Psychometrische Diagnostik

MSc Psychologie

MSc Umweltpsychologie / Mensch-Technik-Interaktion

SoSe 2026

Prof. Dr. Dirk Ostwald

(3) Faktorenanalyse

Einführung

Faktorenanalysemodelle

Parameterschätzung

Modellvergleichskriterien

Selbstkontrollfragen

Einführung

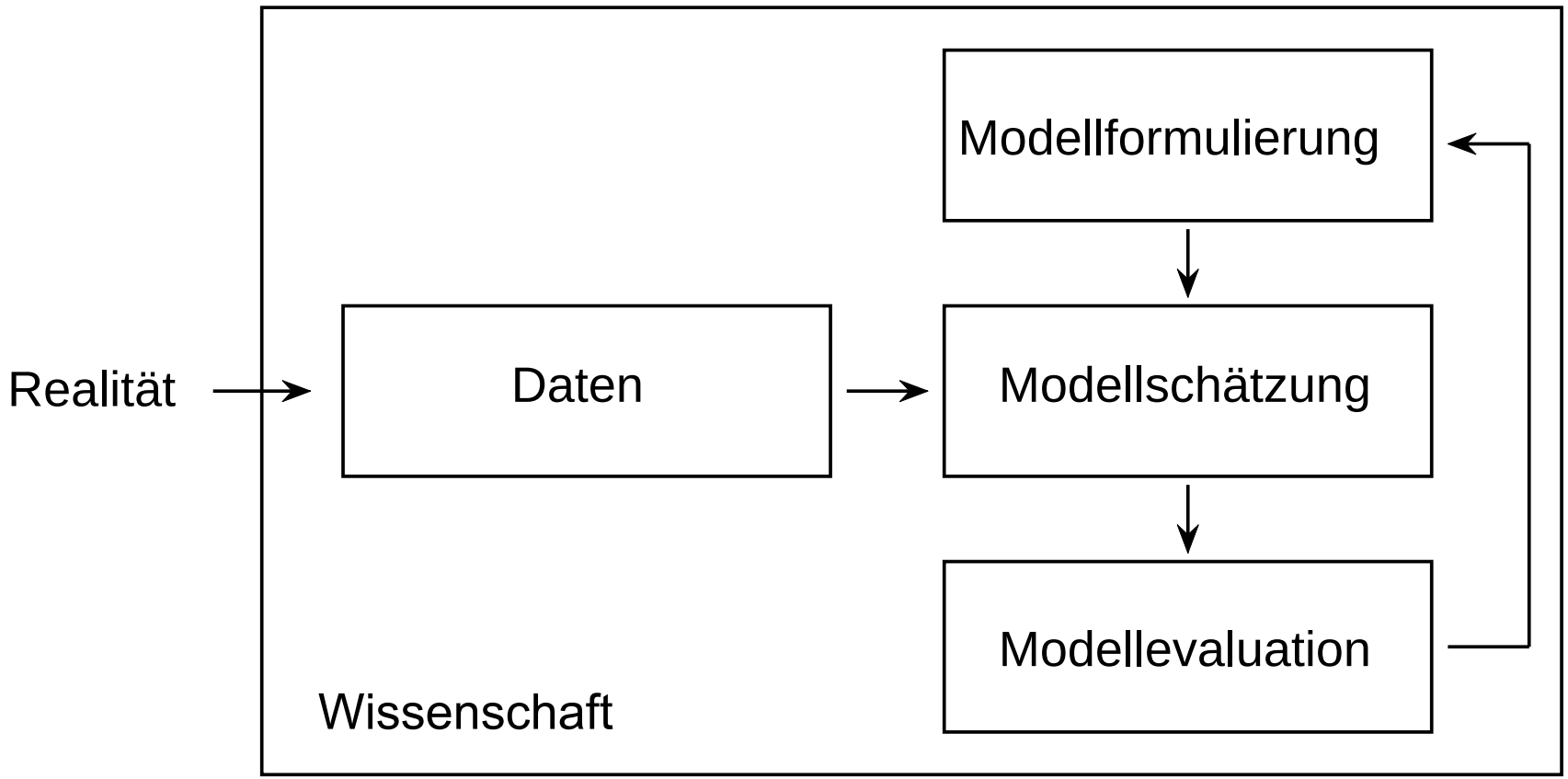
Faktorenanalysemodelle

Parameterschätzung

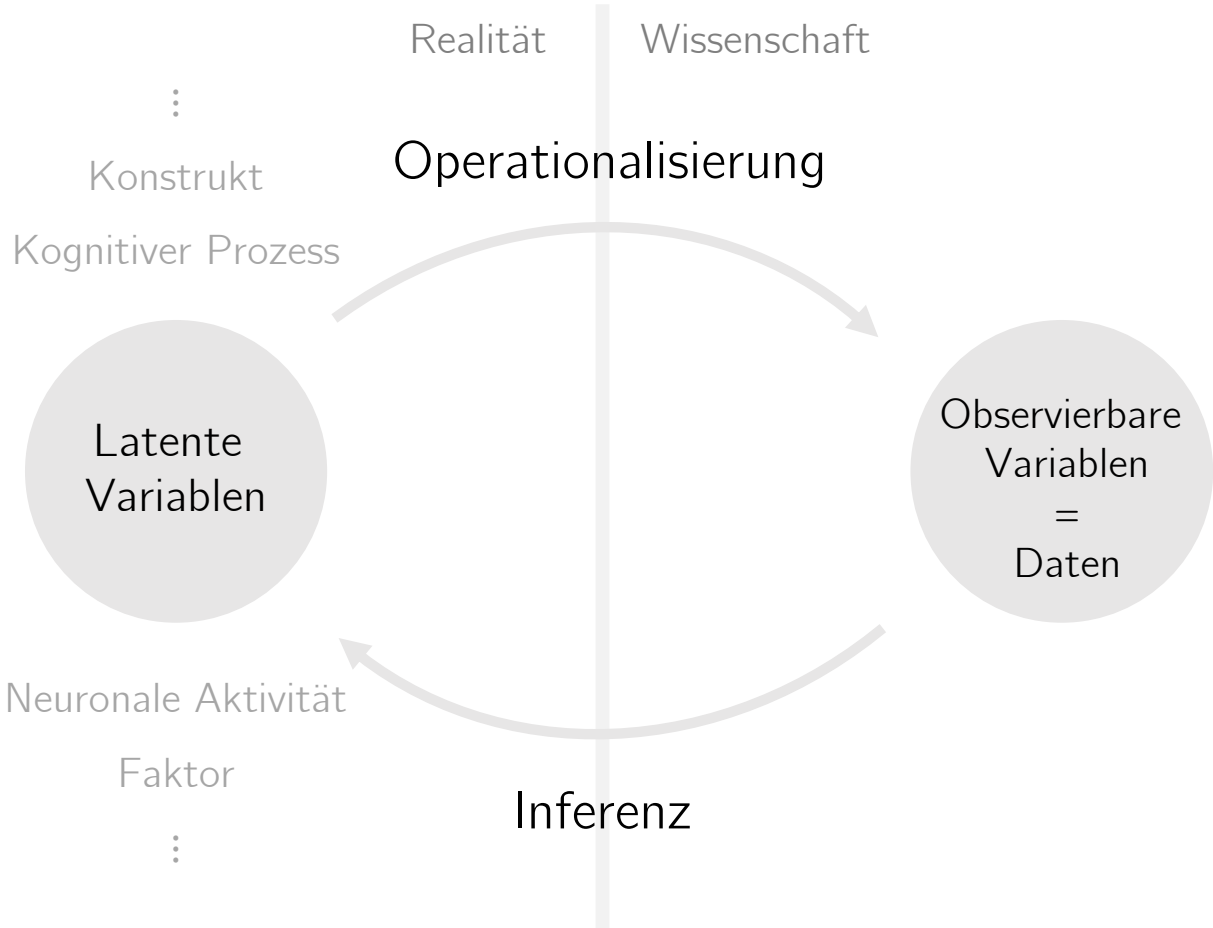
Modellvergleichskriterien

Selbstkontrollfragen

Psychologische Datenwissenschaft



Psychologische Datenwissenschaft



Einführung

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 numerischen Behandlung von Stichprobenkovarianzmatrizen

Konfirmative Faktorenanalyse (CFA)

- Moderneres Verfahren mit expliziter probabilistischer Modellspezifikation
- Fokus auf probabilistischer Parameterschätzung und Modellvergleichskriterien

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
- 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 Strukturgleichungsmodellen (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

Datenanalyseansätze in der angewandten Faktorenanalyseliteratur

Explorativer Ansatz ("Exploratorische Faktorenanalyse")

- Annahme eines normalverteilungsfreien Faktorenanalysemodells
- Hauptkomponentenschätzung bzw. Hauptachsenfaktorisierung
- Eigenwertbasierte Faktorenanzahlwahl durch z.B. Scree-Test oder Kaiser-Guttman-Kriterium
- Verbale Interpretation der faktorbasierten Itemcluster (vgl. Groh & Ostwald (2025))

Konfirmatorischer Ansatz ("Konfirmatorische Faktorenanalyse")

- Annahme des Normalverteilungsmodells 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.
- Typische Beispiele in diesem Sinne sind Stefana & Hill (2025) oder viele anwendungsorientierte Skalenanalysen.

Anwendungsbeispiel

Holzinger & Swineford (1939) Datensatz aus dem R-Paket `lavaan` (Rosseel (2012))

- Mentale Fähigkeitstests von Schüler:innen der siebten und achten Klasse aus zwei Schulen in Wisconsin, USA
- Die `lavaan`-Version enthält 301 Beobachtungen und 9 Indikatoren aus ursprünglich 26 Tests.

Fähigkeit	Kurzbeschreibung	Items
Räumliches Vorstellungsvermögen	Visuelle Wahrnehmung, Würfel, Rauten	y_1, y_2, y_3
Sprachvermögen	Textverständnis, Satzvervollständigung, Wortbedeutung	y_4, y_5, y_6
Kognitive Schnelligkeit	Addition, Punktezählen, Unterscheidung von Großbuchstaben	y_7, y_8, y_9

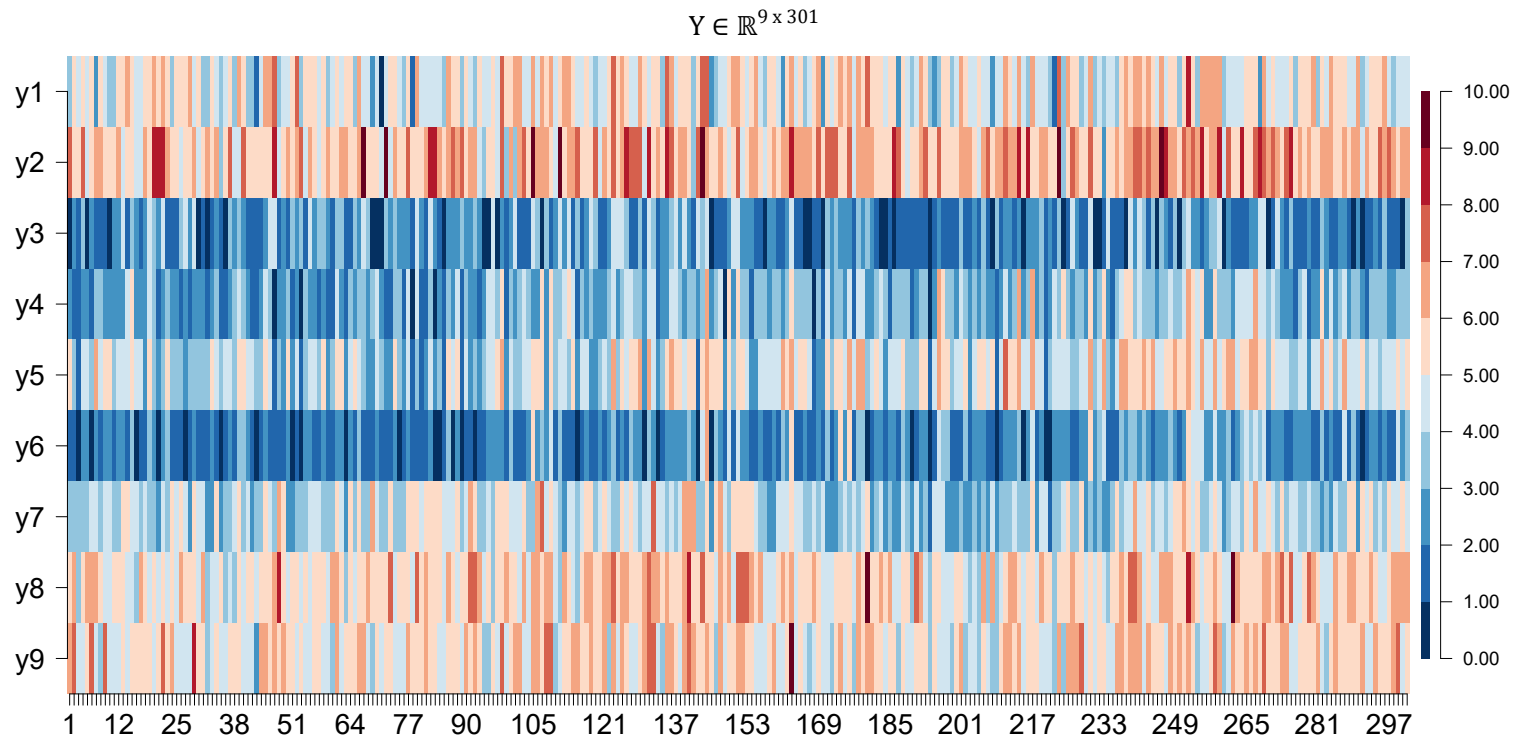
Testpersonen 1 - 10 von n = 301 Schüler:innen

