - log(det(S))                # ln |S|
                  - m)                         # m
  Chisq[s]     = n*F_Y_hat                     # Chi_square Teststatistik
}
```

Anwendungsbeispiel

Verteilung von X^2 bei Gültigkeit des restringierten Modells



- Es tritt eine leichte Verzerrung aufgrund der Datenzensierung auf.

Grundlagen

Faktorenanalysemodelle

Modellschätzung

Modellevaluation

Modellinterpretation

Selbstkontrollfragen

Überblick

Mechanistisch-deduktiver Wissenschaftsansatz

- Sinnvollerweise überlegt man sich die Bedeutung eines Modells bei seiner Formulierung.
- Modellparameterschätzer entsprechen dann per se bedeutungsvollen Datenzusammenfassungen.
- Evaluationsansätze erlauben es dann, die Modellplausibilität im Vergleich zu anderen Modellen zu beurteilen.
- Im Kontext der Faktorenanalyse entspricht dies dem "konfirmatorischen" Ansatz.

Semiformal-induktiver Wissenschaftsansatz

- Man hofft darauf, dass empirische Daten einem die Theorie ihrer Generation selbst mitteilen.
- Informell ist dies natürlich auch die Grundlage des mechanistisch-deduktiven Ansatzes.
- Semiformal wird daraus ein Vorgehen zum Finden einer Modellinterpretationsinspiration.
- Im Kontext der Faktorenanalyse entspricht dies dem "exploratorischen" Ansatz.

Faktorenrotation als semiformaler-induktiver Ansatz der Faktorenanalyse

- *Faktorenrotation* bezeichnet verschiedene Verfahren zur Modifikation von Faktorladungsmatrixschätzern.
- Ziel ist es, bei fester Faktorenanzahl einen "interpretierbaren" Faktorladungsmatrixschätzer zu erhalten.
- "Interpretierbar" meint hier die einfache Identifikation faktorspezifischer Itemcluster.
- Um die Ziele der Faktorenrotation zu erreichen, könnte man auch einfach eine Clusteranalyse durchführen.

Motivation

- Grundlegendes Problem des (normalverteilungsfreien) Faktorenanalysemodells

$$y = Lf + \varepsilon \tag{130}$$

  ist die Nichtidentifizierbarkeit von Faktorladungsmatrix L und Faktorvektor f.

- Nach Theorem Nichtidentifizierbarkeit und Kovarianzinvarianz gilt für

$$\tilde{y} = \tilde{L}\tilde{f} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{f} := Q^T f \tag{131}$$

  mit einer beliebigen orthogonale Matrix $Q \in \mathbb{R}^{k \times k}$, dass

$$y = \tilde{y} \text{ und } \mathbb{C}(\tilde{y}) = \mathbb{C}(y). \tag{132}$$

- Sowohl Lf als auch $\tilde{L}\tilde{f}$ können also zur Erklärung von y genutzt werden.
- Hat man einen Schätzer $\hat{L}$ vorliegen, so ist also auch jedes $\hat{L}Q$ ein möglicher Schätzer.
- Es stellt sich also die Frage, welche Form ein "guter" Faktorladungsmatrixschätzer haben soll.
- L. L. Thurstone (1947) beantwortet diese Frage im Sinne einer "Einfachstruktur" (simple structure).

Einfachstruktur nach L. L. Thurstone (1947) (S.335)

If we eliminate the tests of complexity r, the remaining tests should contribute to the location of the reference frame. We shall describe five useful criteria by which the r reference vectors can be determined. These are as follows:

1) Each row of the oblique factor matrix V should have at least one zero.

2) For each column p of the factor matrix V there should be a distinct set of r linearly independent tests whose factor loadings v_{jp} are zero.

3) For every pair of columns of V there should be several tests whose entries v_{jp} vanish in one column but not in the other.

4) For every pair of columns of V, a large proportion of the tests should have zero entries in both columns. This applies to factor problems with four or five or more common factors.

5) For every pair of columns there should preferably be only a small number of tests with non-vanishing entries in both columns.

When these conditions are satisfied, the plot of each pair of columns shows (1) a large concentration of points in two radial streaks, (2) a large number of points at or near the origin, and (3) only a small number of points off the two radial streaks. For a configuration of r dimensions there are $\frac{1}{2}r(r-1)$ diagrams. When all of them satisfy the three characteristics, we say that the structure is "compelling," and we have good assurance that the simple structure is unique. In the last analysis it is the appearance of the diagrams that determines, more than any other criterion, which of the hyperplanes of the simple structure are convincing and whether the whole configuration is to be accepted as stable and ready for interpretation.*

Moderne Interpretation der Einfachstruktur nach L. L. Thurstone (1947)

- Per Datenkomponente sind Faktorladungen gewünscht, die eine möglichst eindeutige Faktorzuordnung erlauben.
- Statt z.B. untenstehendem $\hat{L}$ wäre also eine geschätzte Faktorladungsmatrix wie $\hat{L}^*$ gewünscht:

$$\hat{L} = \begin{pmatrix} 3.81 & 0.16 \\ 2.54 & 1.35 \\ 3.89 & 0.11 \\ 0.53 & -1.22 \\ 1.66 & -2.32 \end{pmatrix} \quad \Rightarrow \quad \hat{L}^* = \begin{pmatrix} 1.00 & 0.00 \\ 1.00 & 0.00 \\ 1.00 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.00 \end{pmatrix} \tag{133}$$

- Erste Mathematisierungsversuche finden sich bei Carroll (1953), Kaiser (1958), Crawford & Ferguson (1970).
- Heutzutage nennt man Matrizen mit vielen Nullen und einigen wenigen Einsen *sparse matrices*.
- Rohe & Zeng (2023) schreiben entsprechend *"Varimax estimates a sparse basis"*
- Eindeutige Zuordnungen von Items zu Faktoren induzieren dann Cluster von Items.
- Die Items eines Clusters laden jeweils hoch auf einen Faktor.
- Durch Inspektion der Clusteritems hofft man auf Inspiration zu einer inhaltlichen Faktorinterpretation.
- Man hätte also auch einfach eine Clusteranalyse durchführen können.
- Es werden keine Faktoren rotiert, sondern, wenn überhaupt, werden Koordinatensysteme "rotiert".
- Es geht um das Muster der Projektionen der itemspezifischen Faktorladungen auf (nicht) orthogonale Basen.

Faktorenrotationsansätze sind ein weites Feld. Wir fokussieren im Folgenden auf

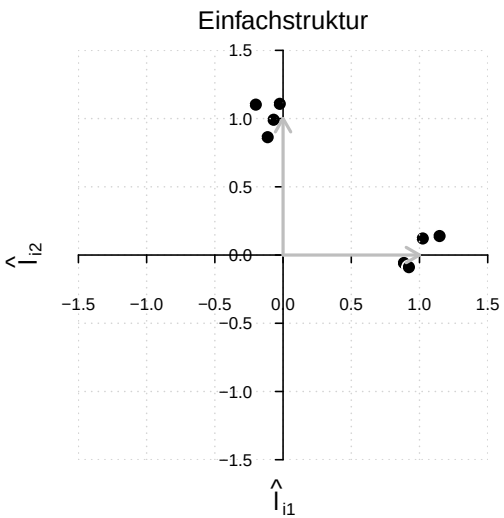
(1) den Begriff einer Vektorraumorthonormalbasis am Beispiel von $\mathbb{R}^2$,
(2) den Begriff der Orthonormalbasisprojektion,
(3) Orthomaxverfahren ("orthogonale Rotationen") zur Auswahl einer Einfachstruktur.

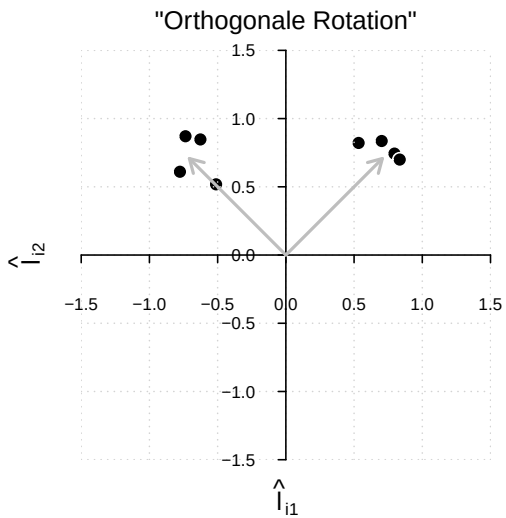
Intuition zu “Orthogonalen und Obliquen Rotationen”

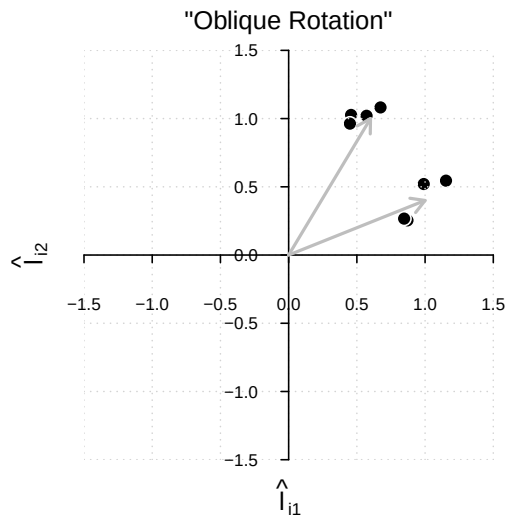
$$\hat{L} = \begin{pmatrix} 0.89 & -0.06 \\ 0.92 & -0.09 \\ 1.15 & 0.14 \\ 1.02 & 0.12 \\ -0.07 & 0.99 \\ -0.11 & 0.86 \\ -0.02 & 1.11 \\ -0.20 & 1.10 \end{pmatrix}$$

$$\hat{L} = \begin{pmatrix} 0.70 & 0.84 \\ 0.80 & 0.74 \\ 0.53 & 0.82 \\ 0.83 & 0.70 \\ -0.63 & 0.85 \\ -0.51 & 0.52 \\ -0.78 & 0.61 \\ -0.74 & 0.87 \end{pmatrix}$$

$$\hat{L} = \begin{pmatrix} 0.87 & 0.25 \\ 0.99 & 0.52 \\ 1.15 & 0.54 \\ 0.85 & 0.27 \\ 0.67 & 1.08 \\ 0.46 & 1.03 \\ 0.45 & 0.96 \\ 0.57 & 1.02 \end{pmatrix}$$







Theorem (Orthonormalbasis)

Eine Menge $B = \{b_1, \ldots, b_k\}$ von k Vektoren in $\mathbb{R}^k$ heißt *Orthonormalbasis von* $\mathbb{R}^k$, wenn mit dem Skalarprodukt

$$\langle x, y \rangle := x^T y \tag{134}$$

für $x, y \in \mathbb{R}^k$ folgende, zum Teil äquivalente, Bedingungen gelten:

(1) Die Vektoren $b_1, \ldots, b_k \in \mathbb{R}^k$ bilden eine Basis von $\mathbb{R}^k$, das heißt, sie sind linear unabhängig und jeder Vektor $x \in \mathbb{R}^k$ lässt sich als Linearkombination $x = \sum_{i=1}^{k} c_i b_i$ mit eindeutigen Koordinaten $c_1, \ldots, c_k \in \mathbb{R}$ schreiben.

