Statistische Analyseverfahren
Abschnitt 6: Faktoranalyse (Faktorenanalyse)

Dr. Andreas Wünsche

TU Bergakademie Freiberg
Institut für Stochastik

Dezember 2019/Januar 2020

6 Faktoranalyse (Faktorenanalyse)

- Die Faktoranalyse ist ein mathematisches Verfahren, welches (im Gegensatz zur Hauptkomponentenanalyse) auf die Erklärung der Korrelationen zwischen vielen messbaren (beobachtbaren bzw. beobachteten) Merkmalen hauptsächlich durch **wenige** zugrundeliegende (nichtbeobachtbare bzw. nicht beobachtete, „latente") sogenannte allgemeine Faktoren abzielt.
- Insbesondere hohe Korrelationen zwischen den Merkmalen können darauf hindeuten, dass diese von einer oder von mehreren latenten Größen beeinflusst werden.
- Zum Beispiel sind Körpergröße und Gewicht (oder Schuhgröße) korreliert und beschreiben in etwa den latenten Faktor „Statur").
- Die Merkmale und Faktoren sind durch lineare Beziehungen verknüpft und aus mathematischer Sicht kann man die Faktoranalyse als Strukturanalyse von Kovarianz- bzw. Korrelationsmatrizen ansehen.

Weitere einführende Bemerkungen

- Die Faktoranalyse wurde zuerst von Psychologen entwickelt, später erfolgten statistische Begründungen, mathematische Untersuchungen und die Entwicklung weiterer Verfahren.
- Bei der Hauptkomponentenanalyse waren die Hauptkomponenten Linearkombinationen der Merkmale (und auch umgekehrt). Bei der Faktoranalyse werden umgekehrt die Merkmale als Linearkombinationen der Faktoren dargestellt.
- Ziel der Faktoranalyse ist die Extraktion der latenten Faktoren.
- Insbesondere wird untersucht, wie viele und welche Faktoren die Zusammenhänge möglichst gut „erklären“.
- Die Faktoren sollten möglichst gut interpretierbar sein.
- Ein typisches Beispiel besteht in der Extraktion von Faktoren wie Kreditwürdigkeit oder Grad der Existenzgefährdung einer Firma aus Realisierungen von Bilanzmerkmalen $X_1, \ldots, X_p$.

6.1 Das k-Faktor-Modell

- **Def. 6.1.1**
  Sei $\underline{\mathbf{X}}$ ein $p-$dimensionaler Zufallsvektor mit $\mathbb{E}\underline{\mathbf{X}} = \underline{\boldsymbol{\mu}}$, $\mathbb{C}\mathbf{ov}\underline{\mathbf{X}} = \underline{\underline{\boldsymbol{\Sigma}}}$.
  $\underline{\mathbf{X}}$ genügt einem k-Faktor-Modell ($k \in \mathbb{N}, k < p$), falls gilt

$$\underline{\mathbf{X}} = \underline{\underline{\boldsymbol{\lambda}}}\,\underline{\mathbf{F}} + \underline{\mathbf{U}} + \underline{\boldsymbol{\mu}} \tag{1}$$

  (**Bezeichnung:** $\underline{\mathbf{X}} \sim F(k, \underline{\underline{\boldsymbol{\lambda}}}, \underline{\underline{\boldsymbol{\psi}}})$) mit

  (i) $\underline{\mathbf{F}}$, ein $k-$dimensionaler Zufallsvektor mit $\mathbb{E}\underline{\mathbf{F}} = \underline{\mathbf{0}}_k$, $\mathbb{C}\mathbf{ov}\underline{\mathbf{F}} = \underline{\underline{\mathbb{I}_{\mathbf{k}}}}$,
  der Vektor der allgemeinen Faktoren,

  (ii) $\underline{\underline{\boldsymbol{\lambda}}}$, die deterministische $p \times k-$Matrix der Faktorladungen,

  (iii) $\underline{\mathbf{U}}$, ein $p-$dimensionaler Zufallsvektor der spezifischen Faktoren mit
  $\mathbb{E}\underline{\mathbf{U}} = \underline{\mathbf{0}}_p$, $\mathbb{C}\mathbf{ov}\underline{\mathbf{U}} = \mathrm{diag}(\psi_1, \ldots, \psi_p) =: \underline{\underline{\boldsymbol{\psi}}}$, $\mathbb{C}\mathbf{ov}[\underline{\mathbf{F}}, \underline{\mathbf{U}}] = \underline{\underline{\mathbf{0}}}_{k \times p}$.