	1	2	3	4	5	6	7	8	9	10
y1	3.33	5.33	4.50	5.33	4.83	5.33	2.83	5.67	4.50	3.50
y2	7.75	5.25	5.25	7.75	4.75	5.00	6.00	6.25	5.75	5.25
y3	0.38	2.12	1.88	3.00	0.88	2.25	1.00	1.88	1.50	0.75
y4	2.33	1.67	1.00	2.67	2.67	1.00	3.33	3.67	2.67	2.67
y5	5.75	3.00	1.75	4.50	4.00	3.00	6.00	4.25	5.75	5.00
y6	1.29	1.29	0.43	2.43	2.57	0.86	2.86	1.29	2.71	2.57
y7	3.39	3.78	3.26	3.00	3.70	4.35	4.70	3.39	4.52	4.13
y8	5.75	6.25	3.90	5.30	6.30	6.65	6.20	5.15	4.65	4.55
y9	6.36	7.92	4.42	4.86	5.92	7.50	4.86	3.67	7.36	4.36

Anwendungsbeispiel

Holzinger-Swineford-Datensatz basierend auf Holzinger & Swineford (1939)

- Gesamter Datensatz (n = 301)



Definition (Kovarianzmatrix eines Zufallsvektors)

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

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

Bemerkungen

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

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 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

Holzinger-Swineford-Datensatz basierend auf Holzinger & Swineford (1939)