(2) Die Vektoren $b_1, \ldots, b_k \in \mathbb{R}^k$ sind orthonormal, das heißt $\langle b_i, b_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$ für $1 \leq i, j \leq k$.

(3) Die Koordinaten $c_1, \ldots, c_k \in \mathbb{R}$ eines $x \in \mathbb{R}^k$ bezüglich B ergeben sich durch die Orthogonalprojektionen

$$c_i = \langle x, b_i \rangle = x^T b_i = \text{ für } i = 1, .., k, \tag{135}$$

es gilt also

$$x = \sum_{i=1}^{k} c_i b_i = \sum_{i=1}^{k} \langle x, b_i \rangle b_i = \sum_{i=1}^{k} \left(x^T b_i \right) b_i \tag{136}$$

Bemerkungen

- Bei Bedarf wird empfohlen, sich der Grundlagen der Vektorrechnung zu vergewissern.
- Siehe dazu auch Probabilistische Datenwissenschaft für die Psychologie, Kapitel 8

Beispiel (1)

Ein Beispiel für eine Orthonormalbasis von $\mathbb{R}^2$ ist die *Standardbasis*

$$\mathsf{B}_1 := \{b_1, b_2\} := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \tag{137}$$

Offenbar gelten

$$b_1^T b_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 + 0 = 1, \quad b_2^T b_2 = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 + 1 = 1 \tag{138}$$

und

$$b_1^T b_2 = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 + 0 = 0. \tag{139}$$

Sei nun beispielsweise

$$x := \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} \in \mathbb{R}^2. \tag{140}$$

Die Koordinaten c_1 und c_2 von x bezüglich B_1 ergeben sich durch die Orthogonalprojektionen

$$c_1 = x^T b_1 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{3} \text{ und } c_2 = x^T b_2 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{2}{3} \tag{141}$$

Offenbar lässt sich $x \in \mathbb{R}^2$ also schreiben als

$$x = \sum_{i=1}^{2} c_i b_i = c_1 b_1 + c_2 b_2 = \frac{1}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$$

Beispiel (2)

Ein weiteres Beispiel für eine Orthonormalbasis von $\mathbb{R}^2$ ist

$$\mathsf{B}_2 := \{b_1, b_2\} := \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\}. \tag{142}$$

Auch hier gelten

$$b_1^T b_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2} + \frac{1}{2} = 1, \quad b_2^T b_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2} + \frac{1}{2} = 1 \tag{143}$$

und

$$b_1^T b_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = -\frac{1}{2} + \frac{1}{2} = 0. \tag{144}$$

Sei nun wiederrum

$$x := \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} \in \mathbb{R}^2. \tag{145}$$

Die Koordinaten c_1 und c_2 von x bezüglich B_2 ergeben sich durch die Orthogonalprojektionen

$$c_1 = x^T b_1 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \approx 0.70 \text{ und } c_2 = x^T b_2 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{3\sqrt{2}} \approx 0.23 \tag{146}$$

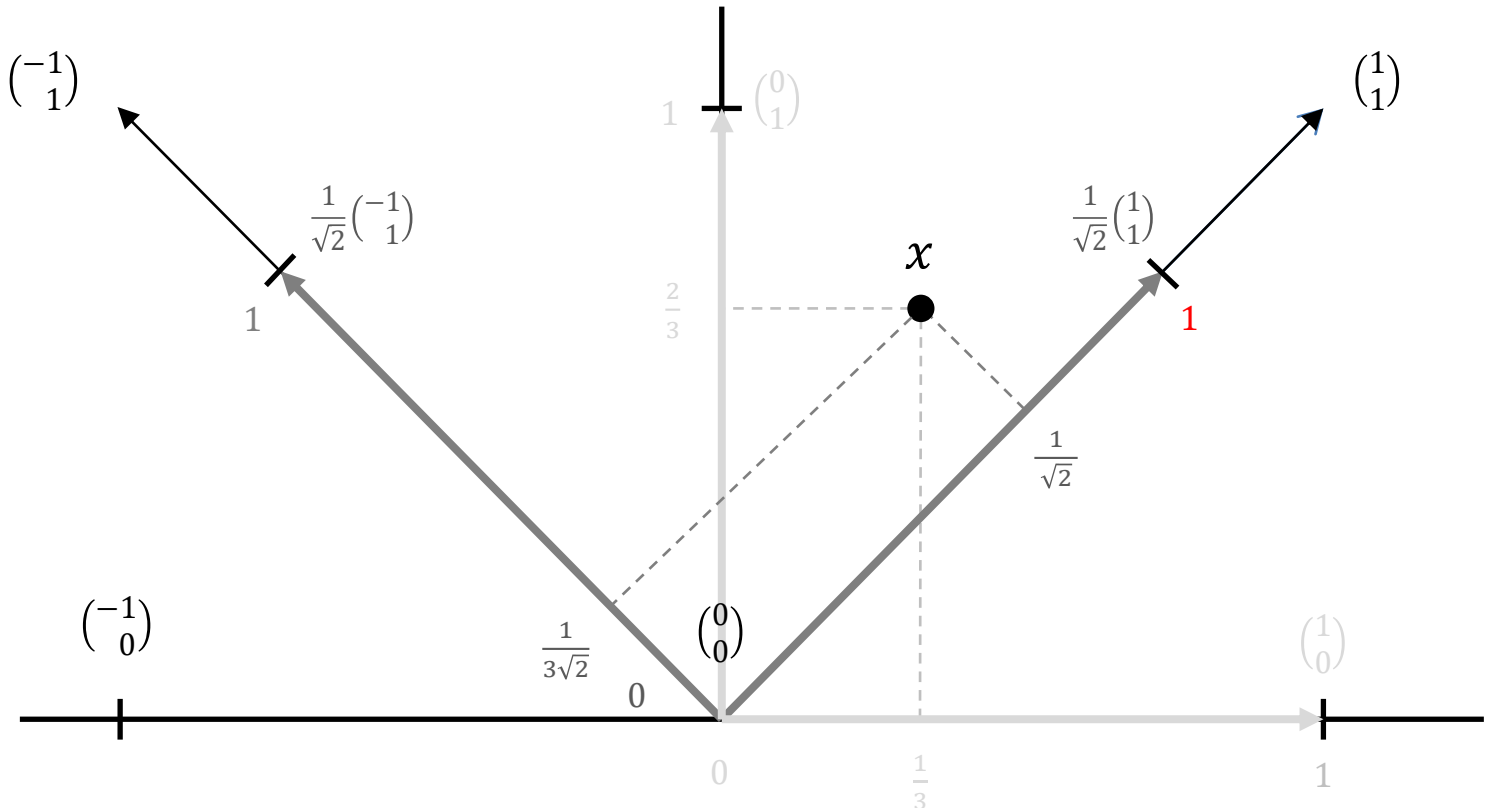
Offenbar lässt sich $x \in \mathbb{R}^2$ also auch schreiben als

$$x = \sum_{i=1}^{2} c_i b_i = c_1 b_1 + c_2 b_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + \frac{1}{3\sqrt{2}} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} + \begin{pmatrix} -\frac{1}{6} \\ \frac{1}{6} \end{pmatrix} = \begin{pmatrix} \frac{3}{6} - \frac{1}{6} \\ \frac{3}{6} + \frac{1}{6} \end{pmatrix} = \begin{pmatrix} \frac{2}{6} \\ \frac{4}{6} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}$$

Visualisierung von Beispiel (1) und Beispiel (2)

$B_1 := \left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\} \Rightarrow x = \sum_{i=1}^{2} c_i, b_i \text{ mit } c_1 = \frac{1}{3}, c_2 = \frac{2}{3}$

$B_2 := \left\{\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}, \begin{pmatrix}-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}\right\} \Rightarrow x = \sum_{i=1}^{2} c_i, b_i \text{ mit } c_1 = \frac{1}{\sqrt{2}}, c_2 = \frac{1}{3\sqrt{2}}$



Motivation und technische Umsetzung von Orthomaxverfahren

Grundidee

- Grundidee ist es, eine Faktorraumbasis zu identifizieren, für die $\hat{L}$ eine Einfachstruktur hat.
- Dazu interpretiert man zunächst die Zeilen von $\hat{L}$ als (transponierte) Vektoren in $\mathbb{R}^k$.
- Man betrachtet also die Faktorladungen jedes Items als Koordinaten bezüglich der Standardbasis von $\mathbb{R}^k$.
- Man möchte einen alternativen Koordinatensatz erhalten, so dass das entsprechende $\hat{L}^*$ Einfachstruktur hat.
- Dazu probiert man verschieden Orthonormalbasen aus und inspiziert die sich ergebenen Formen von $\hat{L}^*$.
- Wenn ein solches $\hat{L}^*$ die bestmögliche Einfachstruktur, stoppt man die Orthonormalbasensuche.

Technische Umsetzung

- Sei $\hat{l}_j^T \in \mathbb{R}^{1\times k}$ eine Zeile von $\hat{L}$, also $\hat{l}_j \in \mathbb{R}^k$ mit Einträgen als Koordinaten bezüglich der Standardbasis
- Sei $M \in \mathbb{R}^{k\times k}$ nun eine Matrix, deren Spalten $m_1, ..., m_k$ eine Orthornormalbasis von $\mathbb{R}^k$ bilden.
- Nach Definition ist M dann eine Orthogonalmatrix, es gilt also $M^T M = I_k$.
- Weiterhin ergeben sich die Koordinaten $c_1, ..., c_k$ von $\hat{l}_j$ bezüglich der von M repräsentierten Basis als

$$\hat{l}_j^{*T} = \begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix} = \begin{pmatrix} \hat{l}_j^T m_1 & \cdots & \hat{l}_j^T m_k \end{pmatrix} = \hat{l}_j^T \begin{pmatrix} m_1 & \cdots & m_k \end{pmatrix} = \hat{l}_j^T M \tag{147}$$

- Für alle Zeilen von $\hat{L}$ simultan, also alle itemspezifischen Faktorladungen ergibt sich $\hat{L}^* = \hat{L}M$.
- Man sucht also ein $M \in \mathbb{R}^{k\times k}$ mit $M^T M = I_k$, so dass $\hat{L}^*$ für festes $\hat{L}$ Einfachstruktur hat.
- Man misst den Grad der Einfachstruktur von $\hat{L}^*$ als Funktion von M dabei mit einer *Orthomaxzielfunktion*

Definition (Orthomaxverfahren)

Es sei $\hat{L} \in \mathbb{R}^{m \times k}$ ein Faktorladungsmatrixschätzer und ϕ eine Funktion, die für eine $M \in \mathbb{R}^{k \times k}$ den Grad der Einfachstruktur von $\hat{L}M$ misst. Dann entspricht das *Orthomaxverfahren* dem restringierten nichtlinearen Optimierungsproblem

$$\max \phi(M) \text{ unter der Nebenbedingung } M^T M = I_k. \tag{148}$$

Bemerkungen

- Die Nebenbedingug $M^T M = I_k$ stellt sicher, dass M eine Orthonormalbasis repräsentiert.