  (iv) $\underline{\boldsymbol{\mu}}$, der $p-$dimensionale deterministische Erwartungswertvektor des
  Merkmalszufallsvektors $\underline{\mathbf{X}}$.

- **Bem.** Die p Komponenten des Zufallsvektors $\underline{\mathbf{X}}$ werden durch
  $k + p$ andere Zufallsgrößen beschrieben.

Komponentenweise Modellgleichungen

Bem. 6.1.2

- Die Modellgleichung (1) komponentenweise lautet

$$X_i = \sum_{j=1}^{k} \lambda_{ij} F_j + U_i + \mu_i \,, \qquad i = 1, \ldots, p \,,$$

  mit der Interpretation: wenige (k) allgemeine Faktoren $F_1, \ldots, F_k$ und jeweils ein spezieller oder spezifischer Faktor U_i „erklären“ das Merkmal X_i (d.h. insgesamt $k+1$ Zufallsgrößen).
- λ_{ij} wird als Ladung des $j-$ten Faktors beim $i-$ten Merkmal bezeichnet (und gibt damit den Einfluss des $j-$ten allgemeinen Faktors auf das $i-$te Merkmal an).
- Die allgemeinen Faktoren besitzen die Varianz 1 und sie sind unkorreliert.
- Die spezifischen Faktoren sind untereinander unkorreliert und unkorreliert zu den allgemeinen Faktoren, der $i-$te spezifische Faktor besitzt die Varianz ψ_i .

Fundamentaltheorem der Faktoranalyse

- **Satz 6.1.3**
  Für ein k-Faktor-Modell gemäß Def. 6.1.1 gelten

$$\mathbb{C}\mathbf{ov}\underline{\mathbf{X}} = \underline{\underline{\mathbf{\Sigma}}} = \underline{\underline{\boldsymbol{\lambda}}}\,\underline{\underline{\boldsymbol{\lambda}}}^{\mathsf{T}} + \underline{\underline{\boldsymbol{\psi}}} = \underline{\underline{\boldsymbol{\lambda}}}\,\underline{\underline{\boldsymbol{\lambda}}}^{\mathsf{T}} + \mathrm{diag}(\psi_1, \ldots, \psi_p)\,; \qquad (2)$$

$$\mathbb{V}\mathbf{ar}X_i = \sigma_{ii} = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_i = h_i^2 + \psi_i\,, \quad i = 1, \ldots, p\,. \qquad (3)$$

Bem. 6.1.4

- Bei gegebenen $k < p$ und $\underline{\underline{\boldsymbol{\psi}}}$ sind notwendig und hinreichend für die Existenz einer solchen Kovarianzstruktur: der Rang der Matrix $\underline{\underline{\mathbf{\Sigma}}} - \underline{\underline{\boldsymbol{\psi}}}$ ist kleiner oder gleich k und $\underline{\underline{\mathbf{\Sigma}}} - \underline{\underline{\boldsymbol{\psi}}}$ ist positiv semidefinit.
- $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$ wird allgemeine Varianz oder Kommunalität des $i-$ten Merkmals genannt, sie misst den Teil der Varianz, der durch die allgemeinen Faktoren erklärt wird.
- ψ_i ist die spezifische Varianz des $i-$ten Merkmals.

Skaleninvarianz

Bem. 6.1.5
Ein k-Faktor-Modell ist (im Gegensatz zur Hauptkomponentenanalyse) skaleninvariant im folgenden Sinn.

(i) Es gelte $\underline{\mathbf{X}} \sim F(k, \underline{\underline{\boldsymbol{\lambda}}}, \underline{\underline{\psi}})$ und $\underline{\underline{\mathbf{C}}} = \mathrm{diag}(c_1, \ldots, c_p)$ sei eine Skalierungsmatrix mit positiven Elementen auf der Hauptdiagonale. Dann gilt für den skalierten Zufallsvektor

$$\underline{\mathbf{Y}} := \underline{\underline{\mathbf{C}}}\,\underline{\mathbf{X}} \sim F(k, \underline{\underline{\mathbf{C}}}\,\underline{\underline{\boldsymbol{\lambda}}}, \underline{\underline{\mathbf{C}}}\,\underline{\underline{\psi}}\,\underline{\underline{\mathbf{C}}}) .$$

(ii) Insbesondere kann $c_i = \dfrac{1}{\sqrt{\sigma_{ii}}}$, $i = 1, \ldots, p$, gewählt werden, so dass man mit standardisierten Zufallsgrößen und der zugehörigen Kovarianzmatrix, d.h. mit der Korrelationsmatrix des Zufallsvektors $\underline{\mathbf{X}}$ arbeiten kann. In diesem Fall sind die Faktorladungen Korrelationskoeffizienten der Merkmale mit den allgemeinen Faktoren und so insbesondere betragsmäßig kleiner (oder gleich) 1.