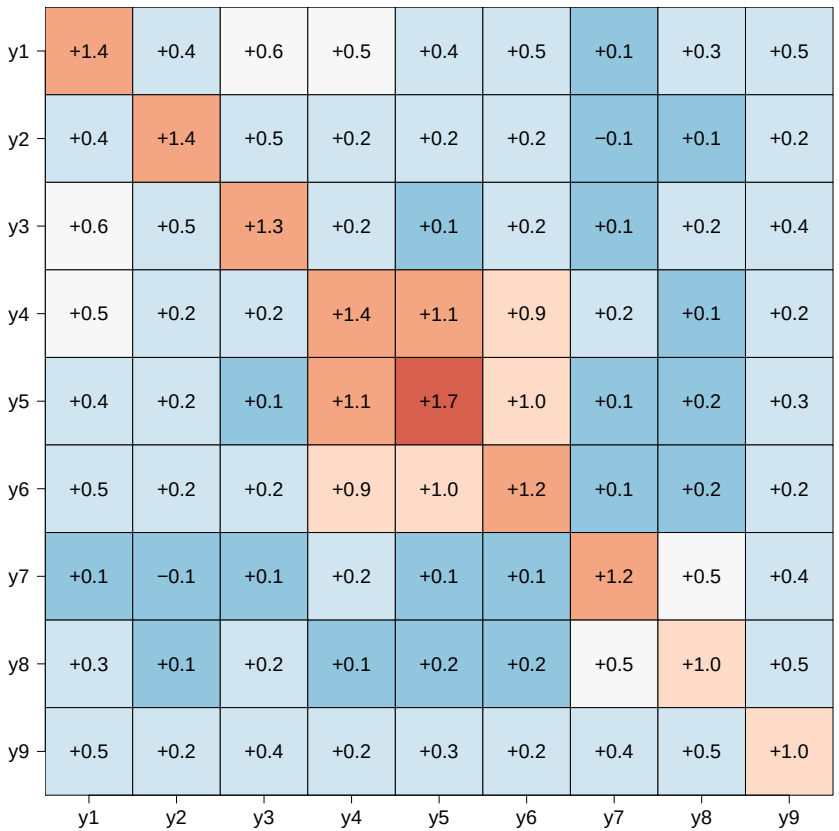
Berechnung von Stichprobenkovarianzmatrix und Stichprobenkorrelationsmatrix

```
Y         = as.matrix(t(read.csv("3-Daten/3-fa-hs.csv")))       # Y \in \mathbb{R}^{m x n}
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

Holzinger-Swineford-Datensatz basierend auf Holzinger & Swineford (1939)

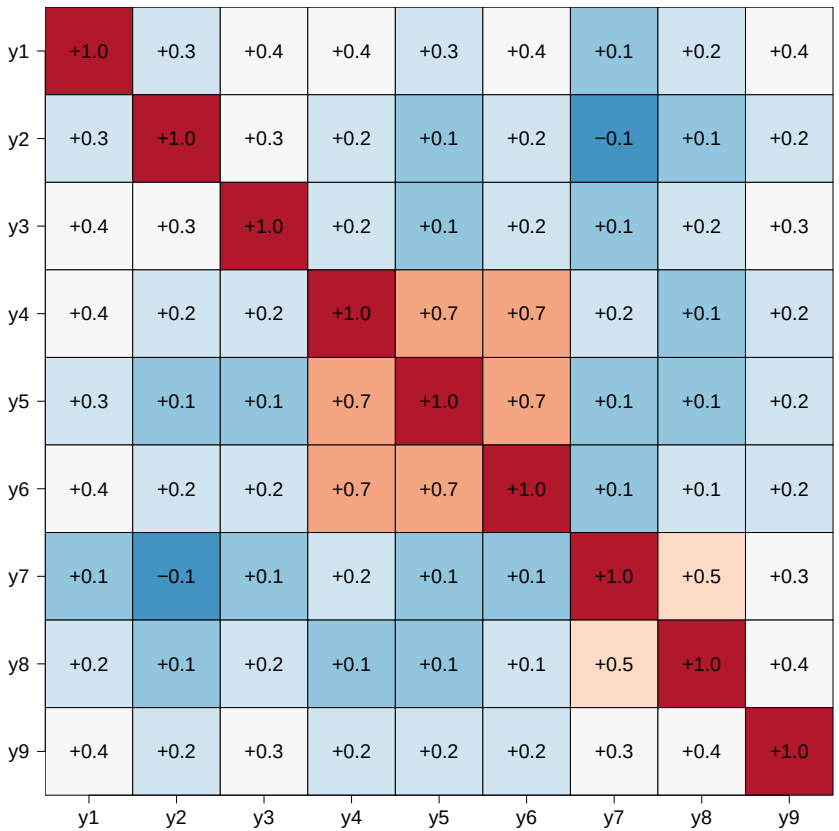
Stichprobenkovarianzmatrix



Anwendungsbeispiel

Holzinger-Swineford-Datensatz basierend auf Holzinger & Swineford (1939)

Stichprobenkorrelationsmatrix



Einführung

Faktorenanalysemodelle

Parameterschätzung

Modellvergleichskriterien

Selbstkontrollfragen

Definition (Faktorenanalysemodelle in struktureller Form)

Es sei

$$y = \mu + Lx + \varepsilon \tag{8}$$

wobei für $m > p$

- y ein m-dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird,
- $\mu \in \mathbb{R}^m$ ein fester, unbekannter Vektor ist, der *Datenerwartungswert* genannt wird,
- $L = (l_{jk}) \in \mathbb{R}^{m \times p}$ eine feste, unbekannte Matrix ist, die *Faktorladungsmatrix* genannt wird,
- x ein p-dimensionaler nicht-beobachtbarer Zufallsvektor ist, dessen Komponenten *Faktoren* genannt werden und für den gilt, dass

$$\mathbb{E}(x) = 0_p \text{ und } \mathbb{C}(x) = I_p, \tag{9}$$

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

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

Dann heißt (4) *Faktorenanalysemodell in struktureller Form*. Gilt darüber hinaus, dass

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

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

Faktorenanalysemodelle

Bemerkungen

- Die $j = 1, \ldots, m$ Komponenten von y modellieren *Itemscores* eines psychologischen Tests einer Testperson.
- Die $k = 1, \ldots, p$ Komponenten von x werden manchmal *gemeinsame Faktoren (common factors)* genannt.
- Die $j = 1, \ldots, m$ Komponenten von ε werden manchmal *spezifische Faktoren (unique factors)* genannt.
- l_{jk} wird die *Faktorladung des $k$ten Faktors auf die $j$te Datenkomponente* genannt.
- Der Eintrag l_{jk} von L entspricht dem Wert y_j für $x = e_k$ ($k$ter kanonischer Basisvektor).
- Für die marginale Kovarianzmatrix von y ist der Wert von μ unerheblich.
- Man nimmt entsprechend häufig an, dass $\mu = 0_m$ gilt, um die Modellformulierung zu vereinfachen.
- Auf Datenebene entspricht dies der Annahme der Datenzentrierung, dass y_i durch $y_i - \bar{y}$ ersetzt wird.
- An Schätzwerten von μ besteht in der Regel kein Interesse.
- Wir schreiben das Faktorenanalysemodell meist in der Kurzform

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim (0_p, I_p) \text{ und } \varepsilon \sim (0_m, \Psi). \tag{11}$$

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

Theorem (Faktorenanalysemodelle in Verteilungsform)

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

$$\mathbb{P}(x, y) = \mathbb{P}(x)\mathbb{P}(y|x) \text{ mit } x \sim (0_p, I_p) \text{ und } y \,|\, x \sim (\mu + Lx, \Psi) \tag{12}$$

und das Normalverteilungsmodell der Faktorenanalyse kann äquivalent in der Wahrscheinlichkeitsdichteform

$$p(x, y) = p(x)p(y|x) \text{ mit } p(x) = N(x; 0_p, \Phi) \text{ und } p(y|x) = N(y; \mu + Lx, \Psi). \tag{13}$$

geschrieben werden.

Bemerkungen

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

(1) Es wird ein p-dimensionaler Wert von x anhand von $\mathbb{P}(x)$ realisiert

(2) x wird durch L transformiert und bildet unter Addition von μ den Erwartungswert von $\mathbb{P}(y|x)$.

(3) Es wird ein m-dimensionaler Wert von ε realisiert

(4) Die m-dimensionalen Werte von μ, Lx und ε werden zu einem Wert von y addiert.

Faktorenanalysemodelle

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 (analog zur Annahme eines wahren Werts der klassischen Testtheorie), dass dieser Zufallsvektorrealisierung eine Realisierung des nicht-beobachtbaren Zufallsvektors x_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 $x_1, ..., x_n$ über Testpersonen 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.

Bemerkungen

- Für $p := 3$ Faktoren, $m := 9$ Testitems und $n := 10$ Testpersonen z.B.

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$
x_{i1}	2	5	1	0	4	3	2	1	5	0
x_{i2}	4	0	2	5	3	4	1	2	0	5
x_{i3}	1	2	4	3	0	5	3	4	1	2
y_{i1}	1	3	2	5	0	4	2	5	3	1
y_{i2}	0	2	5	3	4	1	5	0	2	3
y_{i3}	3	1	4	2	5	0	4	1	3	2
y_{i4}	5	4	0	3	1	2	0	4	5	3
y_{i5}	2	5	3	1	4	0	2	3	1	5
y_{i6}	4	3	1	0	2	5	1	4	2	0
y_{i7}	0	1	3	4	5	2	3	0	4	1
y_{i8}	3	0	5	2	1	4	5	2	0	3
y_{i9}	1	4	0	5	3	1	4	5	2	0

- 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.

Definition (Datensatzmodelle der Faktorenanalyse)

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

$$\mathbb{P}(x_1, y_1, ..., x_n, y_n) = \prod_{i=1}^{n} \mathbb{P}(x_i, y_i) \text{ mit } \mathbb{P}(x_i, y_i) = \mathbb{P}(x_{i'}, y_{i'}) \text{ für } 1 \leq i, i' \leq n \tag{14}$$

oder in Kurzform

$$(x_1, y_1), ..., (x_n, y_n) \sim \mathbb{P}(x, y) \tag{15}$$

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

$$p(x_1, y_1, ..., x_n, y_n) = \prod_{i=1}^{n} p(x_i, y_i) \text{ mit } p(x_i, y_i) = p(x_{i'}, y_{i'}) \text{ für } 1 \leq i, i' \leq n \tag{16}$$

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.
- $p(x_1, y_1, ..., x_n, y_n)$ ist die multivariate Normalverteilung eines $n(m + p)$-dimensionalen Zufallsvektors.

Beispiel (1) Spearmans g-Faktor Modell

Grundlage von Spearmans 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_j mithilfe eines Faktorenanalysemodells der Form

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix} + \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{pmatrix} x + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \Leftrightarrow y_j = l_j x + \varepsilon_j \text{ für } j = 1, \ldots, 6. \tag{18}$$

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

Entsprechend wird auch oft von *Spearmans Zweifaktorenmodell* gesprochen, wobei es im Sinne der modernen Faktorenanalyse allerdings nur einen Faktor und einen Vektor von Zufallsfehlern in diesem Modell gibt.

Beispiel (2) Thurstones Multifaktorenmodell

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

In struktureller Form hat dieses Modell die Form

$$y_j = \mu_j + l_{j1}x_1 + l_{j2}x_2 + \cdots + l_{jp}x_p + \varepsilon_j \text{ für } j = 1, \ldots, m \tag{19}$$

und entspricht damit der allgemeinen Form eines Faktorenanalysemodells, wobei hier die Faktoren $x_1, \ldots, x_p$ als *gemeinsame Faktoren (common factors)* und die Zufallsfehler $\varepsilon_1, \ldots, \varepsilon_m$ als *spezifische Faktoren (unique factors)* bezeichnet werden.

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

Beispiel (3) Das Modell paralleler Testmessungen

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

$$y_j = x + \varepsilon_j \text{ für } j = 1, ..., m \tag{20}$$

mit Messfehler ε_j.

Dann entspricht dieses Modell dem Faktorenanalysemodell

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} x + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{21}$$

Insbesondere entspricht der Faktor x der Faktorenanalyse hier also dem wahren Wert x der klassischen Testtheorie und die Faktorladungsmatrix hat die als bekannt vorausgesetzte Form $L := 1_m$. ->

Beispiel (4) Holzinger-Swineford-Modell

Holzinger & Swineford (1939) berichten Leistungstestdaten aus einer faktorenanalytischen Untersuchung zur Stabilität einer Bi-Faktor-Lösung. Eine heute häufig betrachtete konfirmatorische Spezifikation des in `lavaan` verfügbaren Teildatensatzes modelliert die 9 Indikatoren durch drei korrelierte Faktoren (Jöreskog (1969); Rosseel & Vidal (2025)): einen visuellen Faktor, einen textuellen Faktor und einen Geschwindigkeitsfaktor.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 0 & 0 \\ l_{31} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & l_{52} & 0 \\ 0 & l_{62} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & l_{83} \\ 0 & 0 & l_{93} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix}. \tag{22}$$

Theorem (Marginale Datenkovarianzmatrix des Normalverteilungsmodells der Faktorenanalyse)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{23}$$

Dann gilt für die Verteilung von y, dass

$$p(y) = N(y; \mu, L\Phi L^T + \Psi). \tag{24}$$

und insbesondere für die marginale Kovarianzmatrix von y, dass

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

Beweis

Das Theorem folgt direkt aus den Theoremen zu gemeinsamen und marginalen Normalverteilungen

□

Definition (Orthogonale Matrix)

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

$$Q^T Q = I_m. \tag{26}$$

Bemerkung

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

$$Q = \begin{pmatrix} q_1 & \cdots & q_m \end{pmatrix} \text{ mit } q_i \in \mathbb{R}^m \text{ für } 1 \leq i \leq m, \tag{27}$$

  dass

$$q_i^T q_j = 0 \text{ für } i \neq j \text{ und } q_i^T q_j = 1 \text{ für } i = j \text{ mit } 1 \leq i, j \leq m. \tag{28}$$

- 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

$$Q^{-1} = Q^T. \tag{29}$$

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

$$QQ^T = I_m. \tag{30}$$

(3) (Determinante) Es gilt

$$|Q| = 1 \text{ oder } |Q| = -1. \tag{31}$$

Beweis

(1) Weil $Q^TQ = I_m$ gilt, ist Q injektiv und als quadratische Matrix invertierbar. Damit gilt

$$Q^TQ = I_m \Leftrightarrow Q^TQQ^{-1} = I_mQ^{-1} \Leftrightarrow Q^{-1} = Q^T. \tag{32}$$

(2) Mit Eigenschaft (1) gilt

$$QQ^T = QQ^{-1} = I_m. \tag{33}$$

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

$$|Q|^2 = |Q||Q| = |Q||Q^T| = |QQ^T| = |I_m| = 1. \tag{34}$$

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

Definition (Orthogonale Transformation des Normalverteilungsmodells der Faktorenanalyse)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{35}$$

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

$$\tilde{y} = \mu + \tilde{L}\tilde{x} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{x} := Q^T x \tag{36}$$

eine *orthogonale Transformation des Normalverteilungsmodells der Faktorenanalyse*.

Bemerkungen

- Die orthogonale Transformation eines Faktorenanalysemodells ist Grundlage seiner Rotationsindeterminiertheit.

Theorem (Rotationsindeterminiertheit und Kovarianzinvarianz)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{37}$$

sowie eine seiner orthogonalen Transformationen

$$\tilde{y} = \mu + \tilde{L}\tilde{x} + \varepsilon \text{ mit } \tilde{L} := LQ \text{ und } \tilde{x} := Q^T x \tag{38}$$

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

$$\mathbb{C}(\tilde{y}) = \mathbb{C}(y) \tag{39}$$

Beweis

Wir halten zunächst fest, dass für die Verteilung von $\tilde{x}$ gilt, dass

$$p(\tilde{x}) = N(\tilde{x}; 0_p, Q^T \Phi Q). \tag{40}$$

Mit dem Theorem zu gemeinsamen und marginalen Normalverteilungen gilt dann, dass

$$\begin{aligned} p(\tilde{y}) &= N(\tilde{y}; \mu, \tilde{L} Q^T \Phi Q \tilde{L}^T + \Psi) \\ &= N(\tilde{y}; \mu, LQQ^T \Phi Q (LQ)^T + \Psi) \\ &= N(\tilde{y}; \mu, LQQ^T \Phi QQ^T L^T + \Psi) \\ &= N(\tilde{y}; \mu, L\Phi L^T + \Psi) \\ &= p(y) \end{aligned} \tag{41}$$

Bemerkungen

- Orthogonale Transformationen lassen die Verteilung von y unverändert. Es gibt also unendlich viele verschiedene Faktorladungsmatrizen und Faktorwerte, die die gleiche Datenverteilung und insbesondere die gleiche Datenkovarianz erklären können. Umgekehrt kann aus einer gegebenen Stichprobenkovarianzmatrix also im Allgemeinen nicht eindeutig auf die Werte der Faktorladungsmatrix und der Faktoren geschlossen werden.
- Fundamental ist die Nichtidentifizierbarkeit des Faktorenanalysemodells dadurch bedingt, dass nach Modellannahme weder die Faktorladungsmatrix, noch die Werte der Faktoren, noch die Beobachtungsfehler bekannt sind und Daten daher durch das Zusammenspiel dreier unbekannter Entitäten generiert werden. Nichtsdestotrotz können Faktorenanalysemodelle durch zusätzliche Parameterannahmen prinzipiell identifizierbar gemacht werden, wie im Folgenden diskutiert.

Definition (Identifizierbares Normalverteilungsmodell der Faktorenanalyse)

Gegeben sei ein Normalverteilungsmodell der Faktorenanalyse

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{42}$$

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

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{43}$$

so dass die marginale Datenkovarianz geschrieben werden kann als

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{44}$$

Dann heißen die Kovarianzstruktur des Normalverteilungsmodells der Faktorenanalyse und der Parametervektor θ *identifizierbar*, wenn für alle $\theta_1, \theta_2 \in \Theta$ gilt, dass

$$\Sigma_{\theta_1} = \Sigma_{\theta_2} \Rightarrow \theta_1 = \theta_2. \tag{45}$$

Wenn für $\theta_1 \neq \theta_2$ gilt, dass $\Sigma_{\theta_1} = \Sigma_{\theta_2}$, so heißen die Kovarianzstruktur des Modells und θ *nicht identifizierbar*.

Bemerkungen

- Normalverteilungsmodelle können durch zusätzliche Parameterannahmen identifizierbar gemacht werden.
- Der Erwartungswertparameter μ wird in der folgenden Kovarianzstrukturlogik ausgeklammert.
- Entscheidend ist die Anzahl der Parameter und der beobachteten Statistiken (Kovarianzen).

Theorem (Anzahl unikaler skalarer Parameter und Statistiken)

Gegeben sei ein Normalverteilungsmodell der Faktorenanalyse

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi). \tag{46}$$

Dann gilt für die Anzahl an unikalen skalaren Kovarianzstrukturparametern des Modells

$$n_\theta = mp + \frac{p(p+1)}{2} + m \tag{47}$$

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

$$n_c = \frac{m(m+1)}{2}. \tag{48}$$

Bemerkungen

- n_θ ist die Dimension des unikalen kovarianzstrukturellen 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.
- Der Erwartungswertparameter μ ist kein kovarianzstruktureller Parameter und wird hier nicht berücksichtigt.

Beweis

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

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

$$\frac{p^2 - p}{2} + p = \frac{p^2 - p}{2} + \frac{2p}{2} = \frac{p^2 + p}{2} = \frac{p(p+1)}{2}. \tag{49}$$

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

Für die Anzahl an unikalen skalaren Kovarianzstrukturparametern des Normalverteilungsmodells der Faktorenanalyse ergibt sich also zusammenfassend

$$n_\theta = mp + \frac{p(p+1)}{2} + m. \tag{50}$$

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

$$n_c = \frac{m(m+1)}{2}. \tag{51}$$

Definition (Ordnungsbedingung)

Gegeben sei ein Normalverteilungsmodell der Faktorenanalyse

$$y = \mu + Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{52}$$

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

$$n_\theta \leq n_c. \tag{53}$$

Bemerkungen

- Softwarelösungen implementieren die Ordnungsbedingung meist per default (vgl. Rosseel (2021)).
- Dazu werden typischerweise Einträge von L auf 0 bzw. 1 restringiert, um n_θ zu reduzieren.
- Restringierte Einträge von L werden dann als bekannt vorausgesetzt und nicht geschätzt.
- θ enthält dann nur die Einträge von L, Φ und Ψ, die nicht auf 0 bzw. 1 restringiert sind.

Einführung

Faktorenanalysemodelle

Parameterschätzung

Modellvergleichskriterien

Selbstkontrollfragen

Mittelwertzentrierung

- Vollständige Faktorenanalysemodelle enthalten den Erwartungswertparameter μ,

$$y = \mu + Lx + \varepsilon. \tag{54}$$

- In der Anwendung werden die Daten vor der kovarianzbasierten Faktorenanalyse typischerweise mittelwertzentriert

$$\bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i \text{ und } \tilde{y}_i := y_i - \bar{y}. \tag{55}$$

- Die Schätzung von L, Φ und Ψ basiert ausschließlich auf der Stichprobenkovarianzmatrix.
- Die Berechnung Stichprobenkovarianzmatrix impliziert die Mittelwertzentrierung der Daten, da

$$C = \frac{1}{n-1} Y \left(I_n - \frac{1}{n} 1_{nn} \right) Y^T = \frac{1}{n-1} \tilde{Y}\tilde{Y}^T \text{ mit } \tilde{Y} := Y - \bar{y}1_n^T. \tag{56}$$

- Das Analysemodell für die zentrierten Daten hat also die Form

$$\tilde{y} = y - \bar{y} = Lx + \varepsilon. \tag{57}$$

- Wir setzen im Folgenden die Mittelwertzentrierung der Daten y voraus und schreiben vereinfachend

$$y = Lx + \varepsilon. \tag{58}$$

Parameterschätzung

Überblick

- Die kovarianzstrukturellen Parameter des Faktorenanalysemodells sind L und Ψ (und Φ)
- Die Schätzung dieser Parameter geht von einer vordefinierten Anzahl p von Faktoren aus
- In der psychologischen Literatur wird die Parameterschätzung 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 Thurstone (1935)
- Die Zielsetzung einer *Hauptkomponentenanalyse* 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 Maximierung der Log-Likelihood-Funktion des Faktorenanalyse-Normalverteilungsmodells
- `fa()` mit `fm = "ml"` im `psych` R Paket
- `cfa()` mit `estimator = "ML"` im `lavaan` R Paket

Parameterschätzung

Motivation der Maximum-Likelihood-Schätzung

- Hauptkomponenten- und Hauptachsenfaktorisierungsschätzer sind *ad hoc* 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 mittelwertzentrierter Daten

$$y = Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{59}$$

mit Parametervektor

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{60}$$

und marginaler Datenkovarianzmatrix

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{61}$$

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

$$S := \frac{n-1}{n} C \tag{62}$$

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

$$\ell : \Theta \to \mathbb{R}, \theta \mapsto \ell(\theta) := -\frac{n}{2} \ln |\Sigma_\theta| - \frac{n}{2} \text{tr}\left(S\Sigma_\theta^{-1}\right) - \frac{nm}{2} \ln(2\pi) \tag{63}$$

und eine Maximalstelle von ℓ, also ein Wert

$$\hat{\theta}_{\text{ML}} = \text{argmax}_{\theta \in \Theta} \ell(\theta), \tag{64}$$

heißt *Maximum-Likelihood-Schätzer* für θ.

Beweis

Mit der marginalen Datenverteilung des Normalverteilungsmodells der Faktorenanalyse bei Unabhängigkeit der Datenbeobachtungen $y_1, ..., y_n$ kann die Log-Likelihood-Funktion von Y geschrieben werden als

$$
\begin{aligned}
\ell(\theta) &= \ln\left(\prod_{i=1}^{n} p_\theta(y_i)\right) \\
&= \ln\left(\prod_{i=1}^{n} (2\pi)^{-m/2}|\Sigma_\theta|^{-1/2}\exp\left(-\frac{1}{2}(y_i - 0_m)^T\Sigma_\theta^{-1}(y_i - 0_m)\right)\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n} y_i^T\Sigma_\theta^{-1}y_i \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y} + \bar{y})^T\Sigma_\theta^{-1}(y_i - \bar{y} + \bar{y}) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^T\Sigma_\theta^{-1}(y_i - \bar{y}) - \frac{1}{2}\sum_{i=1}^{n}\bar{y}^T\Sigma_\theta^{-1}\bar{y} - \sum_{i=1}^{n}(y_i - \bar{y})^T\Sigma_\theta^{-1}\bar{y} \qquad (65) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^T\Sigma_\theta^{-1}(y_i - \bar{y}) - \frac{1}{2}\sum_{i=1}^{n}\bar{y}^T\Sigma_\theta^{-1}\bar{y} - \left(\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i'=1}^{n} y_{i'}\right)^T\right)\Sigma_\theta^{-1}\bar{y} \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^T\Sigma_\theta^{-1}(y_i - \bar{y}) - \frac{1}{2}\sum_{i=1}^{n}\bar{y}^T\Sigma_\theta^{-1}\bar{y} - \left(\sum_{i=1}^{n} y_i^T - \frac{n}{n}\sum_{i=1}^{n} y_i^T\right)\Sigma_\theta^{-1}\bar{y} \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^T\Sigma_\theta^{-1}(y_i - \bar{y}) - \frac{1}{2}\sum_{i=1}^{n}\bar{y}^T\Sigma_\theta^{-1}\bar{y} - \left(0_m^T\Sigma_\theta^{-1}\bar{y}\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^T\Sigma_\theta^{-1}(y_i - \bar{y}) - \frac{1}{2}\sum_{i=1}^{n}\bar{y}^T\Sigma_\theta^{-1}\bar{y}
\end{aligned}
$$

Beweis (fortgeführt)

Mit der Mittelwertzentrierung der Daten $\bar{y} = 0_m$, der Definition von S sowie den Matrixspureigenschaften

$$\mathsf{tr}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{tr}\left(A_i\right), \mathsf{tr}\left(AB\right) = \mathsf{tr}\left(BA\right), \mathsf{tr}(a) = a \text{ und } \mathsf{tr}(cA) = c\,\mathsf{tr}(A) \text{ für } A \in \mathbb{R}^{m\times m}, a, c \in \mathbb{R}, \tag{66}$$

gilt dann weiterhin

$$\begin{aligned}
\ell(\theta) &= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}(y_i - 0_m)^T\Sigma_\theta^{-1}(y_i - 0_m) - \frac{1}{2}\sum_{i=1}^{n} 0_m^T\Sigma_\theta^{-1}0_m \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n} y_i^T\Sigma_\theta^{-1}y_i \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\mathsf{tr}\left(\sum_{i=1}^{n} y_i^T\Sigma_\theta^{-1}y_i\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}\mathsf{tr}\left(y_i^T\Sigma_\theta^{-1}y_i\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\sum_{i=1}^{n}\mathsf{tr}\left(\Sigma_\theta^{-1}y_i y_i^T\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\mathsf{tr}\left(\Sigma_\theta^{-1}\sum_{i=1}^{n} y_i y_i^T\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{1}{2}\mathsf{tr}\left(\Sigma_\theta^{-1}nS\right) \\
&= -\frac{mn}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma_\theta| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_\theta^{-1}\right).
\end{aligned} \tag{67}$$

□

Definition (Diskrepanzfunktion der Faktorenanalyse)

Gegeben sei das Normalverteilungsmodell der Faktorenanalyse mittelwertzentrierter Daten

$$y = Lx + \varepsilon \text{ mit } x \sim N(0_p, \Phi) \text{ und } \varepsilon \sim N(0_m, \Psi) \tag{68}$$

mit Parametervektor

$$\theta := \text{vec}(L, \Phi, \Psi), \tag{69}$$

und marginaler Datenkovarianzmatrix

$$\Sigma_\theta := L\Phi L^T + \Psi. \tag{70}$$

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

$$F : \Theta \to \mathbb{R}, \theta \mapsto F(\theta) := \ln|\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln|S| - m \tag{71}$$

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

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} F(\theta), \tag{72}$$

also eine Minimalstelle von F, ein *diskrepanzfunktionsbasierter Faktorenanalyseparameterschätzer*.

Bemerkungen

- Rosseel & Vidal (2025) geben folgende intuitive Interpretation der Diskrepanzfunktion:
- Wenn das Modell optimal ist, wenn also gilt, dass $\Sigma_\theta = S$, dann gilt

$$\begin{aligned} F(\theta) &= \ln|\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln|S| - m \\ &= \ln|S| + \text{tr}\left(SS^{-1}\right) - \ln|S| - m \\ &= \text{tr}\left(I_m\right) - m \\ &= m - m \\ &= 0 \end{aligned} \tag{73}$$

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

Theorem (Diskrepanzfunktionsbasierte und Maximum-Likelihood-Schätzer)

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

Bemerkungen

- Minimierung der Diskrepanzfunktion und Maximierung der Likelihood-Funktion liefern die gleichen Schätzer.
- Diskrepanzfunktionsminimierung ist in der Faktorenanalyse etwas populärer als Likelihood-Maximierung.
- 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

$$a := \frac{n}{2} \text{ und } b := -\frac{n}{2} \ln |S| - \frac{n}{2} m - \frac{nm}{2} \ln(2\pi) \tag{74}$$

gilt, dass

$$\begin{aligned} a\left(-F(\theta)\right) + b &= -\frac{n}{2} \left(\ln |\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln |S| - m \right) - \frac{n}{2} \ln |S| - \frac{n}{2} m - \frac{nm}{2} \ln(2\pi) \\ &= -\frac{n}{2} \ln |\Sigma_\theta| - \frac{n}{2} \text{tr}\left(S\Sigma_\theta^{-1}\right) + \frac{n}{2} \ln |S| + \frac{n}{2} m - \frac{n}{2} \ln |S| - \frac{n}{2} m - \frac{nm}{2} \ln(2\pi) \\ &= -\frac{n}{2} \ln |\Sigma_\theta| - \frac{n}{2} \text{tr}\left(S\Sigma_\theta^{-1}\right) - \frac{nm}{2} \ln(2\pi) \\ &= \ell(\theta). \end{aligned} \tag{75}$$

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

□

Anwendungsbeispiel

Als Beispiel für die Maximum-Likelihood-Schätzung eines Normalverteilungsmodells der Faktorenanalyse folgen wir Rosseel & Vidal (2025) und betrachten den Holzinger-Swineford-Datensatz mit den 9 Indikatoren `y_1` bis `y_9`. Es gelte also $m = 9$ und $C \in \mathbb{R}^{9 \times 9}$. Ein naives Faktorenanalyse-Normalverteilungsmodell für die mittelwertzentrierten Daten mit $p = 3$ Faktoren, die durch die Gruppierung der Indikatoren in drei Fähigkeitsgruppen von je drei Indikatoren motiviert ist, könnte die folgende Form haben:

$$y = Lx + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \\ l_{41} & l_{42} & l_{43} \\ l_{51} & l_{52} & l_{53} \\ l_{61} & l_{62} & l_{63} \\ l_{71} & l_{72} & l_{73} \\ l_{81} & l_{82} & l_{83} \\ l_{91} & l_{92} & l_{93} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix} \tag{76}$$

mit

$$x \sim N(0_3, \Phi) \text{ und } \varepsilon \sim N(0_9, \Psi) \tag{77}$$

und

$$\Phi := \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix} \text{ und } \Psi := \text{diag}\left(\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8, \psi_9\right). \tag{78}$$

Parameterschätzung

Anwendungsbeispiel

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

Ein mögliches restringiertes mittelwertzentriertes Analysemodell für die 9 Indikatoren ist

$$y = Lx + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 0 & 0 \\ l_{31} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & l_{52} & 0 \\ 0 & l_{62} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & l_{83} \\ 0 & 0 & l_{93} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix} \tag{79}$$

mit

$$x \sim N(0_3, \Phi) \text{ und } \varepsilon \sim N(0_9, \Psi) \tag{80}$$

und

$$\Phi := \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix} \text{ und } \Psi := \text{diag}\left(\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8, \psi_9\right). \tag{81}$$

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

$$n_\theta = 6 + 6 + 9 = 21 < 45 = n_c \tag{82}$$

und die Ordnungsbedingung ist erfüllt.

Anwendungsbeispiel

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

$$F : \Theta \to \mathbb{R}, \theta \mapsto F(\theta) := \ln |\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln |S| - m \tag{83}$$

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