- Es gibt verschiedene Zielfunktionen, die den Einfachstrukturgrad von $\hat{L}M$ mehr oder weniger gut messen.
- Wir listen diese Zielfunktionen im Folgenden, betrachten aber nur die Varimaxvariante genauer.

Definition (Orthomaxzielfunktion)

Für einen Faktorladungsmatrixschätzer $\hat{L} \in \mathbb{R}^{m \times k}$ und eine Matrix $M \in \mathbb{R}^{k \times k}$ mit $M^T M = I_k$ sei

$$\hat{L}_M := \hat{L}M =: \left(\ell_{ij}\right)_{1 \leq i \leq m, 1 \leq j \leq k} \in \mathbb{R}^{m \times k} \tag{149}$$

die *transformierte Faktorladungsmatrix*. Dann ist die *Orthomaxzielfunktion* definiert als

$$\phi : \mathbb{R}^{k \times k} \to \mathbb{R}, M \mapsto \phi(M) = \sum_{j=1}^{k} \left(\sum_{i=1}^{m} \frac{1}{m} \ell_{ij}^4 - \frac{w}{m^2} \left(\sum_{i=1}^{m} \ell_{ij}^2 \right)^2 \right) \tag{150}$$

Weiterhin wird f für

- $w := 0$ *Quartimaxzielfunktion*,
- $w := m$ *Faktorparsimonitätszielfunktion*,
- $w := m/2$ *Equamaxzielfunktion*,
- $w := 1$ *Varimaxzielfunktion*.

genannt.

Bemerkungen

- Die Einträge ℓ_{ij} von $\hat{L}_M$ sind von M abhängig, zur Vereinfachung der Notation bleibt dies implizit.
- Es stellt sich die Frage, inwieweit die funktionalen Formen von ϕ die Sparseness von $\hat{L}_M$ messen.
- Zumindest der Varimaxfall kann einigermaßen sinnvoll begründet werden, wie folgendes Theorem zeigt.

Theorem (Varimaxzielfunktion)

Für $j = 1, \ldots, k$ sei

$$\mathsf{SQL}_j = \sum_{i=1}^{m} \left(\ell_{ij}^2 - \frac{1}{m} \sum_{i=1}^{m} \ell_{ij}^2 \right)^2 \tag{151}$$

die Summe der quadrierten Abweichungen der quadrierten Faktorladungen vom Mittelwert der quadrierten Faktorladungen von Faktor j. Dann gilt, dass die Varimax-Zielfunktion die Summenfunktion

$$\varphi : \mathbb{R}^{k \times k} \to \mathbb{R}, M \mapsto \varphi(M) := \sum_{j=1}^{k} \mathsf{SQL}_j \tag{152}$$

maximiert, also eine Lösung des restringierten nichtlinearen Optimierungsproblem

$$\max \varphi(M) \text{ unter der Nebenbedingung } M^T M = I_k. \tag{153}$$

generiert.

Bemerkungen

- Die Summanden ℓ_{ij} von $\varphi(M)$ sind von M abhängig, zur Vereinfachung der Notation bleibt dies implizit.
- Man mag SQL_j als die empirische Varianz der Ladungen von Faktor j über Items interpretieren.
- Allerdings sind die Faktorladungen natürlich keine Realisierungen von Zufallsvariablen.
- Intuitiv präferiert die Varimaxzielfunktion also Lösungen "hoher Varianz der quadrierten Faktorladungen".

Modellinterpretation

Bemerkungen (fortgesetzt)

- Rohe & Zeng (2023) geben folgendes Argument zur Präferenz sparser Faktorladungsmatrizen bei Varimax
- Es sei $k := 2$ und $m := 1$. Dann hat die Varimaxzielfunktion die Form

$$\phi(M) = \sum_{j=1}^{2} \left(\sum_{i=1}^{1} \frac{1}{1} \ell_{ij}^4 - \frac{1}{1^2} \left(\sum_{i=1}^{1} \ell_{ij}^2 \right)^2 \right) = \sum_{j=1}^{2} \left(\ell_{1j}^4 - \left(\ell_{1j}^2 \right)^2 \right) = \ell_{1j}^4 + \ell_{12}^4 - \left(\ell_{11}^2 + \ell_{12}^2 \right) \tag{154}$$

- $\ell_{1j}^2 + \ell_{12}^2$ enstpricht proportional dem Abstand zu $(0, 0)^T$ und bleibt bezüglich jeder Orthonormalbasis konstant.
- Das Maximierungsproblem hat hier also die Form

$$\max_M \ell_{11}^4 + \ell_{12}^4 - c = \max_M \left(\ell_{11}^2 + \ell_{12}^2 \right)^2 - 2\ell_{11}^2 \ell_{12}^2 - c = \max_M c^2 - c - 2\ell_{11}^2 \ell_{12}^2 \tag{155}$$

- Mit $\ell_{11}^2 \geq 0$ und $\ell_{12}^2 \geq 0$ wird $\phi(M)$ also durch $\ell_{11} = 0$ oder $\ell_{12} = 0$ maximiziert
- Zur Optimierung von $\varphi(M)$ stellt **R** die `varimax()` Funktion zur Verfügung, diese wird z.B. von `psych` genutzt.

Beweis

Für $i = 1, \ldots, m$ und $j = 1, \ldots, k$ sei zur Vereinfachung der Notation

$$s_{ij} := \ell_{ij}^2 \tag{156}$$

die quadrierte Ladung des $j$ten Faktors auf das $i$te Item. Weiterhin sei

$$\bar{s}_j := \frac{1}{m} \sum_{i=1}^{m} s_{ij} \tag{157}$$

der Mittelwert der quadrierten Faktorladungen des $j$ten Faktors über Items, also der Mittelwert der $j$ten *Spalte* der transformierten Faktorladungsmatrix. Für jeden Faktor $j = 1, \ldots, k$ kann dann die Summe der quadrierten Mittelwertsabweichungen der quadrierten Faktorladungen geschrieben werden als Term der Orthomaxfunktion ϕ mit $w := 1$, d.h.

$$\mathsf{SQL}_j = \sum_{i=1}^{m} \left(s_{ij} - \bar{s}_j \right)^2 = \sum_{i=1}^{m} \ell_{ij}^4 - \frac{1}{m} \left(\sum_{i=1}^{m} \ell_{ij}^2 \right)^2 . \tag{158}$$

Dann gilt aber

$$\mathsf{SQL} = \sum_{j=1}^{k} \mathsf{SQL}_j = \sum_{j=1}^{k} \left(\sum_{i=1}^{m} \ell_{ij}^4 - \frac{1}{m} \left(\sum_{i=1}^{m} \ell_{ij}^2 \right)^2 \right) . \tag{159}$$

Beweis

Gleichung (158) ergibt sich dabei schließlich durch

$$
\begin{aligned}
\mathsf{SQL}_j &= \sum_{i=1}^{m} \left(s_{ij} - \bar{s}_j \right)^2 \\
&= \sum_{i=1}^{m} \left(s_{ij}^2 - 2 s_{ij} \bar{s}_j + \bar{s}_j^2 \right) \\
&= \sum_{i=1}^{m} s_{ij}^2 - 2 \bar{s}_j \sum_{i=1}^{m} s_{ij} + \sum_{i=1}^{m} \bar{s}_j^2 \\
&= \sum_{i=1}^{m} s_{ij}^2 - 2 \bar{s}_j m \bar{s}_j + m \bar{s}_j^2 \\
&= \sum_{i=1}^{m} s_{ij}^2 - 2 m \bar{s}_j^2 + m \bar{s}_j^2 \\
&= \sum_{i=1}^{m} s_{ij}^2 - m \bar{s}_j^2 \\
&= \sum_{i=1}^{m} s_{ij}^2 - m \left(\frac{1}{m} \sum_{i=1}^{m} s_{ij} \right)^2 \\
&= \sum_{i=1}^{m} s_{ij}^2 - \frac{m}{m^2} \left(\sum_{i=1}^{m} s_{ij} \right)^2 \\
&= \sum_{i=1}^{m} \ell_{ij}^4 - \frac{1}{m} \left(\sum_{i=1}^{m} \ell_{ij}^2 \right)^2
\end{aligned}
\tag{160}
$$

Der `varimax()` Algorithmus

Wie oben gesehen besteht die Grundidee des Varimaxverfahrens darin, die Varimaxzielfunhktion ϕ bezüglich einer orthogonalen Matrix M zu maximieren. Dies entspricht dem nichtlinearen Optimierungsproblem

$$\max \phi(M) \text{ unter der Nebenbedingung } M^T M = I_k. \tag{161}$$

Mittels der Differentialrechnung für Matrizen konnte gezeigt werden (vgl. Horst (1965), Lawley & Maxwell (1971), Neudecker (1969), Magnus & Neudecker (1989)), dass die notwendige Bedingung für ein Maximum der Lagrange-Funktion des Optimierungsproblems (161) in der Gleichung

$$MA_M = B_M \tag{162}$$

ausgedrückt werden kann, wobei $A_M \in \mathbb{R}^{k \times k}$ eine symmetrische und positiv definite Matrix unbekannter Lagrange-Multiplikatoren ist und

$$B_M := \hat{L}^T \left(C_M - \frac{1}{m} \hat{L}_M D_M \right) \tag{163}$$

mit

$$\hat{L}_M := \hat{L}M, C_M := \left(\ell^3\right)_{1 \leq i \leq m, 1 \leq j \leq k} \text{ und } D_M := \text{diag} \left(\sum_{i=1}^{m} \hat{l}_{M_{i1}}^2, \ldots, \sum_{i=1}^{m} \hat{l}_{M_{ik}}^2 \right) \tag{164}$$

ist. Bemerkenswert ist, dass (162) lediglich eine implizite Definition des optimalen M liefert, da beide Seiten der Gleichung von M abhängen. Um dieses Gleichungssystem iterativ nach M zu lösen, geht der in `varimax()` implementierte Ansatz der simultanen Faktoren-Varimax-Lösung nach folgendem Algorithmus vor (laut Rohe & Zeng (2023) entspricht dies einem projizierten Gradientenverfahren (projected gradient ascent)).

