Faktoranalyse

Multivariate Analyse

Mathematischer Hintergrund, durchgerechnetes Beispiel, Handhabung des Tools

Deutsch

Wolfgang Böhme

wolfgang.boehme@economic-modeling.de

www.economic-modeling.de

8. November 2026

Dokumentationen dieser Webseite wurden mit Unterstützung der KI Claude (Anthropic) entwickelt.

Inhaltsverzeichnis

1 Einleitung und Anwendungsgebiete **2**

2 Mathematischer Hintergrund **2**
2.1 Das Faktormodell . . . 2
2.2 Kommunalitaet und Einzigartigkeit . . . 2
2.3 Reproduzierte Korrelationsmatrix . . . 2
2.4 Hauptachsen-Methode (PAF) . . . 3
2.5 KMO-Mass der Stichprobeneignung . . . 3
2.6 Bartlett-Test auf Spharizitaet . . . 3
2.7 Varimax-Rotation . . . 3

3 Durchgerechnetes Beispiel: Bilanzkennzahlen **4**
3.1 Datenbasis ($n = 24$, $p = 6$) . . . 4
3.2 Schritt 1: Korrelationsmatrix, KMO und Bartlett-Test . . . 4
3.3 Schritt 2: Eigenwerte (Kaiser-Kriterium) . . . 4
3.4 Schritt 3: PAF-Extraktion . . . 5

4 Handhabung des interaktiven Faktoranalyse-Tools **5**
4.1 Dateneingabe . . . 5
4.2 Parameterauswahl . . . 6
4.3 Ergebnistabs . . . 6
4.4 Typischer Arbeitsablauf . . . 6

5 Formelübersicht **7**

6 Weiterführende Literatur **7**

Einleitung und Anwendungsgebiete

Die **Faktoranalyse** ist ein Verfahren zur Identifikation *latenter Faktoren*, die den beobachtbaren Variablen zugrunde liegen. Anders als die PCA, die lediglich Varianzmaximierung anstrebt, geht die Faktoranalyse von einem expliziten Messmodell aus: Jede Variable wird als lineare Kombination gemeinsamer Faktoren und einer variablenspezifischen Einzigartigkeit modelliert.

Typische Anwendungsgebiete:

- Skalenkonstruktion in Psychologie und Sozialwissenschaften
- Validierung von Fragebogenitems (z. B. Big Five)
- Reduktion von Indikatoren bei Indizes (z. B. Lebensqualitaet)
- Identifikation latenter Dimensionen in Marktforschungsdaten

Mathematischer Hintergrund

Das Faktormodell

Fuer p standardisierte Variablen $\mathbf{z} = (z_1, \ldots, z_p)^\top$ gilt:

$$z_j = l_{j1}F_1 + l_{j2}F_2 + \cdots + l_{jk}F_k + u_j, \qquad j = 1, \ldots, p \tag{1}$$

mit gemeinsamen Faktoren F_k ($\mathrm{Var}(F_k) = 1$, $\mathrm{Cov}(F_k, F_m) = 0$), Faktorladungen l_{jk} und spezifischer Varianz u_j (Uniqueness).

In Matrixschreibweise:

$$\mathbf{Z} = \mathbf{F}\mathbf{L}^\top + \mathbf{U} \tag{2}$$

Kommunalitaet und Einzigartigkeit

$$h_j^2 = \sum_{k=1}^{K} l_{jk}^2, \qquad u_j = 1 - h_j^2 \tag{3}$$

Die **Kommunalitaet** h_j^2 ist der durch die K Faktoren erklaerte Varianzanteil der j-ten Variable; u_j ist der unerklarte Restanteil.

Reproduzierte Korrelationsmatrix

Das Faktormodell impliziert:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi}, \qquad \boldsymbol{\Psi} = \mathrm{diag}(u_1, \ldots, u_p) \tag{4}$$

Ziel der Extraktion: Residuen $\mathbf{R} - \hat{\mathbf{R}}$ minimieren.