```
YT              = read.csv("3-Daten/3-fa-hs.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 (9)
n               = ncol(Y)                                  # Stichprobenumfang (301)
I_n             = diag(n)                                  # Einheitsmatrix I_n
J_n             = matrix(rep(1,n^2), nrow = n)             # 1_{nn}
S               = (1/n)*(Y %*% (I_n-(1/n)*J_n) %*% t(Y))   # ML Stichprobenkovarianzmatrix
round(S,3)                                                 # Ausgabe der Stichprobenkovarianzmatrix
```

```
      y1     y2    y3    y4    y5    y6     y7    y8    y9
y1 1.358  0.407 0.580 0.505 0.441 0.455  0.085 0.264 0.458
y2 0.407  1.382 0.451 0.209 0.211 0.248 -0.097 0.110 0.244
y3 0.580  0.451 1.275 0.208 0.112 0.244  0.088 0.212 0.374
y4 0.505  0.209 0.208 1.351 1.098 0.896  0.220 0.126 0.243
y5 0.441  0.211 0.112 1.098 1.660 1.015  0.143 0.181 0.295
y6 0.455  0.248 0.244 0.896 1.015 1.196  0.144 0.165 0.236
y7 0.085 -0.097 0.088 0.220 0.143 0.144  1.183 0.535 0.373
y8 0.264  0.110 0.212 0.126 0.181 0.165  0.535 1.022 0.457
y9 0.458  0.244 0.374 0.243 0.295 0.236  0.373 0.457 1.015
```

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.
# ------------------------------------------------------------------
L   = matrix(c(1          , 0         , 0,                          # Faktorladungsmatrix
               theta[1]   , 0         , 0,
               theta[2]   , 0         , 0,
               0          , 1         , 0,
               0          , theta[3]  , 0,
               0          , theta[4]  , 0,
               0          , 0         , 1,
               0          , 0         , theta[5],
               0          , 0         , theta[6]),
               nrow = 9, ncol = 3, byrow = TRUE)

Phi = matrix(c(theta[7]   , theta[8] , theta[9],                     # Faktorkovarianzmatrix
               theta[8] , theta[10], theta[11],
               theta[9] , theta[11], theta[12]),
               nrow = 3, ncol = 3, byrow = TRUE)

Psi = diag(theta[13:21])                                             # Fehlerkovarianzmatrix
return(list(L= L, Phi = Phi, Psi = Psi))}                            # Listenausgabe
```

```
Sgm = function(theta){

# Funktion zur Konstruktion der marginalen Kovarianzmatrix der
# Parameter für das betrachtete Anwendungsbeispiel.
# ------------------------------------------------------------------
M       = mat(theta)                                                 # Matrizenform des Parametervektors
Sigma   = M$L %*% M$Phi %*% t(M$L) + M$Psi}                          # Marginale Datenkovarianzmatrix
```

Anwendungsbeispiel

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

```
theta_0     = c(1,1,1,1,1,1,                    # Startwert für \theta (L-Anteil)
                1,0,0,1,0,1,                    # Startwert für \theta (Phi-Anteil)
                rep(1,9))                       # Startwert für \theta (Psi-Anteil)
Sigma_0     = Sgm(theta_0)                      # Startwert für \Sigma_theta
round(Sigma_0,2)                                # Ausgabe
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,]    2    1    1    0    0    0    0    0    0
 [2,]    1    2    1    0    0    0    0    0    0
 [3,]    1    1    2    0    0    0    0    0    0
 [4,]    0    0    0    2    1    1    0    0    0
 [5,]    0    0    0    1    2    1    0    0    0
 [6,]    0    0    0    1    1    2    0    0    0
 [7,]    0    0    0    0    0    0    2    1    1
 [8,]    0    0    0    0    0    0    1    2    1
 [9,]    0    0    0    0    0    0    1    1    2
```

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

Anwendungsbeispiel

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

```
F_theta = function(theta, S, m, logdet_S){

# Funktion zur Auswertung der Diskrepanzfunktion für das Faktoren-
# analyse Normalverteilungsmodell basierend auf einem gegebenen Wert
# des Parametervektors und einer gegebenen ML-Stichprobenkovarianz.
# ------------------------------------------------------------------
Sigma_theta = Sgm(theta)                                    # \Sigma_\theta
EA          = eigen(Sigma_theta, symmetric = TRUE)          # Check auf positive Definitheit
if (any(EA$values < 1e-6)){                                 # \lambda \approx 0
  return(1e9)}                                              # invalider F Wert
Sigma_i     = chol2inv(chol(Sigma_theta))                   # \Sigma_\theta^{-1}
logdet      = determinant(Sigma_theta, logarithm = TRUE)$modulus    # ln |\Sigma_\theta|
F_theta     = (as.numeric(logdet)                           # ln |\Sigma_\theta|
                + sum(diag(S%*% Sigma_i))                   # tr(S\Sigma_theta^{-1})
                - as.numeric(logdet_S)                      # ln |S|
                - m)                                        # m
return(F_theta)}                                            # Ausgabe
```

Parameterschätzung

Anwendungsbeispiel

```
fa_est = function(Y, theta_0, opts = list()){

# Diese Funktion schätzt die Parameter eines Faktorenanalyse-
# Normalverteilungsmodells durch Minimierung der Diskrepanz-
# funktion F mittels nlminb.
# ------------------------------------------------------------
# Datenstatistiken
m           = nrow(Y)                                    # Datendimension
n           = ncol(Y)                                    # Stichprobenumfang
Y_c         = Y - rowMeans(Y)                            # Zentrierte Daten
S           = (1/n)*(Y_c %*% t(Y_c))                     # ML Kovarianz
logdet_S    = determinant(S, logarithm = TRUE)$modulus   # ln |S|

# ML Parameterschätzung
min         = nlminb(                                    # Optimierung
start       = theta_0,                                   # Startwert
objective   = F_theta,                                   # Zielfunktion
S           = S,                                         # ML Kovarianz
m           = m,                                         # Dimension
logdet_S    = logdet_S,                                  # ln |S|
control     = opts)                                      # Optionen
theta_hat   = min$par                                    # \hat{\theta}
M_hat       = mat(theta_hat)                             # Matrixform
Sgm_hat     = Sgm(theta_hat)                             # \Sigma_\hat{\theta}
F_theta_hat = F_theta(theta_hat, S, m, logdet_S)         # F(\hat{\theta})
nu          = (m*(m+1))/2 - length(theta_0)              # Freiheitsgradparameter
Chi_sq      = n*F_theta_hat                              # \chi^2
p           = 1 - pchisq(Chi_sq, nu)                     # p-Wert

# Listenausgabe
return(list(
theta_hat   = theta_hat,                                 # \hat{\theta}
L_hat       = M_hat$L,                                   # \hat{L}
Phi_hat     = M_hat$Phi,                                 # \hat{\Phi}
Psi_hat     = M_hat$Psi,                                 # \hat{\Psi}
Sgm_hat     = Sgm_hat,                                   # \Sigma_\hat{\theta}
F_theta_hat = F_theta_hat,                               # F(\hat{\theta})
Chi_sq      = Chi_sq,                                    # \chi^2
nu          = nu,                                        # Freiheitsgradparameter
p           = p,                                         # p-Wert
min         = min))}                                     # nlminb Ausgabe
```

Parameterschätzung

Anwendungsbeispiel

Zur Minimierung von F bezüglich θ nutzen wir `fa_est()`.

```
fa_hat = fa_est(Y, theta_0)                                    # ML Parameterschätzung
fa_hat$min                                                     # Ausgabe der Optimierung
```

```
$par
 [1] 0.5535003 0.7293709 1.1130765 0.9261463 1.1799498 1.0815298 0.8093154
 [8] 0.4082323 0.2622247 0.9794913 0.1734948 0.3837481 0.5490545 1.1338397
[15] 0.8443242 0.3711729 0.4462553 0.3562026 0.7993914 0.4876974 0.5661313

$objective
[1] 0.283407

$convergence
[1] 0

$iterations
[1] 29

$evaluations
function gradient
      34      841

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

Parameterschätzung

Anwendungsbeispiel

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