Nicht-Eindeutigkeit der Faktorladungen

Bem. 6.1.6

(i) In einem k-Faktor-Modell ist die Matrix der Faktorladungen für einen gegebenen Zufallsvekor $\underline{\mathbf{X}}$ nicht eindeutig bestimmt. Ist $\underline{\underline{\mathbf{g}}}$ eine orthogonale $k \times k-$Matrix, dann gilt mit

$$\underline{\mathbf{X}} = \underline{\underline{\boldsymbol{\lambda}}}\,\underline{\mathbf{F}} + \underline{\mathbf{U}} + \underline{\boldsymbol{\mu}} \qquad \text{auch} \qquad \underline{\mathbf{X}} = (\underline{\underline{\boldsymbol{\lambda}}}\,\underline{\underline{\mathbf{g}}})(\underline{\underline{\mathbf{g}}}^{\mathsf{T}}\underline{\mathbf{F}}) + \underline{\mathbf{U}} + \underline{\boldsymbol{\mu}}\,,$$

also die Definitionsbeziehung für ein $k-$Faktor-Modell mit der Matrix der Faktorladungen $\underline{\underline{\boldsymbol{\lambda}}}\,\underline{\underline{\mathbf{g}}}$ und dem Vektor der allgemeinen Faktoren $\underline{\underline{\mathbf{g}}}^{\mathsf{T}}\underline{\mathbf{F}}$. Man spricht in diesem Zusammenhang von einer Rotation der Faktoren.

(ii) Durch zusätzliche formale Nebenbedingungen, z.B. $\underline{\underline{\boldsymbol{\lambda}}}^{\mathsf{T}}\underline{\underline{\psi}}^{-1}\underline{\underline{\boldsymbol{\lambda}}}$ ist diagonal oder $\underline{\underline{\boldsymbol{\lambda}}}^{\mathsf{T}}\mathrm{diag}(\sigma_{11},\ldots,\sigma_{pp})^{-1}\underline{\underline{\boldsymbol{\lambda}}}$ ist diagonal, kann $\underline{\underline{\boldsymbol{\lambda}}}$ im Wesentlichen eindeutig bestimmt werden (bis auf Vorzeichenwechsel in den Spalten), so dass das Fundamentaltheorem der Faktoranalyse erfüllt ist.

Parameteranzahl

Bem. 6.1.7

(i) Die Matrix $\underline{\underline{\boldsymbol{\Sigma}}}$ ohne zusätzliche Struktur enthält $\frac{1}{2}p(p+1)$ Parameter, mit der Struktur eines k-Faktor-Modells $pk+p$ Parameter und mit einer zusätzlichen Nebenbedingung wie in 6.1.6 (ii) $pk+p-\frac{1}{2}k(k-1)$ Parameter. Da die Faktoranalyse nur sinnvoll ist, wenn dadurch die Anzahl der unbekannten Parameter verringert wird, fordert man im Allgemeinen

$$\frac{p(p+1)}{2}-\left(pk+p-\frac{k(k-1)}{2}\right)=\frac{(p-k)^2-p-k}{2}:=s>0.$$

(ii) Diese Bedingung ergibt als maximale Anzahlen k von allgemeinen Faktoren in Abhängigkeit von der Anzahl p der Variablen

p	3	4	5	6	7	8	9	10	11	12
max k für $s>0$		1	2	2	3	4	5	5	6	7
max k für $s\geq 0$	1	1	2	3	3	4	5	6	6	7

Fortsetzung Parameteranzahl

(iii) Bei Nutzung standardisierter Zufallsgrößen enthält deren Kovarianzmatrix (also ggf. die Korrelationsmatrix der Ausgangszufallsgrößen) $\frac{1}{2}p(p-1)$ Parameter. Die Bedingung für eine geringere Parameteranzahl in einem k-Faktor-Modell mit einer zusätzlichen Nebenbedingung wie in 6.1.6 (ii) ist dann

$$\frac{p(p-1)}{2} - \left(pk + p - \frac{k(k-1)}{2}\right) = \frac{(p-k)^2 - 3p - k}{2} := s > 0.$$

(iv) Diese Bedingung ergibt als maximale Anzahlen k von allgemeinen Faktoren in Abhängigkeit von der Anzahl p der Variablen

p	3	4	5	6	7	8	9	10	11	12
max k für $s > 0$				1	2	2	3	4	4	5
max k für $s \geq 0$			1	1	2	2	3	4	4	5

(v) Das Auffinden von allgemeinen Faktoren kann aber auch so inhaltlich von Nutzen sein.