Hauptachsen-Methode (PAF)

Die **Principal Axis Factoring**-Methode laeuft iterativ:

1. Initialschaetzung der Kommunalitaeten: $h_j^{(0)} = \max_{m \neq j} |r_{jm}|^2$
2. Reduzierte Korrelationsmatrix: $\mathbf{R}^* = \mathbf{R}$ mit r_{jj} ersetzt durch h_j^2
3. Eigenwertzerlegung von $\mathbf{R}^*$; Extraktion der K groessten positiven Eigenwerte/-vektoren
4. Neue Kommunalitaeten: $h_j^{(t+1)} = \sum_{k=1}^{K} l_{jk}^2$
5. Wiederholen bis Konvergenz ($|h_j^{(t+1)} - h_j^{(t)}| < 10^{-6}$)

KMO-Mass der Stichprobeneignung

Das **Kaiser-Meyer-Olkin-Mass** bewertet, ob die Korrelationsstruktur fuer eine Faktoranalyse geeignet ist:

$$\text{KMO} = \frac{\sum_{j \neq m} r_{jm}^2}{\sum_{j \neq m} r_{jm}^2 + \sum_{j \neq m} p_{jm}^2} \tag{5}$$

wobei p_{jm} die Partialkorrelation zwischen z_j und z_m ist.

Tabelle 1: Bewertung des KMO-Masses

KMO-Wert	Bewertung
≥ 0.90	hervorragend
$0.80 - 0.89$	verdienstvoll (gut)
$0.70 - 0.79$	mittelmaessig
$0.60 - 0.69$	maessig
< 0.60	inakzeptabel

Bartlett-Test auf Spharizitaet

H_0: Die Korrelationsmatrix ist eine Einheitsmatrix ($\mathbf{R} = \mathbf{I}$).

$$\chi^2 \approx -\left(n - 1 - \frac{2p + 5}{6}\right) \ln |\mathbf{R}|, \quad \text{df} = \frac{p(p-1)}{2} \tag{6}$$

Ein signifikantes Ergebnis ($p < 0.05$) zeigt, dass die Korrelationsmatrix sich von der Einheitsmatrix unterscheidet – Voraussetzung fuer FA.

Varimax-Rotation

Identisch zur PCA (siehe Handbuch 1). Ziel: einfache Struktur – jede Variable laedt hoch auf moeglichst einen Faktor.

Durchgerechnetes Beispiel: Bilanzkennzahlen

Datenbasis ($n = 24$, $p = 6$)

Verwendet wird derselbe synthetische Datensatz wie in Handbuch 1 (vollständige Datentabelle dort): 24 Unternehmen mit drei Stabilitätskennzahlen (X_1 Eigenkapitalquote, X_2 Liquidität 2. Grades, X_3 Anlagendeckung II) und drei Ertragskennzahlen (X_4 Umsatzrendite, X_5 Gesamtkapitalrendite, X_6 Umsatzwachstum), alle in Prozent. Der Datensatz ist im Tool vorgeladen.

Ziel der Faktoranalyse: Die latenten Faktoren hinter den Kennzahlen sollen identifiziert werden. Anders als die PCA erklärt die Faktoranalyse nur die *gemeinsame* Varianz der Kennzahlen; der kennzahlenspezifische Rest geht in die Einzigartigkeit u_j ein. Einstellungen: PAF, Kaiser-Kriterium, Varimax (Werkseinstellungen).

Schritt 1: Korrelationsmatrix, KMO und Bartlett-Test

Die Korrelationsmatrix (siehe Handbuch 1, Schritt 2) zeigt zwei Blöcke mit Korrelationen von 0.70–0.79 innerhalb und höchstens 0.31 zwischen den Blöcken.

Tabelle 2: Gütekriterien

Kriterium	Wert	Bewertung
KMO	0.678	mäßig (Schwelle 0.6 erfüllt)
Bartlett χ^2	80.68 (df = 15)	$p < 0.001$, signifikant

Rechenweg Bartlett: $|\mathbf{R}| = 0.0183$, also

$$\chi^2 = -\left(24 - 1 - \tfrac{2\cdot 6+5}{6}\right) \ln 0.0183 = -20.167 \cdot (-4.001) = 80.68, \qquad \text{df} = \tfrac{6\cdot 5}{2} = 15.$$

Die Korrelationsmatrix unterscheidet sich hochsignifikant von der Einheitsmatrix. Der KMO-Wert ist nur mäßig: Jeder Block besteht aus nur drei Kennzahlen, sodass die partiellen Korrelationen innerhalb der Blöcke noch beträchtlich sind. Die Mindestanforderung KMO ≥ 0.6 ist aber erfüllt – die Faktoranalyse ist zulässig.