```
L_hat =

       [,1]  [,2]  [,3]
 [1,] 1.000 0.000 0.000
 [2,] 0.554 0.000 0.000
 [3,] 0.729 0.000 0.000
 [4,] 0.000 1.000 0.000
 [5,] 0.000 1.113 0.000
 [6,] 0.000 0.926 0.000
 [7,] 0.000 0.000 1.000
 [8,] 0.000 0.000 1.180
 [9,] 0.000 0.000 1.082

Phi_hat =

      [,1]  [,2]  [,3]
[1,] 0.809 0.408 0.262
[2,] 0.408 0.979 0.173
[3,] 0.262 0.173 0.384

Psi_hat =

       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]
 [1,] 0.549 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
 [2,] 0.000 1.134 0.000 0.000 0.000 0.000 0.000 0.000 0.000
 [3,] 0.000 0.000 0.844 0.000 0.000 0.000 0.000 0.000 0.000
 [4,] 0.000 0.000 0.000 0.371 0.000 0.000 0.000 0.000 0.000
 [5,] 0.000 0.000 0.000 0.000 0.446 0.000 0.000 0.000 0.000
 [6,] 0.000 0.000 0.000 0.000 0.000 0.356 0.000 0.000 0.000
 [7,] 0.000 0.000 0.000 0.000 0.000 0.000 0.799 0.000 0.000
 [8,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.488 0.000
 [9,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.566
```

Parameterschätzung

Anwendungsbeispiel

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

```
library(lavaan)                                                  # Lavaan Paket
YT              = read.csv("3-Daten/3-fa-hs.csv")                # Y^T \in \mathbb{R}^{n x m}
m               = ncol(YT)                                       # Datendimension
n               = nrow(YT)                                       # Stichprobenumfang
fam             = 'visual  =~ y1 + y2 + y3
                   textual =~ y4 + y5 + y6
                   speed   =~ y7 + y8 + y9'                      # Spezifikation des restringierten Modells
fam             = cfa(fam, data = YT)                            # CFA Faktorenanalysemodelle 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 Faktorkovarianzmatrix
Psi_hat_l       = theta_hat_l$theta                              # Geschätzte Fehlerkovarianzmatrix
```

Parameterschätzung

Anwendungsbeispiel

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

```
L_hat_l =

   visual textul speed
y1  1.000  0.000 0.000
y2  0.554  0.000 0.000
y3  0.729  0.000 0.000
y4  0.000  1.000 0.000
y5  0.000  1.113 0.000
y6  0.000  0.926 0.000
y7  0.000  0.000 1.000
y8  0.000  0.000 1.180
y9  0.000  0.000 1.082

Phi_hat_l =

        visual textul speed
visual   0.809
textual  0.408  0.979
speed    0.262  0.173 0.384

Psi_hat_l =

      y1    y2    y3    y4    y5    y6    y7    y8    y9
y1 0.549
y2 0.000 1.134
y3 0.000 0.000 0.844
y4 0.000 0.000 0.000 0.371
y5 0.000 0.000 0.000 0.000 0.446
y6 0.000 0.000 0.000 0.000 0.000 0.356
y7 0.000 0.000 0.000 0.000 0.000 0.000 0.799
y8 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.488
y9 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.566
```

Einführung

Faktorenanalysemodelle

Parameterschätzung

Modellvergleichskriterien

Selbstkontrollfragen

Modellvergleichskriterien

Wie viele Faktoren sollen zur Modellierung der Daten herangezogen werden?

Populäre Kriterien zur Wahl der Faktorenanzahl in der Anwendungsliteratur

Based on Fabrigar et al. (1999)
Summary Information of Current Practices in the Use of Factor Analysis

	JPSP		JAP	
Variable	***N***	**%**	***N***	**%**
Sample size				
100 or less	30	18.9	8	13.8
101–200	44	27.7	14	24.1
201–300	25	15.7	9	15.5
301–400	13	8.2	2	3.4
More than 400	47	29.6	25	43.1
Ratio of variable to factors				
Less than 3:1	1	0.6	1	1.7
3:1	28	17.6	9	15.5
4:1	26	16.4	10	17.2
5:1	14	8.8	10	17.2
6:1	13	8.2	6	10.3
More than 6:1	74	46.5	18	31.0
Unknown	2	1.3	4	6.9
Type of analysis				
Principal components	84	52.8	28	48.3
Common factors	31	19.5	13	22.4
Multiple methods	8	5.0	2	3.4
Other	1	0.6	0	0.0
Unknown	35	22.0	15	25.9
Factor–component number procedure				
Eigenvalue > 1.0	25	15.7	11	19.0
Scree test	24	15.1	9	15.5
Parallel analysis	1	0.6	0	0.0
Model fit	0	0.0	0	0.0
Theory	2	1.3	4	6.9
Interpretability	4	2.5	0	0.0
Multiple methods	35	22.0	12	20.7
Other	2	1.3	0	0.0
Unknown	66	41.5	22	37.9
Factor–component rotation				
Varimax	87	54.7	28	48.3
Harris–Kaiser	1	0.6	1	1.7
Promax	2	1.3	2	3.4
Direct quartimin	21	13.2	9	15.5
No rotation	23	14.5	4	6.9
Multiple methods	3	1.9	2	3.4
Other	1	0.6	0	0.0
Unknown	21	13.2	12	20.7

Note. Based on Fabrigar et al. (1999), Table 7. JPSP = Journal of Personality and Social Psychology; JAP = Journal of Applied Psychology.

Based on Goretzko et al. (2021)
Summary Information of Current Practices in the Use of Exploratory Factor Analysis

Variable	***N***	**%**
Sample size		
< 100	8	2.6
100–200	42	13.8
200–300	44	14.5
300–400	52	17.1
> 400	153	50.3
Not reported	5	1.6
Item to factor ratio		
< 3:1	10	3.3
3:1–5:1	102	33.6
5:1–7:1	47	15.5
> 7:1	113	37.2
Not reported	32	10.5
Minimum variables per factor		
< 3	32	10.5
3–5	141	46.4
> 5	35	11.5
Not reported	96	31.6
Extraction method		
Principal axis factoring	156	51.3
Maximum likelihood	50	16.4
Unweighted least squares	11	3.6
Weighted least squares	16	5.3
MinRes	3	1.0
Not reported	68	22.4
Factor retention criterion		
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
Rotation method		
Varimax	27	8.9
Promax	98	32.2
Oblimin	44	14.5
Geomin	23	7.6
Equimax	5	1.6
GeoMax	2	0.7
Varimax (oblique)	11	3.6
Other oblique rotations	34	11.2
Not reported	62	20.4

Note. Based on Goretzko et al. (2021), Tables 1, 2, 4, 5, and 6. Percentages for factor retention do not sum to 100 because multiple criteria could be used.

Modellvergleichskriterien

Klassifikation der Kriterien zur Wahl der Faktorenanzahl in der Anwendungsliteratur

Eigenwertbasierte Kriterien

⇒ Hauptkomponentenschätzung- und Explorative Faktorenanalyse-nahe Verfahren

- Scree-Test (Cattell (1966))
- Kaiser-Guttman-Kriterium (Guttman (1954), Kaiser (1960))

Diskrepanzbasierte Kriterien

⇒ Konfirmatorische Faktorenanalyse-nahe Verfahren

- χ^2-Test (Lawley (1940), Jöreskog (1967))
- Root Mean Square Error of Approximation (RMSEA) (Steiger & Lind (1980), Browne & Cudeck (1992))
- Tucker-Lewis Index (TLI) (Tucker & Lewis (1973))
- Comparative Fit Index (CFI) (Bentler (1990))

Likelihood-basierte Kriterien

⇒ Generische Modellklassen-übergreifende Verfahren

- Akaike Information Criterion (AIC) (Akaike (1974))
- Bayesian Information Criterion (BIC) (Schwarz (1978))
- Deviance Information Criterion (DIC) (Spiegelhalter et al. (2002))

Definition (Hypothesen der konfirmatorischen Faktorenanalyse)

Sei

$$M := \left\{ \Sigma_\theta = L\Phi L^T + \Psi \,\middle|\, \theta \in \Theta \right\} \tag{84}$$

ein restringiertes Faktorenanalysemodell. Weiterhin sei Σ die wahre, aber unbekannte Kovarianzmatrix unter festen Hypothesen und Σ_n die wahre, aber unbekannte Kovarianzmatrix unter lokalen Alternativen. Dann unterscheidet man folgende Hypothesen.

- (*Nullhypothese*) Das restringierte Modell gilt exakt.

$$H_0 : \Sigma \in M \Leftrightarrow \text{ Es gibt ein } \theta_0 \in \Theta, \text{ für das gilt } \Sigma = \Sigma_{\theta_0}. \tag{85}$$

- (*Feste Alternative*) Das restringierte Modell gilt nicht.

$$H_1 : \Sigma \notin M. \tag{86}$$

- (*Lokale Alternative*) Das restringierte Modell gilt nicht exakt, aber die Modellverletzung nimmt mit n ab.

$$H_{1,n} : \Sigma_n = \Sigma_{\theta_0} + \frac{\Delta}{\sqrt{n}} \text{ für ein } \theta_0 \in \Theta \text{ und } \Delta \in \mathbb{R}^{m \times m} \text{ mit } \Delta = \Delta^T, \quad \Delta \neq 0_{mm}. \tag{87}$$

  Weiterhin gelte

$$\Sigma_n \notin M \text{ ab einem hinreichend großen } n, \text{ aber } \Sigma_n \to \Sigma_{\theta_0} \in M \text{ für } n \to \infty. \tag{88}$$

Bemerkung

- Die Nullhypothese ist Grundlage des χ^2-Tests, die lokale Alternative ist Grundlage des RMSEA.
- Das Verwerfen der Nullhypothese entspricht dem Verwerfen des aktuell spezifizierten Modells.

Modellvergleichskriterien | χ^2-Test

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 M und $\bar{M}$.

(M) Ein Normalverteilungsmodell der Faktorenanalyse, das die Ordnungsrelation erfüllt und identifizierbar ist,

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

$(\bar{M})$ Ein *gesättigtes* multivariates Normalverteilungsmodell mit beliebigem Kovarianzmatrixparameter,

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

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

$$\Lambda := \ln \left(\frac{\max_{\theta \in \Theta} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta)}{\max_{\gamma \in \Gamma} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma)} \right) \qquad (89)$$

$\Lambda \in \mathbb{R}$ setzt also die maximierten Wahrscheinlichkeitsdichten von Y unter M und $\bar{M}$ ins Verhältnis. Weil M in $\bar{M}$ restringiert ist, gilt $\Lambda \leq 0$. Werte von Λ nahe Null bedeuten, dass Y unter M eine ähnlich große Wahrscheinlichkeitsdichte besitzt wie unter $\bar{M}$. Dies wird allgemein als Evidenz dafür verstanden, dass die Restriktionen von M mit Y vereinbar sind. Dabei ist Λ letztlich die zentrale Motivation für die funktionale Form der Diskrepanzfunktion (vgl. Lawley (1940)). Das Vorgehen ist hier also ähnlich dem eines varianzanalytischen F-Tests, allerdings unter Umkehrung der üblichen Interpretation von großen und kleinen Werten des Testkriteriums und bei Umkehrung der Interpretation von M und $\bar{M}$ als Hypothesen von Interesse.

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

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

$$\Lambda := \ln \left(\frac{\max_{\theta \in \Theta} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta)}{\max_{\gamma \in \Gamma} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma)} \right), \tag{90}$$

wobei

$$\Sigma_\theta = L\Phi L^T + \Psi, \theta = \text{vec}(L, \Phi, \Psi) \in \Theta, \Theta \subset \mathbb{R}^{n\theta} \text{ und } n_\theta \leq m(m+1)/2 \tag{91}$$

und

$$\Sigma_\gamma = \text{vec}^{-1}(\gamma), \gamma \in \Gamma \text{ und } \Gamma \subset \mathbb{R}^{m(m+1)/2} \tag{92}$$

seien. Weiterhin sei für die verzerrte Stichprobenkovarianzmatrix S von Y

$$F(\theta) := \ln |\Sigma_\theta| + \text{tr}\left(S\Sigma_\theta^{-1}\right) - \ln |S| - m \tag{93}$$

die Diskrepanzfunktion der Faktorenanalyse und $\hat{\theta}$ eine Minimalstelle von F. Dann gilt

$$-2\Lambda = nF(\hat{\theta}) \tag{94}$$

Bemerkungen

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

Beweis

Wir halten zunächst fest, dass

$$\max_{\theta \in \Theta} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\theta) = \prod_{i=1}^{n} N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) \tag{95}$$

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

$$\max_{\gamma \in \Gamma} \prod_{i=1}^{n} N(y_i; 0_m, \Sigma_\gamma) = \prod_{i=1}^{n} N(y_i; 0_m, S) \tag{96}$$

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

$$\Lambda = \ln \left(\frac{\prod_{i=1}^{n} N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right)}{\prod_{i=1}^{n} N(y_i; 0_m, S)} \right) = \sum_{i=1}^{n} \ln N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) - \sum_{i=1}^{n} \ln N(y_i; 0_m, S) \tag{97}$$

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

$$\begin{aligned}
\Lambda =& \sum_{i=1}^{n} \ln N\left(y_i; 0_m, \Sigma_{\hat{\theta}}\right) - \sum_{i=1}^{n} \ln N(y_i; 0_m, S) \\
=& -\frac{mn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma_{\hat{\theta}}| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) - \frac{n}{2}\bar{y}^T\Sigma_{\hat{\theta}}^{-1}\bar{y} \\
&+ \frac{mn}{2}\ln(2\pi) + \frac{n}{2}\ln|S| + \frac{n}{2}\mathsf{tr}\left(SS^{-1}\right) + \frac{n}{2}\bar{y}^T S^{-1}\bar{y} \\
=& -\frac{n}{2}\ln|\Sigma_{\hat{\theta}}| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) - \frac{n}{2}0_m^T\Sigma_{\hat{\theta}}^{-1}0_m \\
&+ \frac{n}{2}\ln|S| + \frac{n}{2}\mathsf{tr}\left(SS^{-1}\right) + \frac{n}{2}0_m^T S^{-1}0_m \\
=& -\frac{n}{2}\ln|\Sigma_{\hat{\theta}}| - \frac{n}{2}\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) + \frac{n}{2}\ln|S| + \frac{mn}{2}
\end{aligned} \tag{98}$$

Multiplikation mit -2 ergibt schließlich

$$-2\Lambda = n\ln|\Sigma_{\hat{\theta}}| + n\,\mathsf{tr}\left(S\Sigma_{\hat{\theta}}^{-1}\right) - n\ln|S| - mn = nF(\hat{\theta}). \tag{99}$$

□

Theorem (X^2-Statistik)

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

$$\mathrm{X}^2 := nF(\hat{\theta}) \tag{100}$$

unter der Nullhypothese

$$H_0 : \Sigma \in M \Leftrightarrow \text{Es gibt ein } \theta_0 \in \Theta, \text{ für das gilt } \Sigma = \Sigma_{\theta_0} \tag{101}$$

dass für $\nu := n_c - n_\theta$ gilt:

$$\mathrm{X}^2 \sim \chi^2(\nu) \text{ für } n \to \infty \tag{102}$$

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 asymptotische χ^2-Verteilung allen Log-Likelihood-Ratio-Kriterien gemein ist.
- Wir validieren die asymptotische Verteilung für $n = 301$ untenstehend in einem Anwendungsbeispiel.
- ν wird *Freiheitsgradparameter* der χ^2-Verteilung genannt.
- n_c ist die Anzahl der unikalen skalaren Datenstatistiken, n_θ die der unikalen skalaren Modellparameter.
- Höhere Modellkomplexität, also mehr Modellparameter n_θ bei gleichem $n_c \Rightarrow \nu \downarrow$
- Geringere Modellkomplexität, also mehr Kovarianzmodellrestriktionen, bei gleichem $n_c \Rightarrow \nu \uparrow$

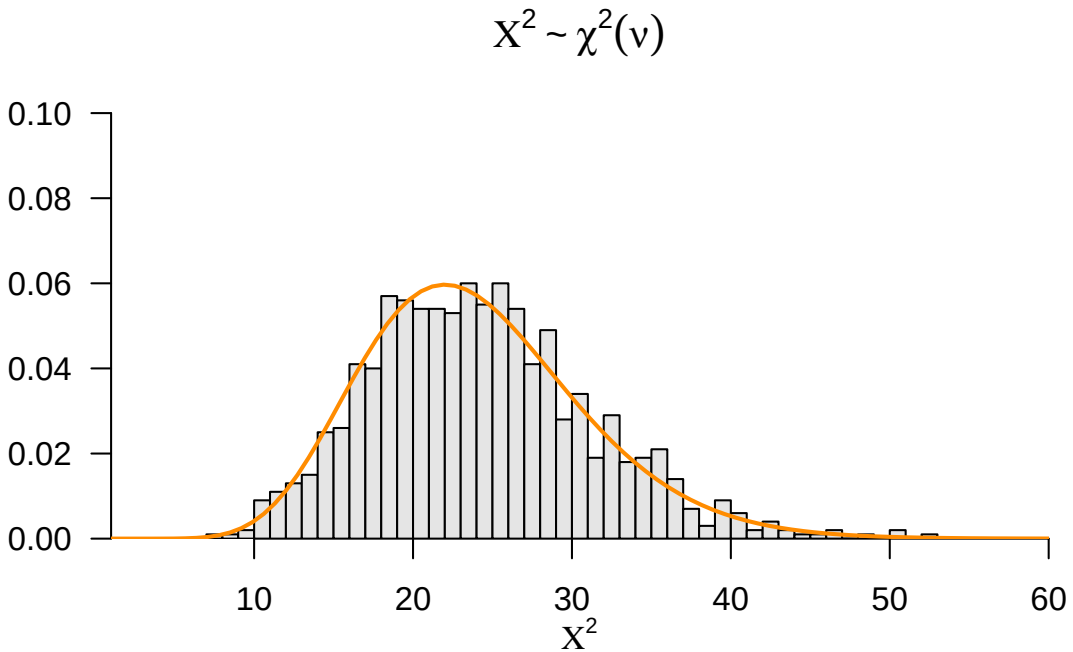
Datensimulation anhand des Normalverteilungsmodells der Faktorenanalyse