Definition (`varimax()` Algorithmus)

Gegeben sei ein Schätzer $\hat{L}$ für die Faktorladungsmatrix eines Faktoranalysemodells. Dann hat der Algorithmus der **R** Funktion `varimax()` die folgende Form:

Initialisierung

(SN) Setze $\hat{L}_N := N\hat{L}$ mit $N :=$ diag $\left(\left(\sum_{j=1}^{k} \hat{l}_{1j}^2 \right)^{-\frac{1}{2}}, ..., \left(\sum_{j=1}^{k} \hat{l}_{mj}^2 \right)^{-\frac{1}{2}} \right)$

(S0) Setze $M^{(0)} := I_k, \delta^{(0)} := 0, \delta^{(1)} := 1$ und definiere $\varepsilon^{(0)}$

Iterationen

Für $i = 1, ..., 1000$ und solange $d^{(i)} \geq d^{(i-1)}(1 + \varepsilon)$

(S1) Setze $\hat{L}_{M^{(i)}} = \hat{L}_N M^{(i-1)}$

(S2) Setze $B_{M^{(i)}} = \hat{L}_N^T \left(C_{M^{(i)}} - \frac{1}{m} \hat{L}_{M^{(i)}} D_{M^{(i)}} \right)$

(S3) Evaluiere $B_{M^{(i)}} = U^{(i)} \Delta^{(i)} V^{(i)^T}$

(S4) Setze $M^{(i)} = U^{(i)} V^{(i)^T}$

(S5) Evaluiere $d^{(i)} :=$ tr$(\Delta^{(i)})$

Finalisierung

(S6) Setze $\hat{L}_{NM} := \hat{L}_N M^{(i)}$ und $\hat{L}_M := N^{-1} \hat{L}_{NM}$

Der `varimax()` Algorithmus

In der Definition des `varimax()` Algorithmus normalisiert der Initialisierungsschritt (SN) jede Zeile der Faktorladungsmatrixschätzung $\hat{L}$ auf die euklidische Einheitslänge durch Multiplikation mit einer $m \times m$ diagonalen Normalisierungsmatrix N, deren Diagonalelemente den reziproken euklidischen Längen der Zeilen von $\hat{L}$ entsprechen. Obwohl nicht unumstritten (vgl. Kaiser (1970), Kaiser & Rice (1974), Rohe & Zeng (2023)), wurde diese zeilenweise Normalisierung bei der Einführung des Varimaxverfahrens durch Kaiser (1958) vorgeschlagen und bildet das Standardverfahren in der **R** Implementierung des Varimaxverfahrens. Die Orthonormalbasismatrix M wird dann zunächst als $k \times k$ Einheitsmatrix initialisiert, die Konvergenzvariablen δ werden auf 0 gesetzt und ein geeigneter Wert für das Konvergenzkriterium ε wird definiert. Die Iterationsschritte (S1) und (S2) evaluieren dann die relevanten Iterandenmatrizen für die rechte Seite von (162), $L_{M^{(i)}}$, $D_{M^{(i)}}$, $C_{M^{(i)}}$, und $B_{M^{(i)}}$, wodurch die rekursive Gleichung

$$M^{(i+1)} A_{M^{(i)}} = B_{M^{(i)}} \tag{165}$$

implizit definiert wird. Diese implizite rekursive Gleichung wird dann für $M^{(i+1)}$ mittels einer Singulärwertzerlegung von $B_{M^{(i)}}$ gelöst, wobei $M^{(i+1)}$ in den Iterationsschritten (S3) und (S4) auf $U^{(i)}V^{(i)}$ gesetzt wird. Hierbei sind $U^{(i)}$ und $V^{(i)}$ die Matrizen der linken bzw. rechten Singulärvektoren von $B_{M^{(i)}}$, und wir begründen diese Lösung untenstehend. Die Konvergenz des Algorithmus wird dann anhand der Spur der Singulärwertmatrix $\Delta^{(i)}$ im Iterationsschritt (S5) bewertet. Schließlich wird nach der Konvergenz die optimierte Orthonormalbasismatrix $M^{(i)}$ auf die normalisierte Faktorladungsmatrixschätzung angewendet und die Normalisierungsskalierung in Schritt (S6) rückgängig gemacht.

Der `varimax()` Algorithmus

Wir wollen schließlich den Iterationsschritt (S4) begründen, welche die notwendige Bedingung für ein Maximum der Lagrange-Funktion des restringierten Optimierungsproblems Gleichung (161) in der Form von (162) iterativ mittels einer Singulärwertzerlegung der Matrix auf der rechten Seite von (162) löst. Dieser Ansatz nutzt die Eigenschaften der beteiligten Matrizen und deren Beziehungen. Zur Vereinfachung der Notation verzichten wir im Folgenden auf die Matrixabhängigkeit von M sowie auf den Iterations-Superskript und betrachten die Lösung von $MA = B$ für bekanntes B und unbekannte (aber positiv-definite und symmetrische) Lagrange-Multiplikatormatrix A. Zunächst stellen wir fest, dass die Prämultiplikation von $MA = B$ mit der Transponierten von B ergibt, dass

$$MA = B \Leftrightarrow B^T MA = B^T B \Leftrightarrow (MA)^T MAB^T B \Leftrightarrow A^T M^T MA = B^T B \Leftrightarrow A^2 = B^T B, \qquad (166)$$

wobei die letzte Äquivalenz aus der Orthogonalität von M und der Symmetrie von A folgt. Basierend auf (166) formulieren wir als Nächstes zwei Ausdrücke für die unbekannte Matrix A^2. Erstens betrachten wir die Singulärwertzerlegung von $B = USV^T$, wobei U die orthogonale Matrix der linken Singulärvektoren, S die Diagonalmatrix der Singulärwerte und V die orthogonale Matrix der rechten Singulärvektoren ist. Wir erhalten dann:

$$A^2 = B^T B = \left(USV^T\right)^T USV^T = VSU^T USV^T = VS^2V^T. \qquad (167)$$

Details des `varimax()` Algorithmus

Da A eine positiv-definite und symmetrische Matrix ist, kann sie als Zerlegung in ihre orthogonale Matrix von Eigenvektoren Q und ihre Diagonalmatrix von Eigenwerten Λ geschrieben werden. Wir haben also auch:

$$A^2 = AA = Q\Lambda Q^T Q\Lambda Q^T = Q\Lambda^2 Q^T \tag{168}$$

und wir erhalten:

$$A^2 = VS^2V^T = Q\Lambda^2 Q^T = A^2. \tag{169}$$

Ohne Beschränkung der Allgemeinheit können wir $V = Q$ annehmen, sodass $S^2 = \Lambda^2$ und damit $S = \Lambda$ bis auf Vorzeichenpermutationen gilt. Wir können nun (162) wie folgt lösen:

$$MA = B \Leftrightarrow M = BA^{-1} \Leftrightarrow M = USV^T Q\Lambda^{-1}Q^T \Leftrightarrow M = USQ^T QS^{-1}V^T \Leftrightarrow M = UV^T, \tag{170}$$

was schließlich den Iterationsschritt (S4) rechtfertigt.

Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

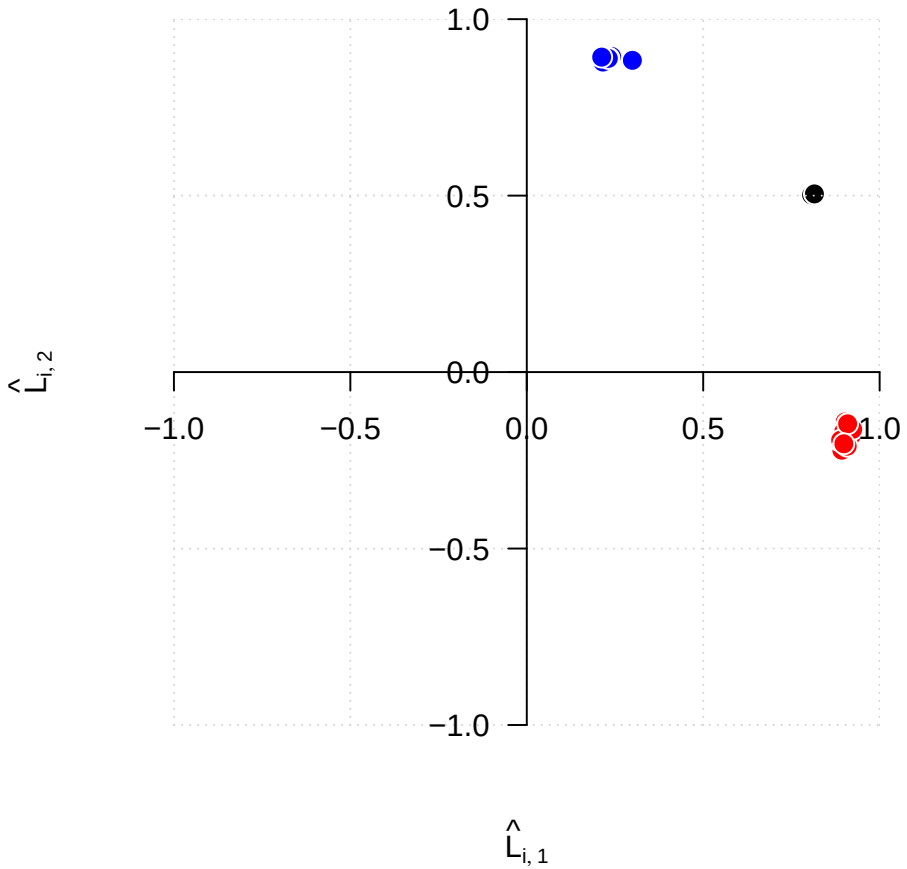
⇒ Iterierte Hauptachsenfaktorisierung

```
Y           = as.matrix(read.csv("3-Daten/3-fa-keller.csv"))          # Y \in \mathbb{R}^{m x n}
R           = cor(Y)                                                  # m x m Korrelationsmatrix
k           = 2                                                       # Faktorenzahl
eps_min     = 0.01                                                    # Konvergenzkriterium
i_max       = 25                                                      # Maximale Anzahl an IPAF Iterationen
i           = 0                                                       # Iterationszähler
R_i         = R                                                       # Initiale Korrelationsmatrix R^{(0)}
h_i         = sum(diag(R_i))                                          # Summe der Kommunalitätsschätzer
eps_i       = h_i                                                     # Fehlerinitialisierung
while(eps_i > eps_min && i <= i_max){                                 # Iterationen
    QLQ_i       = eigen(R_i)                                          # Eigenanalyse
    L_hat_i     = QLQ_i$vectors[,1:k] %*% diag(sqrt(QLQ_i$values[1:k])) # Hauptkomponentenschätzer kter Ordnung
    diag(R_i)   = diag(L_hat_i %*% t(L_hat_i) )                       # Kommunalitätsschätzer Update
    h_ii        = sum(diag(R_i))                                      # Kommunalitätschätzersummen Update
    eps_i       = abs(h_ii - h_i)                                     # Fehlerkonvergenkriterium
    h_i         = h_ii                                                # Kommunalitätsschätzersummen Update