Schritt 2: Eigenwerte (Kaiser-Kriterium)

Tabelle 3: Eigenwerte der Korrelationsmatrix

Faktor	Eigenwert	Varianzanteil	Kumuliert
F1	2.971	49.52 %	49.52 %
F2	2.010	33.50 %	83.01 %
F3	0.397	6.61 %	89.62 %
F4–F6	< 0.3	–	–

Kaiser-Kriterium: **2 Faktoren** ($\lambda \geq 1$).

Schritt 3: PAF-Extraktion

Startwerte der Kommunalitäten (größte quadrierte Korrelation je Zeile): $h_1^{(0)} = 0.743^2 = 0.553$, $h_2^{(0)} = h_3^{(0)} = 0.771^2 = 0.595$, $h_4^{(0)} = h_6^{(0)} = 0.793^2 = 0.628$, $h_5^{(0)} = 0.727^2 = 0.529$. Die Diagonale von $\mathbf{R}$ wird durch diese Werte ersetzt, die reduzierte Matrix zerlegt und das Verfahren bis zur Konvergenz wiederholt.

Tabelle 4: PAF-Ladungsmatrix: unrotiert und Varimax-rotiert

	unrotiert		Varimax			
Variable	F1	F2	F1	F2	h_j^2	u_j
X_1 EK-Quote	0.673	−0.497	**0.827**	0.122	0.699	0.301
X_2 Liquidität	0.683	−0.522	**0.853**	0.112	0.739	0.261
X_3 Anlagendeckung	0.664	−0.595	**0.890**	0.047	0.795	0.205
X_4 Umsatzrendite	0.741	0.486	0.183	**0.867**	0.785	0.215
X_5 GK-Rendite	0.618	0.506	0.082	**0.795**	0.638	0.362
X_6 Umsatzwachstum	0.654	0.631	0.019	**0.909**	0.827	0.173
Erkl. Varianz	45.3 %	29.4 %	37.4 %	37.3 %	Σ 74.7 %	

Beispiel: $h_1^2 = 0.827^2 + 0.122^2 = 0.699$, $u_1 = 1 - 0.699 = 0.301$. Die Faktoren erklären zusammen 74.7 % der Gesamtvarianz – weniger als die zwei Hauptkomponenten der PCA (83.0 %), weil PAF nur die gemeinsame Varianz modelliert.

Reproduzierte Korrelation: Das Faktormodell reproduziert z. B. $\hat{r}_{12} = 0.827 \cdot 0.853 + 0.122 \cdot 0.112 = 0.719$ (beobachtet: 0.710, Residuum −0.009) und $\hat{r}_{14} = 0.827 \cdot 0.183 + 0.122 \cdot 0.867 = 0.257$ (beobachtet: 0.311). Die kleinen Residuen bestätigen die Zwei-Faktoren-Lösung.

Interpretation:

- **Faktor 1 – Finanzielle Stabilität:** Eigenkapitalquote, Liquidität und Anlagendeckung (Ladungen 0.83–0.89). Er misst die Substanz und Krisenfestigkeit eines Unternehmens.
- **Faktor 2 – Ertragskraft:** Umsatzrendite, Gesamtkapitalrendite und Umsatzwachstum (Ladungen 0.80–0.91). Er misst die Fähigkeit, Gewinne zu erwirtschaften und zu wachsen.
- Querladungen liegen durchweg unter 0.2: Die Varimax-Lösung erfüllt das Kriterium der Einfachstruktur sehr gut. Die Gesamtkapitalrendite hat die niedrigste Kommunalität (0.638) und ist damit am schwächsten durch die Faktoren erklärt, liegt aber deutlich über der Grenze von 0.4.

Handhabung des interaktiven Faktoranalyse-Tools

Dateneingabe

Identisch zur PCA (Manuell, CSV, Beispiel). Vorgeladen sind die Bilanzkennzahlen der 24 Unternehmen ($n = 24$, $p = 6$). Weitere Beispiele: Arbeitszufriedenheit ($n = 20$, $p = 6$), Persönlichkeitsskala (Big Five, $n = 25$, $p = 8$) und Konsumverhalten ($n = 18$,

$p = 5$). Zu jedem Beispiel erscheint über der Eingabetabelle ein Erläuterungskasten mit den Daten, dem Ziel der Analyse und dem erwarteten Ergebnis; er verschwindet bei »Leer« oder beim Laden einer CSV-Datei.