```
library(MASS)
fa_sim = function(theta, n){
  # Diese Funktion simuliert Daten aus einem Normalverteilungsmodell
  # der Faktorenanalyse.
  # Inputs
  #   theta : Parametervektor
  #   n     : Stichprobenumfang
  #
  # Output
  #   Y     : simulierte Datenmatrix \in \mathbb{R}^{m x n}
  # ------------------------------------------------------------
  Sigma_theta = Sgm(theta)                                  # \Sigma_\theta
  m           = nrow(Sigma_theta)                           # Datendimension
  YT          = mvrnorm(                                    # multivariate Normalverteilung
  n           = n,                                          # Stichprobenumfang
  mu          = rep(0,m),                                   # Erwartungswertparameter
  Sigma       = Sigma_theta)                                # Kovarianzmatrixparameter
  Y           = t(YT)                                       # Y \in R^{m x n}
  rownames(Y) = paste("y", 1:m, sep = "")                   # Itemnamen
  colnames(Y) = 1:n                                         # Personenlabel
  return(Y)}                                                # Ausgabe
```

Validierung der X^2-Verteilung bei Gültigkeit von H_0

```
set.seed(0)                                          # Reproduzierbarkeit
theta       = theta_hat                              # Parameter
Sigma       = Sgm(theta)                             # Kovarianzmatrix
m           = nrow(Sigma)                            # Datendimension
n           = ncol(Y)                                # Stichprobenumfang
mu          = rep(0, m)                              # Erwartungswert
nsim        = 1e3                                    # Anzahl Simulationen
d           = (m*(m+1))/2 - length(theta)            # Freiheitsgradparameter
Chisq       = rep(NaN, nsim)                         # \chi^2 Werte
theta_0     = c(1,1,1,1,1,1,1,0,0,1,0,1,rep(1,m))    # lavaan starting point
opts        = list(iter.max = 50, eval.max = 100)    # Optimierungsoptionen
for(s in 1:nsim){                                    # Simulationen
  Y         = fa_sim(theta, n)                       # Datensatz
  fit       = fa_est(Y, theta_0, opts)               # ML Schätzung
  Chisq[s]  = fit$Chi_sq}                            # \chi^2 Wert
```

Validierung der X^2-Verteilung bei Gültigkeit von H_0



Modellvergleichskriterien | χ^2-Test

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

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

Zum Vergleich betrachten wir die `lavaan`-Ausgabe für das geschätzte Modell.

```
# Eigene Schätzung
F_theta_hat = fa_hat$F_theta_hat                # F(\hat{\theta})
Chi_square  = fa_hat$Chi_sq                     # Chi-Square-Teststatistik
nu          = fa_hat$nu                         # Freiheitsgradparameter
p           = fa_hat$p                          # p-Wert
```

Wir erhalten

```
Test statistic          85.306

Degrees of freedom      24

P-value (Chi-square)    0
```

- H_0 wird mit einem X^2-Wert von 85.306 (p = 0) abgelehnt.
- Das Modell weist nach dem χ^2-Test keine akzeptable Passung auf.

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 35 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        21

  Number of observations                           301

Model Test User Model:

  Test statistic                                85.306
  Degrees of freedom                                24
  P-value (Chi-square)                           0.000
```

Nachteile von X^2

- Für $\mathrm{X}^2 := nF(\hat{\theta})$ gilt bei positivem Approximationsfehler, dass $\mathrm{X}^2 \to \infty$ für $n \to \infty$.
- Für große Stichprobengrößen wird X^2 daher zunehmend sensitiv gegenüber kleinen Modellabweichungen.
- $\Rightarrow$ Das Modell kann auch bei praktisch geringem Approximationsfehler verworfen werden.
- Bei X^2 erfolgt keine explizite Normierung der Residualdiskrepanz bezüglich der Modellkomplexität.

Root-Mean-Square-Error-of-Approximation (RMSEA)

- Mit dem *Approximationsfehler* $\varphi(\Sigma_{\theta_*}, \Sigma)$ zwischen modellbasierter Σ_{θ_*} und wahrer, aber unbekannter, Σ gilt

$$\text{RMSEA} := \sqrt{\frac{\varphi(\Sigma_{\theta_*}, \Sigma)}{\nu}} \tag{103}$$

- Bei konstantem ν impliziert ein hoher Approximationsfehler einen hohen RMSEA.
- Hohe RMSEA-Werte bei konstantem ν implizieren eine schlechte Modellanpassung
- Bei konstantem $\varphi(\Sigma_{\theta_*}, \Sigma)$ impliziert eine geringe Modellkomplexität ein hohes ν, also einen geringen RMSEA.
- Geringe RMSEA-Werte implizieren einen geringen Approximationsfehler relativ zu ν.
- In aller Regel sinkt mit höherer Modellkomplexität neben ν aber natürlich auch $\varphi(\Sigma_{\theta_*}, \Sigma)$.
- Sinkt $\varphi(\Sigma_{\theta_*}, \Sigma)$ mehr als ν durch Modellkomplexitätszunahme, dann sinkt der RMSEA.

Allerdings:

- Da Σ unbekannt ist, ist $\varphi(\Sigma_{\theta_*}, \Sigma)$ nicht beobachtbar und muss aus den Daten geschätzt werden.
- RMSEA wird geschätzt durch

$$\widehat{\text{RMSEA}} = \sqrt{\frac{\max\left(\mathrm{X}^2 - \nu, 0\right)}{n\nu}} \tag{104}$$

- Im Folgenden wollen wir diese Schätzung motivieren.

Anwendungsbeispiel

Wir bestimmen den RMSEA-Schätzer und vergleichen ihn mit der `lavaan` Ausgabe.

```
# Eigene Schätzung
F_theta_hat = fa_hat$F_theta_hat                        # F(\hat{\theta})
Chi_square  = fa_hat$Chi_sq                             # Chi-Square-Teststatistik
nu          = fa_hat$nu                                 # Freiheitsgradparameter
n           = ncol(Y)                                   # Stichprobenumfang
RMSEA_hat   = sqrt(max(Chi_square - nu, 0)/(n*nu))      # RMSEA-Schätzung

# lavaan-Schätzung
RMSEA_l     = fitMeasures(fam, "rmsea")                 # lavaan RMSEA
```

Wir erhalten

```
RMSEA Schätzer         0.092

RMSEA Schätzer lavaan 0.092
```

Generell sagt man

- $0.00 \leq$ RMSEA $\leq 0.05 \Rightarrow$ "Gute Modellpassung"
- $0.05 <$ RMSEA $\leq 0.08 \Rightarrow$ "Akzeptable Modellpassung"
- $0.08 <$ RMSEA $\leq 0.10 \Rightarrow$ "Mäßige Modellpassung"
- $0.10 <$ RMSEA $< \infty \quad \Rightarrow$ "Schlechte Modellpassung"

Definition (Approximationsfehler, Schätzfehler, Gesamtfehler)

Gegeben sei ein Faktorenanalysemodell mit der Menge der durch das Modell darstellbaren Kovarianzmatrizen

$$M := \left\{ \Sigma_\theta \mid \Sigma_\theta = L\Phi L^T + \Psi \text{ und } \theta \in \Theta \right\}. \tag{105}$$

Weiterhin sei

$$\varphi : \mathbb{S}^{m\times m} \times \mathbb{S}^{m\times m} \to \mathbb{R}_{\geq 0}, \quad (A, B) \mapsto \varphi(A, B) \tag{106}$$

eine *Diskrepanzfunktion* zweier positiv semidefiniter Matrizen in $\mathbb{S}^{m\times m}$ mit der Eigenschaft, dass

$$A = B \Leftrightarrow \varphi(A, B) = 0. \tag{107}$$

Schließlich seien Σ die wahre, aber unbekannte Kovarianzmatrix, S die Stichprobenkovarianzmatrix und

$$\theta_* \in \text{argmin}_{\theta\in\Theta}\varphi(\Sigma_\theta, \Sigma), \qquad \hat{\theta} \in \text{argmin}_{\theta\in\Theta}\varphi(\Sigma_\theta, S). \tag{108}$$

Dann werden

$$\varphi(\Sigma_{\theta_*}, \Sigma) \text{ als } \textit{Approximationsfehler}, \varphi(\Sigma_{\hat{\theta}}, \Sigma_{\theta_*}) \text{ als } \textit{Schätzfehler} \text{ und } \varphi(\Sigma_{\hat{\theta}}, \Sigma) \text{ als } \textit{Gesamtfehler} \tag{109}$$

bezeichnet.

Bemerkungen

- $\varphi(\Sigma_{\theta_*}, \Sigma)$ misst die Diskrepanz zwischen wahrer, aber unbekannter Σ und einer optimierten Σ_{θ_*} aus M.
- $\varphi(\Sigma_{\hat{\theta}}, \Sigma_{\theta_*})$ misst die Diskrepanz zwischen MLE $\Sigma_{\hat{\theta}}$ und einer optimierten Σ_{θ_*} aus M.
- $\varphi(\Sigma_{\hat{\theta}}, \Sigma)$ misst die Diskrepanz zwischen MLE $\Sigma_{\hat{\theta}}$ und wahrer, aber unbekannter Σ.
- Keine der drei Fehlerkomponenten ist beobachtbar, da Σ unbekannt ist.
- Die Begriffe sind terminologisch, eine additive Zerlegung des Gesamtfehlers gilt im Allgemeinen nicht.