    i           = i + 1}                                              # Iterationszähler Update
D           = diag(sign(colSums(L_hat_i)))                            # Vorzeichenfinalisierung
L_hat       = L_hat_i %*% D                                           # IHAF Faktorladungsmatrixschätzer
```

Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

$\Rightarrow$ Visualisierung der geschätzten Faktorladungen zu jedem Item



Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

$\Rightarrow$ Visualisierung der geschätzten Faktorladungen zu jedem Item

Zeilen von $\hat{L}_2$ mit • ("Affektiver Faktor")

```
 [1] "Pessimismus"                   "Versagensgefühle"
 [3] "VerlustAnFreude"               "Unruhe"
 [5] "Interessenverlust"             "Entschlusslosigkeit"
 [7] "Energieverlust"                "Schlaf"
 [9] "Reizbarkeit"                   "Appetitveränderung"
[11] "Konzentrationsschwierigkeiten" "Ermüdung"
[13] "VerlustSexuellenInteresses"
```

Zeilen von $\hat{L}_2$ mit • ("Kognitiver Faktor")

```
[1] "Schuldgefühle"      "Bestrafungsgefühle" "Selbstablehnung"
[4] "Selbstkritik"       "Suizidgedanken"     "Wertlosigkeit"
```

Zeilen von $\hat{L}_2$ mit •

```
[1] "Traurigkeit" "Weinen"
```

Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

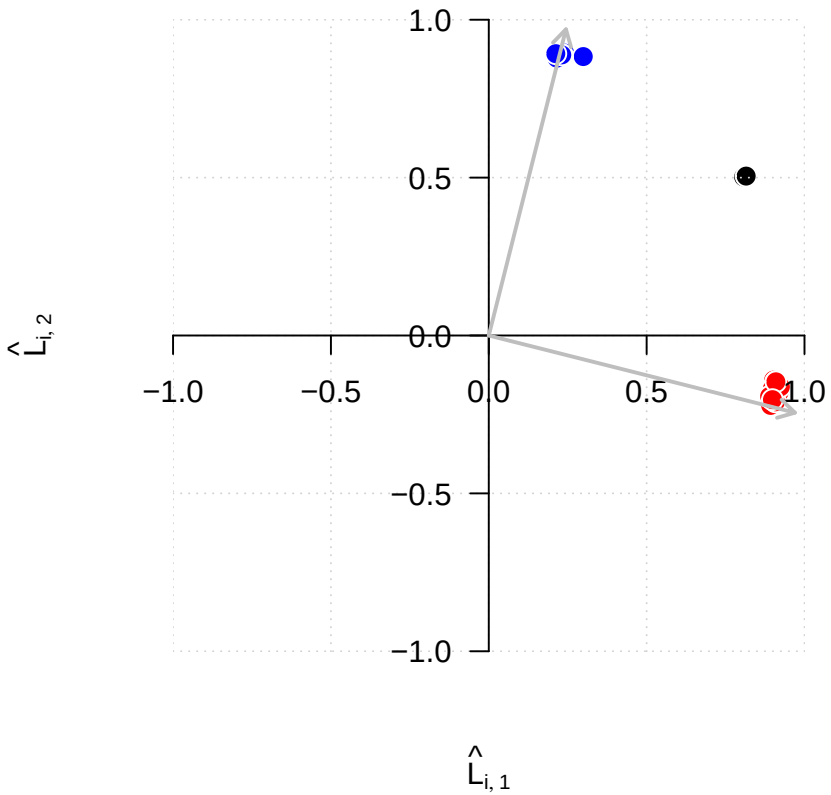
$\Rightarrow$ Evaluation des `varimax()` Algorithmus für $\hat{L}$

```
N             = solve(diag(sqrt(diag(L_hat %*% t(L_hat)))))     # Zeilennormalisierende Matrix
L_hat_N       = N %*% L_hat                                     # zeilennormalisierte Faktorladungsmatrixschätzer
m             = nrow(L_hat_N)                                   # Zeilenanzahl
k             = ncol(L_hat_N)                                   # Spaltenanzahl
Mi            = diag(k)                                         # Orthonormalbasismatrixinitialisierung
d             = 0                                               # Konvergenzvariableninitialisierung
eps           = 1e-05                                           # Konvergenzkriterium (change factor)
for (i in 1L:100L){                                             # Iterationen
    L_hat_Mi  = L_hat_N %*% Mi                                  # \hat{L}_M^{(i)}
    Ci        = L_hat_Mi^3                                      # C^{(i)}
    Di        = diag(drop(rep(1, m) %*% L_hat_Mi^2))            # D^{(i)}
    Bi        = t(L_hat_N) %*% (Ci - (1/m) * L_hat_Mi %*% Di)   # B^{(i)}
    UDVi      = svd(Bi)                                         # Singulärwertzerlegung von B^{(i)}
    Mi        = UDVi$u %*% t(UDVi$v)                            # Orthonormalbasismatrix Update M^{(i)}
    dd        = sum(UDVi$d)                                     # Konvergenzkriterium Update
    if (dd < d * (1 + eps)){break}                              # Konvergenztest
    d         = dd}                                             # Konvergenzvariablen Update
L_hat_NM      = L_hat_N %*% Mi                                  # Transformierter Faktorladungsmatrixschätzer
L_hat_M       = solve(N) %*% L_hat_NM                           # Zeilenormalisierungsaufhebung
```

Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

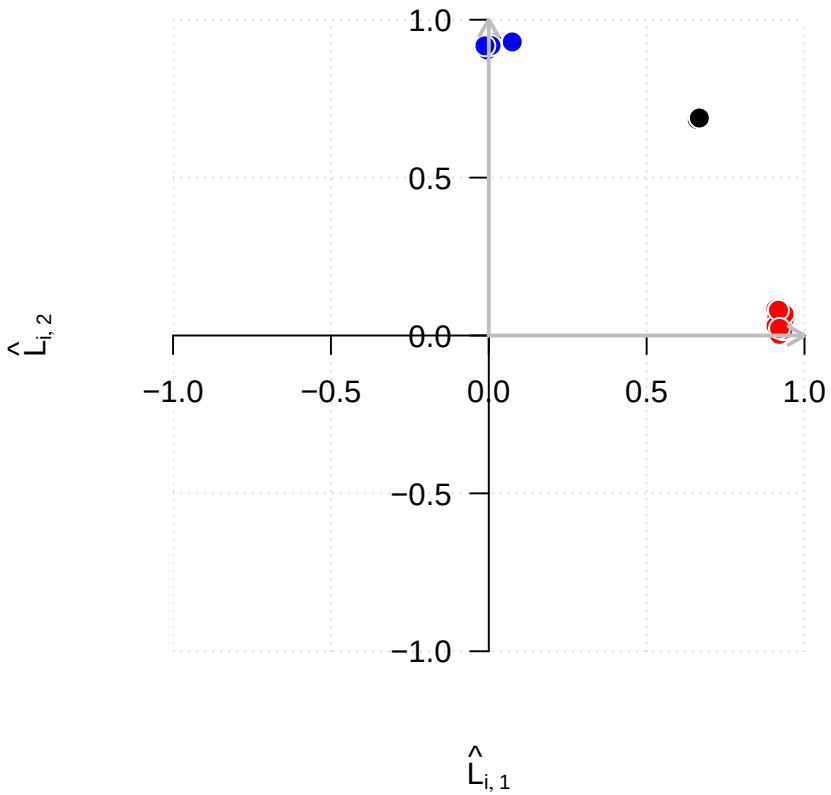
⇒ Faktorladungen bezüglich der kanonischen Basis und Varimaxbasis (→)



Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

$\Rightarrow$ Faktorladungen bezüglich der Varimax-Basisvektoren ($\rightarrow$)



Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

⇒ Durchführung mithilfe des `psych` Pakets (Revelle (2024))

```
library(psych)
Y           = as.matrix(read.csv("3-Daten/3-fa-keller.csv"))          # Y \in \mathbb{R}^{m x n}
R           = cor(Y)                                                  # m x m Korrelationsmatrix
res         =                                                         # Ergebnisobjekt
fa(                                                                   # fa() Funktion
R,                                                                    # Stichprobenkorrelationsmatrix
nfactors    = 2,                                                      # Anzahl an Faktoren
rotate      = "varimax",                                              # "Varimaxrotation"
SMC         = TRUE,                                                   # Kommunalitätsschätzerinitialisierung
min.err     = 0.01,                                                   # Konvergenzkritierium
max.iter    = 25,                                                     # Maximale Anzahl Iterationen
fm          = "pa")                                                   # "Iterierte Hauptachsenfaktorisierung"
L_hat_psych = res$loadings                                            # Varimax-Faktorladungsmatrixschätzer
```

Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

⇒ Durchführung mit Standardcode

```
round(L_hat_M,2)
```

```
                            [,1] [,2]
Traurigkeit                 0.66 0.68
Pessimismus                 0.93 0.08
Versagensgefühle            0.94 0.05
VerlustAnFreude             0.91 0.03
Schuldgefühle               0.01 0.93
Bestrafungsgefühle          0.01 0.92
Selbstablehnung             0.07 0.93
Selbstkritik               -0.01 0.90
Suizidgedanken              0.01 0.92
Weinen                      0.67 0.69
Unruhe                      0.91 0.08
Interessenverlust           0.92 0.00
Entschlusslosigkeit         0.92 0.08
Wertlosigkeit              -0.01 0.92
Energieverlust              0.93 0.02
Schlaf                      0.91 0.03
Reizbarkeit                 0.91 0.06
Appetitveränderung          0.91 0.03
Konzentrationsschwierigkeiten 0.94 0.07
Ermüdung                    0.92 0.08
VerlustSexuellenInteresses  0.92 0.02
```

Anwendungsbeispiel

"Faktorenextration mit iterierter Hauptachsenfaktorisierung und anschließender Varimaxrotation"

⇒ Durchführung mit psych

```
round(L_hat_psych[,1:2],2)
```

```
                              PA1  PA2
Traurigkeit                  0.66 0.68
Pessimismus                  0.93 0.08
Versagensgefühle             0.94 0.05
VerlustAnFreude              0.91 0.03
Schuldgefühle                0.01 0.93
Bestrafungsgefühle           0.01 0.92
Selbstablehnung              0.07 0.93
Selbstkritik                -0.01 0.90
Suizidgedanken               0.01 0.92
Weinen                       0.67 0.69
Unruhe                       0.91 0.08
Interessenverlust            0.92 0.00
Entschlusslosigkeit          0.92 0.08
Wertlosigkeit               -0.01 0.92
Energieverlust               0.93 0.02
Schlaf                       0.91 0.03
Reizbarkeit                  0.91 0.06
Appetitveränderung           0.91 0.03
Konzentrationsschwierigkeiten 0.94 0.07
Ermüdung                     0.92 0.08
VerlustSexuellenInteresses   0.92 0.02
```

Grundlagen

Modellformulierung

Modellschätzung

Modellevaluation

Modellinterpretation

Selbstkontrollfragen

Selbstkontrollfragen I