Parameterauswahl

Extraktionsmethode
Hauptachsen-Methode (PAF): Empfohlen; modelliert latente Faktoren korrekt durch iterative Kommunalitaetsschaetzung. *Hauptkomponentenmethode:* Schnell, aber erklaert auch spezifische Varianz.

Faktoranzahl
Automatisch: Kaiser-Kriterium fuer die Korrelationsmatrix. Alternativ: 1–5 Faktoren manuell vorgeben.

Rotation
Varimax (Standard): Orthogonale Rotation; Faktoren bleiben unkorreliert. *Keine:* Unrotierte Loesung fuer direkte Vergleichbarkeit mit PCA.

Ladungsschwelle
Hervorhebung ab $|l| \geq 0.4$ (anpassbar). Werte ≥ 0.5 gelten als substanziell.

KMO und Bartlett
Checkbox aktiviert die Berechnung der Voraussetzungspruefung.

Ergebnistabs

Uebersicht
Kennzahlen + Eigenwert-Tabelle mit Kaiser-Badge.

Ladungen
Faktorladungsmatrix mit Farbhervorhebung, Kommunalitaeten und Einzigartigkeit je Variable.

Guete
KMO-Ampel, Bartlett-Test mit p-Wert, Kommunalitaetentabelle mit Qualitaetsbewertung.

Grafiken
Scree-Plot, Varianzbalken mit kumulativer Linie, Ladungsplot (F1 vs. F2) mit Variablenbeschriftung.

Interpretation
Automatischer Bewertungstext mit KMO, Bartlett und Rotationshinweis.

Export
CSV: Ladungen, Kommunalitaeten, Eigenwerte. PNG: Scree-Plot, Ladungsplot.

Typischer Arbeitsablauf

1. Daten laden und Extraktionsmethode PAF waehlen
2. KMO-Checkbox aktiviert lassen
3. »**Faktoranalyse berechnen**« klicken

4. Guete-Tab: KMO ≥ 0.6 und Bartlett signifikant pruefen
5. Scree-Plot: Faktoranzahl ggf. manuell anpassen
6. Ladungen-Tab: inhaltliche Interpretation der Faktoren
7. Variablen mit $h_j^2 < 0.4$ ggf. entfernen und neu berechnen

Wichtige Hinweise:
- Mindestens $n = 5$ vollstaendige Beobachtungen erforderlich.
- Faustformel: $n \geq 10 \cdot p$ fuer zuverlassige Ergebnisse.
- Variablen mit $h_j^2 < 0.3$ sind schlecht erklaert – Entfernung erwaegen.
- Bei Cross-Loadings ($|l_{jk}| \geq 0.3$ auf mehreren Faktoren): oblique Rotation (Oblimin) erwaegen.

Formelübersicht

Tabelle 5: Formeln der Faktoranalyse

Bezeichnung	**Formel**
Faktormodell	$z_j = \sum_k l_{jk} F_k + u_j$
Reproduz. Korrelation	$\mathbf{R} = \mathbf{L}\mathbf{L}^\top + \mathbf{\Psi}$
Kommunalitaet	$h_j^2 = \sum_k l_{jk}^2$
Einzigartigkeit	$u_j = 1 - h_j^2$
KMO-Mass	$\text{KMO} = \sum r_{jm}^2 / (\sum r_{jm}^2 + \sum p_{jm}^2)$
Bartlett χ^2	$-(n - 1 - \frac{2p+5}{6}) \ln \lvert\mathbf{R}\rvert$, $\text{df} = p(p-1)/2$

Weiterführende Literatur

- Comrey, A. L. & Lee, H. B. (1992). *A First Course in Factor Analysis*. LEA.
- Backhaus, K. et al. (2021). *Multivariate Analysemethoden*, 16. Aufl. Springer Gabler.
- Fabrigar, L. R. et al. (1999). Evaluating the use of exploratory FA in research. *Psychological Methods*, 4(3), 272–299.