Theorem (X^2-Statistik unter lokaler Alternative)

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

$$\mathrm{X}_n^2 := nF(\hat{\theta}_n) \tag{110}$$

unter der lokalen Alternative

$$H_{1,n} : \Sigma_n = \Sigma_{\theta_0} + \frac{\Delta}{\sqrt{n}}, \tag{111}$$

dass

$$\mathrm{X}_n^2 \overset{a}{\sim} \chi^2(\delta, \nu) \quad \text{für } n \to \infty \text{ mit } \nu := n_c - n_\theta \tag{112}$$

Bezeichne weiterhin

$$\theta_{*,n} \in \operatorname{argmin}_{\theta \in \Theta} \varphi(\Sigma_\theta, \Sigma_n) \tag{113}$$

den zur wahren Kovarianzmatrix Σ_n nächstgelegenen Modellparameter. Dann bezeichnet

$$\delta := \lim_{n \to \infty} n\varphi(\Sigma_{\theta_{*,n}}, \Sigma_n) \tag{114}$$

den Nichtzentralitätsparameter der asymptotischen Verteilung.

Bemerkungen

- Für einen Beweis verweisen wir auf Browne (1984), Satorra & Saris (1985), Steiger et al. (1985).
- Unter H_0 ergibt sich als Spezialfall $\delta = 0$ und damit die zentrale χ^2-Verteilung.
- δ ist der Grenzwert des mit n skalierten Approximationsfehlers unter lokaler Fehlspezifikation.
- Diese Interpretation von δ bildet die Grundlage des Approximationsfehlerschätzers.

Motivation des RMSEA-Schätzers

Nach obigem Theorem gilt approximativ für große n unter der lokalen Alternative

$$\mathrm{X}^2 \overset{a}{\sim} \chi^2(\delta, \nu). \tag{115}$$

Mit dem Erwartungswert der nichtzentralen χ^2-Verteilung gilt dann

$$\mathbb{E}(\mathrm{X}^2) = \nu + \delta. \tag{116}$$

Betrachtet man nun eine Realisierung von X^2 als "Momentenschätzer" für $\mathbb{E}(\mathrm{X}^2)$, so ergibt sich

$$\mathrm{X}^2 \approx \nu + \delta \Leftrightarrow \delta \approx \mathrm{X}^2 - \nu. \tag{117}$$

Als zensierter Momentenschätzer für δ ergibt sich daher

$$\hat{\delta} := \max\left(\mathrm{X}^2 - \nu, 0\right). \tag{118}$$

Gleichzeitig gilt nach dem vorherigen Theorem

$$\delta = \lim_{n \to \infty} n\varphi(\Sigma_{\theta_*,n}, \Sigma_n). \tag{119}$$

Für große Stichproben kann daher heuristisch

$$\delta \approx n\varphi(\Sigma_{\theta_*}, \Sigma) \tag{120}$$

interpretiert werden, sodass

$$\hat{\varphi}(\Sigma_{\theta_*}, \Sigma) = \max\left(\frac{\mathrm{X}^2 - \nu}{n}, 0\right) \tag{121}$$

als zensierter Momentenschätzer des Approximationsfehlers motiviert ist.

Motivation des RMSEA-Schätzers

Einsetzen dieses Schätzers in die RMSEA-Definition ergibt dann

$$\widehat{\text{RMSEA}} = \sqrt{\frac{\hat{\delta}}{n\nu}} = \sqrt{\frac{\max\left(\text{X}^2 - \nu, 0\right)}{n\nu}}. \tag{122}$$

Bemerkungen

- Die asymptotische Theorie des RMSEA basiert auf lokalen Alternativen der Form $\Sigma_n = \Sigma_{\theta_0} + \Delta/\sqrt{n}$. Solche lokalen Alternativen beschreiben Modellverletzungen, die mit wachsendem Stichprobenumfang kleiner werden und gewährleisten eine nichttriviale Grenzverteilung $\text{X}^2_n \overset{a}{\sim} \chi^2(\delta, \nu)$. Unter dieser Annahme schätzt $\text{X}^2_n - \nu$ im Momentensinn den Nichtzentralitätsparameter δ, sodass sich der RMSEA als Schätzer eines normierten Approximationsfehlers motivieren lässt.
- Die Annahme lokaler Alternativen ist jedoch primär ein mathematisches Hilfskonstrukt zur asymptotischen Analyse. In empirischen Anwendungen wächst typischerweise die Stichprobe, nicht aber die Nähe der Populationskovarianzmatrix zum postulierten Modell. Die nichtzentrale χ^2-Theorie liefert daher eine theoretische Grundlage für die Konstruktion des RMSEA, begründet aber nicht unmittelbar die Interpretation konkreter RMSEA-Werte in realen Anwendungen.
- Aus theoretischer Sicht entsteht damit eine gewisse Spannung: Die Motivation des RMSEA beruht auf einem asymptotischen Szenario, das in der Praxis selten plausibel erscheint, während die praktische Verwendung des RMSEA meist gerade für feste, von der Stichprobengröße unabhängige Modellabweichungen erfolgt.
- Die breite Verwendung des RMSEA beruht daher nicht allein auf seiner asymptotischen Herleitung, sondern auch auf empirischen Konventionen und der Erfahrung, dass das Maß in vielen Anwendungen nützliche Informationen über den Grad des Modellmissfits liefert.

Einführung

Faktorenanalysemodelle

Parameterschätzung

Modellvergleichskriterien

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 struktureller Form wieder.
6. Geben Sie das Theorem zu den Faktorenanalysemodellen in Verteilungsform wieder.
7. Erläutern Sie die datengenerative Sicht eines Faktorenanalysemodells.
8. Geben Sie die Definition der Datensatzmodelle der Faktorenanalyse wieder.
9. Geben Sie das Theorem zur marginalen Datenkovarianzmatrix des Faktorenanalysemodells wieder.
10. Geben Sie die Definition einer orthogonalen Matrix wieder.
11. Geben Sie die Definition der orthogonalen Transformation des Faktorenanalyse-Normalverteilungsmodells wieder.
12. Geben Sie das Theorem zur Rotationsindeterminiertheit und Kovarianzinvarianz wieder.
13. Geben Sie die Definition eines identifizierbaren Faktorenanalyse-Normalverteilungsmodells wieder.
14. Geben Sie das Theorem zur Anzahl unikaler Parameter eines Faktorenanalyse-Normalverteilungsmodells wieder.
15. Geben Sie die Definition der Ordnungsbedingung wieder.
16. Erläutern Sie die Motivation des Maximum-Likelihood-Ansatzes zur Schätzung eines Faktorenanalysemodells.
17. Geben Sie das Theorem zu den Maximum-Likelihood-Schätzern eines Faktorenanalysemodells wieder.
18. Geben Sie die Definition der Diskrepanzfunktion eines Faktorenanalysemodells wieder.
19. Erläutern Sie die Interpretation hoher und niedriger Werte der Diskrepanzfunktion.
20. Geben Sie das Theorem zur Äquivalenz von Diskrepanzfunktionsbasierten und ML-Schätzern wieder.
21. Erläutern Sie die Motivation des χ^2-Tests der Faktorenanalyse.
22. Geben Sie das Theorem zum Log-Likelihood-Ratio-Kriterium und zur Diskrepanzfunktion wieder.
23. Geben Sie das Theorem zur X^2-Statistik eines Faktorenanalysemodells wieder.
24. Erläutern Sie die Interpretation hoher und niedriger p-Werte der X^2-Statistik.
25. Erläutern Sie die Motivation und Interpretation des RMSEA.

Ergänzungen aus der Linearen Algebra

Definition (Nullmatrizen, Einsmatrizen)

- Wir bezeichnen *Nullmatrizen* mit

$$0_{nm} := (0)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 0_n := (0)_{1 \leq i \leq n} \in \mathbb{R}^n \tag{123}$$

- Wir bezeichnen die *Einsmatrizen* mit

$$1_{nm} := (1)_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{R}^{n \times m} \text{ und } 1_n := (1)_{1 \leq i \leq n} \in \mathbb{R}^n \tag{124}$$

Bemerkungen

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

$$0_n 0_n^T = 0_{nn} \text{ und } 1_n 1_n^T = 1_{nn}. \tag{125}$$

Definition (Einheitsmatrizen und Einheitsvektoren)

- Wir bezeichnen die *Einheitsmatrix* mit

$$I_n := (i_{jk})_{1 \leq j,k \leq n} \in \mathbb{R}^{n \times n} \text{ mit } i_{jk} = 1 \text{ für } j = k \text{ und } i_{jk} = 0 \text{ für } j \neq k \tag{126}$$

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

$$e_i := (e_{i_j})_{1 \leq j \leq n} \in \mathbb{R}^n \text{ mit } e_{i_j} = 1 \text{ für } i = j \text{ und } e_{i_j} = 0 \text{ für } i \neq j \tag{127}$$

Bemerkungen

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

$$I_n = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix} \tag{128}$$

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

$$e_i^T e_j = 0 \text{ für } i \neq j, e_i^T e_i = 1 \text{ und } e_i^T v = v^T e_i = v_i \text{ für } v \in \mathbb{R}^n. \tag{129}$$

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

$$D = \text{diag}(d_1, ..., d_n). \tag{130}$$

- 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

$$D := \text{diag}(d_1, ..., d_n) \Rightarrow |D| = \prod_{i=1}^{n} d_i \tag{131}$$

- 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.
- Beispielsweise gilt für die Summe zweier symmetrischer Matrizen, dass auch diese wieder symmetrisch ist

$$A = A^T \text{ und } B = B^T \Rightarrow A + B = (A + B)^T \tag{132}$$

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

$$S^T = S \Rightarrow \left(S^{-1}\right)^T = S^{-1}. \tag{133}$$

- 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

$$x^T C x > 0 \text{ für alle } x \in \mathbb{R}^n \text{ mit } x \neq 0_n. \tag{134}$$

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

$$x^T C x \geq 0 \text{ für alle } x \in \mathbb{R}^n \text{ mit } x \neq 0_n. \tag{135}$$

Wenn $C \in \mathbb{R}^{n \times n}$ positiv-semidefinit 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

$$Av = \lambda v \tag{136}$$

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 skaliert.
- 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 $c \in \mathbb{R}$ mit $c \neq 0$ auch $cv \in \mathbb{R}^m$ Eigenvektor von A und zwar wiederum mit Eigenwert $\lambda \in \mathbb{R}$.

Beweis

Es gilt

$$Av = \lambda v \Leftrightarrow cAv = c\lambda v \Leftrightarrow A(cv) = \lambda(cv) \tag{137}$$

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 $m \times m$ Matrix 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 verschiedene Eigenwerte von S mit zugehörigen Eigenvektoren q_i und q_j, respektive. Dann ergibt sich wie unten gezeigt, dass

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \tag{138}$$

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 es keine andere Zahl c als die Null gibt, für die bei $a, b \in \mathbb{R}$ und $a \neq b$ gilt, dass

$$ac = bc. \tag{139}$$

Um

$$\lambda_i q_i^T q_j = \lambda_j q_i^T q_j. \tag{140}$$

zu zeigen, halten wir zunächst fest, dass

$$Sq_i = \lambda_i q_i \Leftrightarrow (Sq_i)^T = (\lambda_i q_i)^T \Leftrightarrow q_i^T S^T = q_i^T \lambda_i^T \Leftrightarrow q_i^T S = q_i^T \lambda_i \Leftrightarrow q_i^T S q_j = \lambda_i q_i^T q_j \tag{141}$$

und

$$Sq_j = \lambda_j q_j \Leftrightarrow q_i^T S q_j = \lambda_j q_i^T q_j \tag{142}$$

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 kann S geschrieben werden als

$$S = Q \Lambda Q^T, \tag{143}$$

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

$$Q := \begin{pmatrix} q_1 & q_2 & \cdots & q_m \end{pmatrix} \in \mathbb{R}^{m \times m} \text{ und } \Lambda := \text{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_m\right) \in \mathbb{R}^{m \times m}, \tag{144}$$

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

$$S q_i = \lambda_i q_i \text{ für } i = 1, \ldots, m \Leftrightarrow SQ = Q\Lambda. \tag{145}$$

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

$$SQQ^T = Q\Lambda Q^T \Leftrightarrow SI_m = Q\Lambda Q^T \Leftrightarrow S = Q\Lambda Q^T. \tag{146}$$

□

Bemerkungen

- $S = Q\Lambda Q^T$ heißt auch *Diagonalisierung* von S.
- Man wählt als Diagonalelemente von Λ die der Größe nach geordneten Eigenwerte von S.
- Man wählt 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 verschiedenen Eigenwerten $\lambda_1, ..., \lambda_m$. Dann gelten

$$|S| = \prod_{i=1}^{m} \lambda_i \text{ und } \text{tr}(S) = \sum_{i=1}^{m} \lambda_i \tag{147}$$

Beweis

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

$$|S| = |Q \Lambda Q^T| \tag{148}$$

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

$$|S| = |Q \Lambda Q^T| = |Q||\Lambda||Q^T| = |\Lambda| = \prod_{i=1}^{m} \lambda_i. \tag{149}$$

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

$$\text{tr}(S) = \text{tr}\left(Q \Lambda Q^T\right). \tag{150}$$

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

$$\text{tr}(S) = \text{tr}\left(Q \Lambda Q^T\right) = \text{tr}\left(Q^T Q \Lambda\right) = \text{tr}(\Lambda) = \sum_{i=1}^{m} \lambda_i. \tag{151}$$

□

Ergänzungen aus der Wahrscheinlichkeitstheorie

Definition (χ^2-Zufallsvariable)

Sei X^2 eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p : \mathbb{R}_{>0} \to \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}x\right), \tag{152}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X^2 einer χ^2-Verteilung mit Freiheitsgradparameter n unterliegt und nennen X^2 eine χ^2-Zufallsvariable mit Freiheitsgradparameter n. Wir kürzen dies mit

$$\mathrm{X}^2 \sim \chi^2(n) \tag{153}$$

ab. Die WDF einer χ^2-Zufallsvariable bezeichnen wir mit

$$\chi^2(x; n) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}x\right). \tag{154}$$

Bemerkungen

- Die χ^2-Verteilung ist ein Spezialfall der Gammaverteilung.
- Der Freiheitsgradparameter n kontrolliert Lage, Streuung und Schiefe der WDF.