1. Erläutern Sie die Begriffe der latenten und observierbaren Variablen.
2. Erläutern Sie die Begriffe der explorativen und der konfirmatorischen Faktorenanalyse.
3. Geben Sie die Definitionen der Kovarianzmatrix und der Korrelationsmatrix eines Zufallsvektors wieder.
4. Geben Sie die Definition von Stichprobenkovarianzmatrix und Stichprobenkorrelationsmatrix wieder.
5. Geben Sie die Definition der Faktorenanalysemodelle in strukturelle Form wieder.
6. Geben Sie das Theorem zu den Faktorenanalysmodellen in Verteilungsform wieder.
7. Erläutern Sie die datengenerative Sicht eines Faktorenanalysemodells.
8. Geben Sie die Definition der Datensatzmodelle der Faktorenanalyse wieder.
9. Geben Sie Spearman's g-Faktor Modell in struktureller Form wieder und erläutern Sie es.
10. Geben Sie Thurstones's Multifaktorenmodell in struktureller Form wieder und erläutern Sie es.
11. Geben Sie das Modell paralleler Testmessungen in struktureller Faktorenanalysemodellform wieder.
12. Geben Sie das Theorem zur marginalen Datenkovarianzmatrix des Faktorenanalysemodells wieder.
13. Geben Sie das Theorem zur Varianzzerlegung der faktoranalytischen Datenkomponenten wieder.
14. Geben Sie die Definitionen von Kommunalität, Spezifität und Gesamtvarianz wieder.
15. Geben Sie die Definition einer orthogonalen Matrix wieder.
16. Geben Sie die Definition der Orthogonalen Transformation eines Faktorenanalysemodells wieder.
17. Geben Sie das Theorem zur Nichtidentifizierbarket und Kovarianzinvarianz des Faktorenanalysemodells wieder.
18. Geben Sie das Theorem zur Orthonormalzerlegung einer symmetrischen Matrix wieder.
19. Erläutern Sie die Begriffe Eigenvektor und Eigenwert bei Orthonormalzerlegung einer symmetrischen Matrix.
20. Erläutern Sie die Motivation der Hauptkomponentenschätzung der Parameter eines Faktorenanalysemodells.
21. Geben Sie die Definition der Hauptkomponentenschätzer $k$ter Ordnung von L und Ψ wieder.
22. Geben Sie die Definition von Kommunalitäts-, Spezifitäts-, und Gesamtvarianzschätzer wieder.
23. Erläutern Sie die Motivation des Maximum-Likelihood-Ansatzes zur Schätzung eines Faktorenanalysemodells.

Selbstkontrollfragen II

24. Geben Sie das Theorem zur Maximum-Likelihood-Schätzer eines Faktorenanalysemodells wieder.
25. Geben Sie die Definition der Diskrepanzfunktion eines Faktorenanalysemodells wieder.
26. Erläutern Sie die Interpretation hoher und niedriger Werte der Diskrepanzfunktion.
27. Geben Sie das Theorem zur Äquivalenz von Diskrepanzfunktionsbasierten und ML-Schätzern wieder.
28. Geben Sie die Definition der Stichprobenvarianzzerlegung der Faktorenanalyse wieder.
29. Geben Sie das Theorem zur Stichprobenvarianzzerlegung der Faktorenanalyse wieder.
30. Erläutern Sie die Motivation und das Vorgehen des Scree-Tests zur Bestimmung der Faktoranzahl.
31. Erläutern Sie die Motivation des χ^2-Tests der Faktorenanalyse.
32. Geben Sie das Theorem zum Log-Likelihood-Ratio-Kriterium und Diskrepanzfunktion wieder.
33. Geben Sie das Theorem zur X^2-Statistik eines Faktorenanalysemodells wieder.
34. Erläutern Sie die Interpretation hoher und niedriger P-Werte der X^2-Statistik.
35. Erläutern Sie die Motivation für Faktorenrotationen im Kontext exploratorischer Faktorenanalysen.
36. Erläutern Sie den Begriff der Einfachstruktur einer Faktorladungsmatrix.
37. Geben Sie das Theorem zum Begriff der Orthonormalbasis wieder.
38. Geben Sie die Definition des Begriffs des Orthomaxverfahrens wieder.
39. Erläutern Sie die Motivation und die technische Umsetzung von Orthomaxverfahren.
40. Geben Sie die Definition der Orthomaxzielfunktion wieder.
41. Geben Sie das Theorem zur Varimaxzielfunktion wieder.

Beispielklausurfrage

Welche Aussage zur Varimax-Zielfunktion trifft **nicht** zu?

a) Der Wert der Varimaxzielfunktion hängt nicht von den Faktorwerten der Testpersonen ab.
b) Der Wert der Varimaxzielfunktion hängt von einer (geschätzten) Faktorladungsmatrix ab.
c) Der Wert der Varimaxzielfunktion hängt von einer durch eine Matrix repräsentierten Orthogonalbasis ab.
d) Der Wert der Varimaxzielfunktion wird maximal, wenn alle Einträge der Faktorladungsmatrix gleich sind.

Ergänzungen aus der Linearen Algebra

Definition (Nullmatrizen, Einsmatrizen)

- Wir bezeichnen *Nullmatrizen* mit

$$0_{nm} := (0)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n \tag{171}$$

- Wir bezeichnen den *Einsmatrizen* mit

$$1_{nm} := (1)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 1_n := (1)_{1 \leq i \leq n} \in \mathbb{R}^n \tag{172}$$

Bemerkungen

- 0_{nm} und 0_n bestehen nur aus Nullen.
- 1_{nm} und 1_n bestehen nur aus Nullen.
- Es gelten zum Beispiel

$$0_n 0_n^T = 0_{nn} \text{ und } 1_n 1_n^T = 1_{nn}. \tag{173}$$

Definition (Einheitsmatrizen und Einheitsvektoren)

- Wir bezeichnen die *Einheitsmatrix* mit

$$I_n := (i_{jk})_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ für } j = k \text{ und } i_{jk} = 0 \text{ für } j \neq k \tag{174}$$

- Wir bezeichnen die *Einheitsvektoren* $e_i, i = 1, ..., n$ mit

$$e_i := (e_{i_j})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{i_j} = 1 \text{ für } i = j \text{ und } e_{i_j} = 0 \text{ für } i \neq j \tag{175}$$

Bemerkungen

- I_n besteht nur aus Nullen und Diagonalelementen gleich Eins.
- $e_i, i = 1, ..., n$ besteht nur aus Nullen und einer Eins in der $i$ten Komponente.
- Es gilt

$$I_n = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix} \tag{176}$$

- Es gelten weiterhin zum Beispiel für $1 \leq i, j \leq n$

$$e_i^T e_j = 0 \text{ für } i \neq j, e_i^T e_i = 1 \text{ und } e_i^T v = v^T e_i = v_i \text{ für } v \in \mathbb{R}^n. \tag{177}$$

Definition (Diagonalmatrix)

Eine Matrix $D \in \mathbb{R}^{n \times m}$ heißt *Diagonalmatrix*, wenn $d_{ij} = 0$ für $1 \leq i \leq n, 1 \leq j \leq m$ mit $i \neq j$.

Bemerkungen

- Eine Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ mit Diagonalelementen $d_1, ..., d_n$ schreibt man auch als

$$D = \text{diag}(d_1, ..., d_n). \tag{178}$$

- Diagonalmatrizen haben viele "gute" Eigenschaften.
- Zum Beispiel überzeugt man sich leicht davon, dass Multiplikation einer Matrix A von links mit einer Diagonalmatrix D der Multiplikation der Zeilen der Matrix A mit den entsprechenden Diagonaleinträgen von D entspricht. Die entsprechende Multiplikation von rechts entspricht der Multiplikation der Spalten von A mit entsprechenden Diagonaleinträgen von D.
- Eine weitere wichtige Eigenschaft ist

$$D := \text{diag}(d_1, ..., d_n) \Rightarrow |D| = \prod_{i=1}^{n} d_i \tag{179}$$

- Für Beweise dieser Eigenschaften wird auf die weiterführende Literatur, z.B. Searle (1982) verwiesen.

Definition (Symmetrische Matrix)

Eine Matrix $S \in \mathbb{R}^{n \times n}$ heißt *symmetrisch*, wenn gilt dass $S^T = S$.

Bemerkungen

- Symmetrische Matrizen sind spezielle quadratische Matrizen.
- Symmetrische Matrizen haben viele "gute" Eigenschaften.
- Beispielweise gilt für die Summe zweier symmetrischer Matrizen, dass auch diese wieder symmetrisch ist

$$A = A^T \text{ und } B = B^T \Rightarrow A + B = (A + B)^T \tag{180}$$

  und das die Inverse einer symmetrischen Matrix, sofern sie existiert, auch symmetrisch ist,

$$S^T = S \Rightarrow \left(S^{-1}\right)^T = S^{-1}. \tag{181}$$

- Für Beweise dieser Eigenschaften wird auf die weiterführende Literatur, z.B. Searle (1982), verwiesen.

Definition (Positiv-definite Matrizen und positiv-semidefinite Matrizen)

Eine quadratische Matrix $C \in \mathbb{R}^{n \times n}$ heißt *positiv-definit*, wenn C symmetrisch ist und gilt, dass

$$x^T C x > 0 \text{ für alle } x \in \mathbb{R}^n \text{ mit } x \neq 0_n. \tag{182}$$

Wenn $C \in \mathbb{R}^{n \times n}$ positiv-definit ist, so schreiben wir $C \in \mathbb{R}^{n \times n}$ pd.

Eine quadratische Matrix $C \in \mathbb{R}^{n \times n}$ heißt *positiv-semidefinit*, wenn C symmetrisch ist und gilt, dass

$$x^T C x \geq 0 \text{ für alle } x \in \mathbb{R}^n \text{ mit } x \neq 0_n. \tag{183}$$

Wenn $C \in \mathbb{R}^{n \times n}$ positiv-definit ist, so schreiben wir $C \in \mathbb{R}^{n \times n}$ psd.

Bemerkungen

- Positive-definite und positiv-semidefinite Matrizen sind spezielle symmetrische Matrizen.
- Kovarianzmatrixparameter von multivariaten Normalverteilungen sind positiv definite Matrizen.
- Kovarianzmatrizen von Zufallsvektoren sind positiv-semidefinite Matrizen.
- Positiv-definite Matrizen haben viele “gute” Eigenschaften.
- Beispielsweise existiert die Inverse C^{-1} einer positiv-definiten Matrix und ist selbst positiv-definit, für einen Beweis dieser Eigenschaft verweisen wir auf die weiterführende Literatur, z.B. Searle (1982).

Definition (Eigenvektor, Eigenwert)

$A \in \mathbb{R}^{m \times m}$ sei eine quadratische Matrix. Dann heißt jeder Vektor $v \in \mathbb{R}^m, v \neq 0_m$, für den gilt, dass

$$Av = \lambda v \tag{184}$$

mit $\lambda \in \mathbb{R}$ ein *Eigenvektor* von A. λ heißt zugehöriger *Eigenwert* von A.

Bemerkungen

- Ein Eigenvektor v von A wird durch A mit einem Faktor λ verlängert.
- Jeder Eigenvektor hat einen zugehörigen Eigenwert.
- Die Eigenwerte verschiedener Eigenvektoren können identisch sein.

Theorem (Multiplikativität von Eigenvektoren)

$A \in \mathbb{R}^{m \times m}$ sei eine quadratische Matrix. Wenn $v \in \mathbb{R}^m$ Eigenvektor von A mit Eigenwert $\lambda \in \mathbb{R}$ ist, dann ist für $c \in \mathbb{R}$ auch $cv \in \mathbb{R}^m$ Eigenvektor von A und zwar wiederum mit Eigenwert $\lambda \in \mathbb{R}$.

Beweis

Es gilt

$$Av = \lambda v \Leftrightarrow cAv = c\lambda v \Leftrightarrow A(cv) = \lambda(cv) \tag{185}$$

Also ist cv ein Eigenvektor von A mit Eigenwert λ.

□

Konvention

Wir betrachten nur Eigenvektoren mit $\|v\| = 1$.

Theorem (Eigenwerte und Eigenvektoren symmetrischer Matrizen)

$S \in \mathbb{R}^{m \times m}$ sei eine symmetrische Matrix. Dann gelten

(1) Die Eigenwerte von S sind reell.

(2) Die Eigenvektoren zu je zwei verschiedenen Eigenwerten von S sind orthogonal.

Bemerkung

- In nachfolgendem Beweis setzen wir die Tatsache dass eine symmetrische m reelle Eigenwerte hat als gegeben voraus und zeigen lediglich, dass die Eigenvektoren zu je zwei verschiedenen Eigenwerten einer symmetrischen Matrix orthogonal sind. Ein vollständiger Beweis des Theorems findet sich in Strang (2009), Kapitel 6.4.
- Da wir als Eigenvektoren nur Eigenvektoren der Länge 1 betrachten, sind die hier angesprochenen orthogonalen Eigenvektoren insbesondere auch orthonormal.

Beweis

Ohne Beschränkung der Allgemeinheit seien λ_i und λ_j mit $1 \leq i, j \leq m$ und $\lambda_i \neq \lambda_j$ zwei verschiedenen Eigenwerte von S mit zugehörigen Eigenvektoren q_i und q_j, respektive. Dann ergibt sich wie unten gezeigt, dass

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \tag{186}$$

Mit $q_i \neq 0_m, q_j \neq 0_m$ und $\lambda_i \neq \lambda_j$ folgt damit $q_i^T q_j = 0$, weil weil es keine andere Zahl c als die Null gibt, für die bei $a, b \in \mathbb{R}$ und $a \neq b$ gilt, dass

$$ac = bc. \tag{187}$$

Um

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \tag{188}$$

zu zeigen, halten wir zunächst fest, dass

$$Sq_i = \lambda_i q_i \Leftrightarrow (Sq_i)^T = (\lambda_i q_i)^T \Leftrightarrow q_i^T S^T = q_i^T \lambda_i^T \Leftrightarrow q_i^T S = q_i^T \lambda_i \Leftrightarrow q_i^T S q_j = \lambda_i q_i^T q_j \tag{189}$$

und

$$Sq_j = \lambda_j q_j \Leftrightarrow q_i^T S q_j = \lambda_j q_i^T q_j \tag{190}$$

gelten. Sowohl $\lambda_i q_i^T q_j$ als auch $\lambda_j q_i^T q_j$ sind also mit $q_i^T S q_j$ und damit auch miteinander identisch.

□

Theorem (Orthonormalzerlegung einer symmetrischen Matrix)

$S \in \mathbb{R}^{m \times m}$ sei eine symmetrische Matrix mit m verschiedenen Eigenwerten. Dann kannn S geschrieben werden als

$$S = Q \Lambda Q^T, \tag{191}$$

wobei $Q \in \mathbb{R}^{m \times m}$ eine orthogonale Matrix ist und $\Lambda \in \mathbb{R}^{m \times m}$ eine Diagonalmatrix ist.

Beweis

Es seien $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ die der Größe nach geordneten Eigenwerte von S und $q_1, q_2, \ldots, q_m$ seien die jeweils zugehörigen orthonormalen Eigenvektoren. Mit

$$Q := \begin{pmatrix} q_1 & q_2 & \cdots & q_m \end{pmatrix} \in \mathbb{R}^{m \times m} \text{ und } \Lambda := \text{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_m\right) \in \mathbb{R}^{m \times m}, \tag{192}$$

folgt dann mit den Definitionen von Eigenwerten und Eigenvektoren zunächst, dass

$$S q_i = \lambda_i q_i \text{ für } i = 1, \ldots, m \Leftrightarrow SQ = Q\Lambda. \tag{193}$$

Rechtseitige Multiplikation mit Q^T ergibt dann mit $QQ^T = I_m$, dass

$$SQQ^T = Q\Lambda Q^T \Leftrightarrow SI_m = Q\Lambda Q^T \Leftrightarrow S = Q\Lambda Q^T. \tag{194}$$

□

Bemerkungen

- $S = Q\Lambda Q^T$ heißt auch *Diagonalisierung* von S.
- Man wählt man als Diagonalelemente von Λ die der Größe nach geordneten Eigenwerte von S.
- Man wählt man als Spalten von Q die zugehörigen orthonormalen Eigenvektoren von S.

Theorem (Spur und Determinante einer symmetrischen Matrix)

$S \in \mathbb{R}^{m \times m}$ sei eine symmetrische Matrix mit verschiedenden Eigenwerten $\lambda_1, ..., \lambda_m$. Dann gelten

$$|S| = \prod_{i=1}^{m} \lambda_i \text{ und } \text{tr}(S) = \sum_{i=1}^{m} \lambda_i \tag{195}$$

Beweis

Mit dem Theorem zur Zerlegung einer symmetrischen Matrix mit verschiedenen Eigenwerten gilt, dass

$$|S| = |Q \Lambda Q^T| \tag{196}$$

wobei $Q \in \mathbb{R}^{m \times m}$ eine orthogonale Matrix und $\Lambda \in \mathbb{R}^{m \times m}$ die Diagonalmatrix der m verschiedenen Eigenwerte von S ist. Mit dem Determinantenmultiplikationssatz, der Determinanteneigenschaft von orthogonalen Matrizen und der Tatsache, dass die Determinante einer Diagonalmatrix dem Produkt ihrer Diagonalelemente entspricht, gilt dann weiterhin

$$|S| = |Q \Lambda Q^T| = |Q||\Lambda||Q^T| = |\Lambda| = \prod_{i=1}^{m} \lambda_i. \tag{197}$$

Wiederrum mit dem Theorem zur Zerlegung einer symmetrischen Matrix mit verschiedenen Eigenwerten gilt, dass

$$\text{tr}(S) = \text{tr}\left(Q \Lambda Q^T\right). \tag{198}$$

Mit der zyklischen Permutationsinvarianz der Spur, der Inversionseigenschaft orthogonaler Matrizen und der Definition der Spur gilt dann weiterhin

$$\text{tr}(S) = \text{tr}\left(Q \Lambda Q^T\right) = \text{tr}\left(Q^T Q \Lambda\right) = \text{tr}(\Lambda) = \sum_{i=1}^{m} \lambda_i. \tag{199}$$

□

Ergänzungen aus der Wahrscheinlichkeitstheorie

Theorem (Eigenschaften der Kovarianzmatrix)

ξ sei ein m-dimensionaler Zufallsvektor und $\mathbb{C}(\xi)$ sei seine Kovarianzmatrix. Dann gelten
(1) (Elemente) Die Elemente von $\mathbb{C}(\xi)$ sind die Kovarianzen der Komponenten von ξ,

$$\mathbb{C}(\xi) = \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1 \leq i,j \leq m}. \tag{200}$$

(2) (Kovarianzmatrixverschiebungssatz) Es gilt

$$\mathbb{C}(\xi) = \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T. \tag{201}$$

(3) (Linear-affine Transformation) Für $A \in \mathbb{R}^{n \times m}$ und $b \in \mathbb{R}^n$ gilt

$$\mathbb{C}(A\xi + b) = A\mathbb{C}(\xi)A^T. \tag{202}$$

(4) (Matrixeigenschaften) $\mathbb{C}(\xi)$ ist symmetrisch und positiv-semidefinit.

Beweis

(1) Es gilt

$$
\begin{aligned}
\mathbb{C}(\xi) &:= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) \\
&= \mathbb{E}\left(\left(\begin{pmatrix}\xi_1\\ \vdots \\ \xi_n\end{pmatrix} - \begin{pmatrix}\mathbb{E}(\xi_1)\\ \vdots \\ \mathbb{E}(\xi_n)\end{pmatrix}\right)\left(\begin{pmatrix}\xi_1\\ \vdots \\ \xi_n\end{pmatrix} - \begin{pmatrix}\mathbb{E}(\xi_1)\\ \vdots \\ \mathbb{E}(\xi_n)\end{pmatrix}\right)^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1)\\ \vdots \\ \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1)\\ \vdots \\ \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1)\\ \vdots \\ \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1) & \dots & \xi_n - \mathbb{E}(\xi_n)\end{pmatrix}\right) \\
&= \mathbb{E}\begin{pmatrix}(\xi_1 - \mathbb{E}(\xi_1))(\xi_1 - \mathbb{E}(\xi_1)) & \dots & (\xi_1 - \mathbb{E}(\xi_1))(\xi_n - \mathbb{E}(\xi_n)\\ \vdots & \ddots & \vdots \\ (\xi_n - \mathbb{E}(\xi_n))(\xi_1 - \mathbb{E}(\xi_1)) & \dots & (\xi_n - \mathbb{E}(\xi_n))(\xi_n - \mathbb{E}(\xi_n))\end{pmatrix} \\
&= \left(\mathbb{E}\left((\xi_i - \mathbb{E}(\xi_i))(\xi_j - \mathbb{E}(\xi_j))\right)\right)_{1\le i,j\le n} \\
&= \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1\le i,j\le n}.
\end{aligned}
\tag{203}
$$

Beweis (fortgeführt)

(2) Mit den Eigenschaften von Erwartungswerten gilt

$$\begin{aligned}\mathbb{C}(\xi) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) \\ &= \mathbb{E}\left(\xi\xi^T - \xi\mathbb{E}(\xi)^T - \mathbb{E}(\xi)\xi^T + \mathbb{E}(\xi)\mathbb{E}(\xi)^T\right) \\ &= \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T - \mathbb{E}(\xi)\mathbb{E}(\xi)^T + \mathbb{E}(\xi)\mathbb{E}(\xi)^T \\ &= \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T.\end{aligned} \tag{204}$$

(3) Mit den Eigenschaften von Erwartungswerten gilt

$$\begin{aligned}\mathbb{C}(A\xi + b) &= \mathbb{E}\left((A\xi + b - \mathbb{E}(A\xi + b))(A\xi + b - \mathbb{E}(A\xi + b))^T\right) \\ &= \mathbb{E}\left((A\xi + b - A\mathbb{E}(\xi) - b)(A\xi + b - A\mathbb{E}(\xi) - b)^T\right) \\ &= \mathbb{E}\left((A(\xi - \mathbb{E}(\xi)))(A(\xi - \mathbb{E}(\xi)))^T\right) \\ &= \mathbb{E}\left(A(\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T A^T\right) \\ &= A\mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) A^T \\ &= A\mathbb{C}(\xi)A^T.\end{aligned} \tag{205}$$

Beweis (fortgeführt)

(4) Die Symmetrie von $\mathbb{C}(\xi)$ folgt aus der Symmetrie der Kovarianz einer Zufallsvariable mit

$$\mathbb{C}(\xi_i, \xi_j) = \mathbb{C}(\xi_j, \xi_i) \text{ für alle } i = 1, ..., m, j = 1, ..., m. \tag{206}$$

Um die positive Semidefinitheit von $\mathbb{C}(\xi)$ nachzuweisen, ist zu zeigen, dass $a^T \mathbb{C}(\xi) a \geq 0$ für alle $a \in \mathbb{R}^m$ mit $a \neq 0_m$. Sei also $a \in \mathbb{R}^m$ mit $a \neq 0$. Dann gilt mit Aussage (3) für $A := a^T \in \mathbb{R}^{1 \times m}$, dass

$$a^T \mathbb{C}(\xi) a = \mathbb{C}(a^T \xi). \tag{207}$$

Weiterhin gilt mit der Definition der Kovarianzmatrix aber, dass

$$\mathbb{C}(a^T \xi) = \mathbb{E}\left(\left(a^T \xi - \mathbb{E}(a^T \xi)\right)^2\right) = \mathbb{V}\left(a^T \xi\right). \tag{208}$$

Da mit den Eigenschaften der Varianz die Varianz der Zufallsvariable $a^T \xi$ aber immer nichtnegativ ist, folgt

$$a^T \mathbb{C}(\xi) a = \mathbb{V}(a^T \xi) \geq 0 \tag{209}$$

und damit die positive Semidefinitheit von $\mathbb{C}(\xi)$.