Definition (Nichtzentrale χ^2-Zufallsvariable)

Sei X^2 eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p : \mathbb{R}_{>0} \to \mathbb{R}_{>0}, x \mapsto p(x) := \sum_{j=0}^{\infty} \exp\left(-\frac{\delta}{2}\right) \frac{\left(\frac{\delta}{2}\right)^j}{j!} \chi^2(x; n + 2j). \tag{155}$$

Dann sagen wir, dass X^2 einer nichtzentralen χ^2-Verteilung mit Freiheitsgradparameter n und Nichtzentralitätsparameter $\delta \geq 0$ unterliegt und nennen X^2 eine *nichtzentrale χ^2-Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n*. Wir kürzen dies mit

$$\mathrm{X}^2 \sim \chi^2(\delta, n) \tag{156}$$

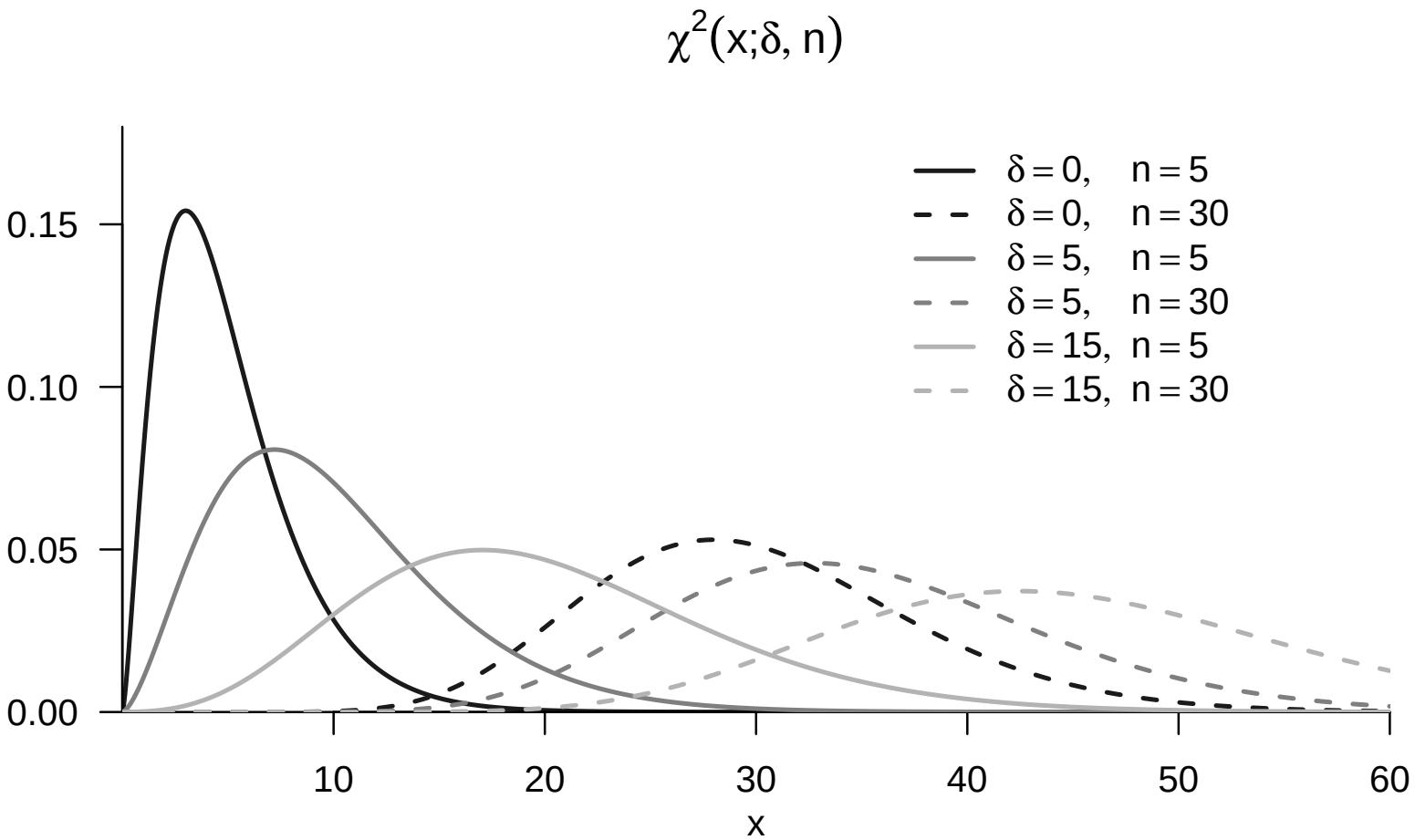
ab. Die WDF einer nichtzentralen χ^2-Zufallsvariable bezeichnen wir mit

$$\chi^2(x; \delta, n). \tag{157}$$

Bemerkungen

- Für $\delta = 0$ entspricht eine nichtzentrale χ^2-Zufallsvariable einer χ^2-Zufallsvariable mit Freiheitsgradparameter n.
- Der Nichtzentralitätsparameter δ verschiebt Wahrscheinlichkeitsmasse zu größeren Werten von X^2.

Wahrscheinlichkeitsdichten nichtzentraler χ^2-Zufallsvariablen



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,

$$\mathbb{C}(\xi) = \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1 \leq i,j \leq m}. \tag{158}$$

(2) (Kovarianzmatrixverschiebungssatz) Es gilt

$$\mathbb{C}(\xi) = \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T. \tag{159}$$

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

$$\mathbb{C}(A\xi + b) = A\mathbb{C}(\xi)A^T. \tag{160}$$

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

Beweis

(1) Es gilt

$$
\begin{aligned}
\mathbb{C}(\xi) &:= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) \\
&= \mathbb{E}\left(\left(\begin{pmatrix}\xi_1 \\ \vdots \\ \xi_m\end{pmatrix} - \begin{pmatrix}\mathbb{E}(\xi_1) \\ \vdots \\ \mathbb{E}(\xi_m)\end{pmatrix}\right)\left(\begin{pmatrix}\xi_1 \\ \vdots \\ \xi_m\end{pmatrix} - \begin{pmatrix}\mathbb{E}(\xi_1) \\ \vdots \\ \mathbb{E}(\xi_m)\end{pmatrix}\right)^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1) \\ \vdots \\ \xi_m - \mathbb{E}(\xi_m)\end{pmatrix}\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1) \\ \vdots \\ \xi_m - \mathbb{E}(\xi_m)\end{pmatrix}^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1) \\ \vdots \\ \xi_m - \mathbb{E}(\xi_m)\end{pmatrix}\begin{pmatrix}\xi_1 - \mathbb{E}(\xi_1) & \dots & \xi_m - \mathbb{E}(\xi_m)\end{pmatrix}\right) \\
&= \mathbb{E}\begin{pmatrix}(\xi_1 - \mathbb{E}(\xi_1))(\xi_1 - \mathbb{E}(\xi_1)) & \dots & (\xi_1 - \mathbb{E}(\xi_1))(\xi_m - \mathbb{E}(\xi_m)) \\ \vdots & \ddots & \vdots \\ (\xi_m - \mathbb{E}(\xi_m))(\xi_1 - \mathbb{E}(\xi_1)) & \dots & (\xi_m - \mathbb{E}(\xi_m))(\xi_m - \mathbb{E}(\xi_m))\end{pmatrix} \\
&= \left(\mathbb{E}\left((\xi_i - \mathbb{E}(\xi_i))(\xi_j - \mathbb{E}(\xi_j))\right)\right)_{1 \le i,j \le m} \\
&= \left(\mathbb{C}(\xi_i, \xi_j)\right)_{1 \le i,j \le m}.
\end{aligned}
\tag{161}
$$

Beweis (fortgeführt)

(2) Mit den Eigenschaften von Erwartungswerten gilt

$$\begin{aligned}\mathbb{C}(\xi) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) \\ &= \mathbb{E}\left(\xi\xi^T - \xi\mathbb{E}(\xi)^T - \mathbb{E}(\xi)\xi^T + \mathbb{E}(\xi)\mathbb{E}(\xi)^T\right) \\ &= \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T - \mathbb{E}(\xi)\mathbb{E}(\xi)^T + \mathbb{E}(\xi)\mathbb{E}(\xi)^T \\ &= \mathbb{E}\left(\xi\xi^T\right) - \mathbb{E}(\xi)\mathbb{E}(\xi)^T.\end{aligned} \tag{162}$$

(3) Mit den Eigenschaften von Erwartungswerten gilt

$$\begin{aligned}\mathbb{C}(A\xi + b) &= \mathbb{E}\left((A\xi + b - \mathbb{E}(A\xi + b))(A\xi + b - \mathbb{E}(A\xi + b))^T\right) \\ &= \mathbb{E}\left((A\xi + b - A\mathbb{E}(\xi) - b)(A\xi + b - A\mathbb{E}(\xi) - b)^T\right) \\ &= \mathbb{E}\left((A(\xi - \mathbb{E}(\xi)))(A(\xi - \mathbb{E}(\xi)))^T\right) \\ &= \mathbb{E}\left(A(\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T A^T\right) \\ &= A\mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) A^T \\ &= A\mathbb{C}(\xi)A^T.\end{aligned} \tag{163}$$

Beweis (fortgeführt)

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

$$\mathbb{C}(\xi_i, \xi_j) = \mathbb{C}(\xi_j, \xi_i) \text{ für alle } i = 1, ..., m, j = 1, ..., m. \tag{164}$$

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

$$a^T \mathbb{C}(\xi) a = \mathbb{C}(a^T \xi). \tag{165}$$

Weiterhin gilt mit der Definition der Kovarianzmatrix aber, dass

$$\mathbb{C}(a^T \xi) = \mathbb{E}\left(\left(a^T \xi - \mathbb{E}(a^T \xi)\right)^2\right) = \mathbb{V}\left(a^T \xi\right). \tag{166}$$

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

$$a^T \mathbb{C}(\xi) a = \mathbb{V}(a^T \xi) \geq 0 \tag{167}$$

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

Theorem (Datenmatrix und Stichprobenstatistiken)

Es sei

$$y := \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \tag{168}$$

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

- für das Stichprobenmittel

$$\bar{y} = \frac{1}{n} y 1_n, \tag{169}$$

- für die Stichprobenkovarianzmatrix

$$C = \frac{1}{n-1} \left(y \left(I_n - \frac{1}{n} 1_{nn} \right) y^T \right), \tag{170}$$

- und für die Stichprobenkorrelationsmatrix mit

$$D := \text{diag} \left((C)_{ii}^{-1/2}, i = 1, \ldots, m \right), \tag{171}$$

dass

$$R = DCD. \tag{172}$$

Beweis

Die Darstellung des Stichprobenmittels ergibt sich aus

$$\begin{aligned}
\bar{y} &:= \frac{1}{n}\sum_{i=1}^{n} y_i \\
&= \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{n} y_{i1} \\ \vdots \\ \sum_{i=1}^{n} y_{im} \end{pmatrix} \\
&= \frac{1}{n}\left(\begin{pmatrix} y_{11} & \cdots & y_{n1} \\ \vdots & \ddots & \vdots \\ y_{1m} & \cdots & y_{nm} \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right) \\
&= \frac{1}{n} y 1_n.
\end{aligned} \tag{173}$$

Beweis (fortgeführt)

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

$$
\begin{aligned}
C &:= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})^T \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i y_i^T - y_i \bar{y}^T - \bar{y} y_i^T + \bar{y} \bar{y}^T \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i y_i^T - \sum_{i=1}^{n} y_i \bar{y}^T - \sum_{i=1}^{n} \bar{y} y_i^T + \sum_{i=1}^{n} \bar{y} \bar{y}^T \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i y_i^T - n \bar{y} \bar{y}^T - n \bar{y} \bar{y}^T + n \bar{y} \bar{y}^T \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i y_i^T - n \bar{y} \bar{y}^T \right).
\end{aligned}
\tag{174}
$$

Beweis

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

$$
\begin{aligned}
y\left(I_n - \frac{1}{n}1_{nn}\right)y^T &= \left(yI_n - \frac{1}{n}y1_{nn}\right)y^T \\
&= yy^T - \frac{1}{n}y1_{nn}y^T \\
&= \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}\begin{pmatrix} y_1^T \\ \vdots \\ y_n^T \end{pmatrix} - \frac{1}{n}y1_n 1_n^T y^T \\
&= \sum_{i=1}^{n} y_i y_i^T - n\left(\frac{1}{n}y1_n\right)\left(\frac{1}{n}1_n^T y^T\right) \\
&= \sum_{i=1}^{n} y_i y_i^T - n\bar{y}\bar{y}^T \\
&= \sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^T \\
&= (n-1)C.
\end{aligned}
\tag{175}
$$

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

$$\begin{aligned} R_{ij} &= \frac{(C)_{ij}}{\sqrt{(C)_{ii}}\sqrt{(C)_{jj}}} \\ &= \frac{1}{\sqrt{(C)_{ii}}}(C)_{ij}\frac{1}{\sqrt{(C)_{jj}}} \\ &= (DCD)_{ij}. \end{aligned} \tag{176}$$

Referenzen I

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

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

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

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

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258.

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

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

Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, *48*(xi + 91).

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

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

Jöreskog, K. G. (1969). *A general approach to confirmatory maximum likelihood factor analysis*.

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

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

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

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*.

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

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

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

Referenzen IV

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

Satorra, A., & Saris, W. E. (1985). Power of the Likelihood Ratio Test in Covariance Structure Analysis. *Psychometrika*, *50*(1), 83–90. https://doi.org/10.1007/BF02294150

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

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

Steiger, J. H., & Lind, J. C. (1980). Statistically-Based Tests for the Number of Common Factors. *Annual Spring Meeting of the Psychometric Society*.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the Multivariate Asymptotic Distribution of Sequential Chi-Square Statistics. *Psychometrika*, *50*(3), 253–263. https://doi.org/10.1007/BF02294104

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

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

Referenzen V

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

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

Tucker, L. R., & Lewis, C. (1973). A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

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.