Theorem (Datenmatrix und Stichprobenstatistiken)

Es sei

$$y := \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \tag{210}$$

eine $m \times n$ *Datenmatrix*, die durch die spaltenweise Konkatenation von n m-dimensionalen Zufallvektoren $y_1, \ldots, y_n$ gegeben sei. Dann ergeben sich

- für das Stichprobenmittel

$$\bar{y} = \frac{1}{n} y 1_n, \tag{211}$$

- für die Stichprobenkovarianzmatrix

$$C = \frac{1}{n-1} \left(y \left(I_n - \frac{1}{n} 1_{nn} \right) y^T \right), \tag{212}$$

- und für Stichprobenkorrelationsmatrix mit

$$D := \text{diag} \left(\sqrt{C_{y_{ii}}}^{-1}, i = 1, \ldots, m \right), \tag{213}$$

dass

$$R = DCD. \tag{214}$$

Beweis

Die Darstellung des Stichprobenmittels ergibt sich aus

$$\begin{aligned}
\bar{y} &:= \frac{1}{n}\sum_{i=1}^{n} y_i \\
&= \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{n} y_{i1} \\ \vdots \\ \sum_{i=1}^{n} y_{im} \end{pmatrix} \\
&= \frac{1}{n}\left(\begin{pmatrix} y_{11} & \cdots & y_{n1} \\ \vdots & \ddots & \vdots \\ y_{1m} & \cdots & y_{nm} \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right) \\
&= \frac{1}{n} y 1_n.
\end{aligned} \tag{215}$$

Beweis (fortgeführt)

Hinsichtlich der Darstellung der Stichprobenkovarianzmatrix halten wir zunächst fest, dass nach Definition gilt, dass

$$
\begin{aligned}
C &:= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})^T \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i y_i^T - y_i \bar{y}^T - \bar{y} y_i^T + \bar{y} \bar{y}^T \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i y_i^T - \sum_{i=1}^{n} y_i \bar{y}^T - \sum_{i=1}^{n} \bar{y} y_i^T + \sum_{i=1}^{n} \bar{y} \bar{y}^T \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i y_i^T - n \bar{y} \bar{y}^T - n \bar{y} \bar{y}^T + n \bar{y} \bar{y}^T \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i y_i^T - n \bar{y} \bar{y}^T \right).
\end{aligned}
\tag{216}
$$

Beweis

Mit $1_n 1_n^T = 1_{nn}$ ergibt sich dann weiterhin

$$
\begin{aligned}
y\left(I_n - \frac{1}{n}1_{nn}\right)y^T &= \left(yI_n - \frac{1}{n}y1_{nn}\right)y^T \\
&= yy^T - \frac{1}{n}y1_{nn}y^T \\
&= \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}\begin{pmatrix} y_1^T \\ \vdots \\ y_n^T \end{pmatrix} - \frac{1}{n}y1_n 1_n^T y^T \\
&= \sum_{i=1}^{n} y_i y_i^T - n\left(\frac{1}{n}y1_n\right)\left(\frac{1}{n}1_n^T y^T\right) \\
&= \sum_{i=1}^{n} y_i y_i^T - n\bar{y}\bar{y}^T \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^T \\
&= C.
\end{aligned}
\tag{217}
$$

Beweis (fortgeführt)

Hinsichtlich der Korrelationsmatrix ergibt sich nach Definition und für ein beliebiges Indexpaar i, j mit $1 \leq i, j \leq m$ schließlich, dass

$$\begin{aligned} R_{y_{ij}} &= \frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \\ &= \frac{1}{\sqrt{(C)_{ii}}}(C)_{ij}\frac{1}{\sqrt{(C)_{jj}}} \\ &= (DCD)_{ij}. \end{aligned} \tag{218}$$

Referenzen I

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed). Wiley.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*, *67*(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley New York.

Braeken, J., & Van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Carroll, J. B. (1953). An Analytical Solution for Approximating Simple Structure in Factor Analysis. *Psychometrika*, *18*(1), 23–38. https://doi.org/10.1007/BF02289025

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Cliff, N. (1988). The Eigenvalues-Greater-Than-One Rule and the Reliability of Components. *Psychological Bulletin*, *103*(2), 276–279.

Crawford, C. B., & Ferguson, G. A. (1970). A General Rotation Criterion and Its Use in Orthogonal Rotation. *Psychometrika*, *35*(3), 321–332. https://doi.org/10.1007/BF02310792

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, *4*(3), 272–299.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*(594-604), 309–368. https://doi.org/10.1098/rsta.1922.0009

Referenzen II

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, *40*(7), 3510–3521. https://doi.org/10.1007/s12144-019-00300-2

Groh, D., & Ostwald, D. (2025). *The Psychometric Construction of Empathy: An Exploratory Factor Analysis of Empathy Self-Rating Subscales*. PsyArXiv. https://doi.org/10.31234/osf.io/ehncx_v3

Guttman, L. (1954). Some Necessary Conditions for Common-Factor Analysis. *Psychometrika*, *19*(2), 149–161. https://doi.org/10.1007/BF02289162

Hautzinger, M., Keller, F., & Kühner, C. (2006). *BDI-II Beck Depressions-Inventar*. Pearson.

Horst, P. (1965). *Factor analysis of data matrices*. Holt, Rinehart and Winston, Inc.

Horst, P., Wallin, P., Guttman, L., Wallin, F. B., Clausen, J. A., Reed, R., & Rosenthal, E. (1941). *The prediction of personal adjustment: A survey of logical problems and research techniques, with illustrative application to problems of vocational selection, school success, marriage, and crime.* Social Science Research Council. https://doi.org/10.1037/11521-000

Hotelling, H. (1933). Analysis of complex variables into principal components. *Journal of Educational Psychology*, *24*, 417-441 and 498-520.

Jöreskog, K. G. (1967). *Some contributions to maximum likelihood factor analysis*.

Jöreskog, K. G. (1970). A General Method for Analysis of Covariance Structures. *Biometrika*, *57*(2), 239. https://doi.org/10.2307/2334833

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. https://doi.org/10.1007/BF02289233

Referenzen III

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. https://doi.org/10.1007/BF02291817

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Keller, F., Hautzinger, M., & Kühner, C. (2008). Zur faktoriellen Struktur des deutschsprachigen BDI-II. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *37*(4), 245–254. https://doi.org/10.1026/1616-3443.37.4.245

Lawley, D. (1940). The Estimation of Factor Loadings by the Method of Maximum Likelihood. *Proceedings of the Royal Society of Edinburgh. Section B: Biological Sciences*.

Lawley, D. N. (1943). On Problems connected with Item Selection and Test Construction. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, *61*(3), 273–287. https://doi.org/10.1017/S0080454100006282

Lawley, D., & Maxwell, A. (1962). Factor Analysis as a Statistical Method. *The Statistician*, *12*(3), 209. https://doi.org/10.2307/2986915

Lawley, D., & Maxwell, A. (1971). *Factor analysis as a statistical method* (2nd ed.). London Butterworths.

Magnus, J. R., & Neudecker, H. (1989). Matrix Differential Calculus with Applications in Statistics and Econometrics. *Journal of the American Statistical Association*, *84*(408), 1103. https://doi.org/10.2307/2290109

Mulaik, S. A. (2010). *Foundations of Factor Analysis*. CRC Press, Taylor & Francis Group.

Muthén, L. K., & Muthén, B. O. (1998--2017). *Mplus User's Guide. Eighth Edition*.

Referenzen IV

Neudecker, H. (1969). Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products. *Journal of the American Statistical Association*, *64*.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Raîche, G., Walls, T. A., Magis, D., Riopel, M., & Blais, J.-G. (2013). Non-Graphical Solutions for Cattell's Scree Test. *Methodology*, *9*(1), 23–29. https://doi.org/10.1027/1614-2241/a000051

Revelle, W. (2024). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University.

Rippe, D. D. (1953). Application of a large sampling criterion to some sampling problems in factor analysis. *Psychometrika*, *18*(3), 191–205. https://doi.org/10.1007/BF02289056

Rohe, K., & Zeng, M. (2023). Vintage factor analysis with Varimax performs statistical inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *85*(4), 1037–1060. https://doi.org/10.1093/jrsssb/qkad029

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

Rosseel, Y. (2021). Evaluating the Observed Log-Likelihood Function in Two-Level Structural Equation Modeling with Missing Data: From Formulas to R Code. *Psych*, *3*(2), 197–232. https://doi.org/10.3390/psych3020017

Rosseel, Y., & Vidal, M. (2025). A tutorial for understanding SEM using R: Where do all the numbers come from? *British Journal of Mathematical and Statistical Psychology*, bmsp.70003. https://doi.org/10.1111/bmsp.70003

Searle, S. (1982). *Matrix Algebra Useful for Statistics*. Wiley-Interscience.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201. https://doi.org/10.2307/1412107

Referenzen V

Stefana, A., & Hill, C. E. (2025). Session evaluation scale: Psychometric evaluation and development of short versions. *Psychotherapy Research*, 1–17. https://doi.org/10.1080/10503307.2025.2587694

Strang, G. (2009). *Introduction to Linear Algebra*. Cambridge University Press.

Thurstone, L. (1936). The Factorial Isolation of Primary Abilities. *Psychometrika*, *1*(3), 175–182. https://doi.org/10.1007/BF02288363

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* University of Chicago Press. https://doi.org/10.1037/10018-000

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. University of Chicago Press.

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62. https://doi.org/10.1214/aoms/1177732360

Wishart, J. (1928). The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, *20A*(1/2), 32. https://doi.org/10.2307/2331939

Yanai, H., & Ichikawa, M. (2007). Factor Analysis. In *Handbook of statistics, Vol 26* (1st ed). Elsevier.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of Five Rules for Determining the Number of Components to Retain. *Psychological Bulletin*, *99*(3), 452–442.