Beispiel 1-Faktor-Modell

Bsp. 6.1.8

- Es seien $p = 3$ und $k = 1$, dann lautet das 1-Faktor-Modell

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \cdot F + \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} .$$

- Für (2) erhält man mit $\sigma_{ii} = \sigma_i^2 > 0$, $i = 1, 2, 3$,

$$\begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1^2 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_1\lambda_2 & \lambda_2^2 & \lambda_2\lambda_3 \\ \lambda_1\lambda_3 & \lambda_2\lambda_3 & \lambda_3^2 \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{pmatrix} .$$

Fortsetzung Beispiel 1-Faktor-Modell

- Falls $\rho_{ij} \neq 0,\ i,j = 1,2,3\ (i \neq j)$ und

$$0 \le \frac{\rho_{12}\rho_{13}}{\rho_{23}} \le 1\,, \quad 0 \le \frac{\rho_{12}\rho_{23}}{\rho_{13}} \le 1\,, \quad 0 \le \frac{\rho_{13}\rho_{23}}{\rho_{12}} \le 1$$

existiert eine eindeutige Lösung (bis auf die Vorzeichen der λ_i)

$$\lambda_1^2 = \frac{\rho_{12}\rho_{13}}{\rho_{23}}\sigma_1^2\,, \quad \lambda_2^2 = \frac{\rho_{12}\rho_{23}}{\rho_{13}}\sigma_2^2\,, \quad \lambda_3^2 = \frac{\rho_{13}\rho_{23}}{\rho_{12}}\sigma_3^2\,,$$

$$\psi_1 = \sigma_1^2 - \lambda_1^2\,, \quad \psi_2 = \sigma_2^2 - \lambda_2^2\,, \quad \psi_3 = \sigma_3^2 - \lambda_3^2\,.$$

- Nutzt man standardisierte Merkmalszufallsgrößen kann man mit der Korrelationsmatrix rechnen und obige Formeln gelten mit $\sigma_1 = \sigma_2 = \sigma_3 = 1\,.$
- Dieses Verfahren kann auch für Zufallsvektoren mit $p > 3$ Komponenten angewandt werden. Ist die erste Zeile der Korrelationsmatrix gleich $(1, \rho_{12}, \rho_{13}, \ldots, \rho_{1p})$, muss z.B. für die zweite Zeile eine Konstante c existieren, so dass diese lautet $(\rho_{12}, 1, c\rho_{13}, \ldots, c\rho_{1p})$, analog für die anderen Zeilen.

6.2 Schätzung von $\underline{\underline{\boldsymbol{\lambda}}}$ und $\underline{\underline{\boldsymbol{\psi}}}$

Bem. 6.2.1

(i) Seien $\underline{\underline{\mathbf{x}}}$ eine Datenmatrix, sowie $\underline{\hat{\boldsymbol{\mu}}} = \underline{\bar{\mathbf{x}}}$ und $\underline{\underline{\hat{\boldsymbol{\Sigma}}}} = \underline{\underline{\mathbf{s}}}_{\underline{\underline{\mathbf{x}}}}$ mit Diagonalelementen $s_{11}, \ldots, s_{pp}$ Schätzer der Parameter.

(ii) Aufgrund der Skaleninvarianz kann man mit der standardisierten Datenmatrix $\underline{\underline{\mathbf{y}}} := (\underline{\underline{\mathbf{I}}}_p - \frac{1}{n}\underline{\mathbf{1}}_n\underline{\mathbf{1}}^\mathsf{T})\underline{\underline{\mathbf{x}}}\,\underline{\underline{\mathbf{d}}}^{-1}$ mit $\underline{\underline{\mathbf{d}}} = \operatorname{diag}(s_{11}, \ldots, s_{pp})$ arbeiten. Die empirische Kovarianzmatrix zu $\underline{\underline{\mathbf{y}}}$ ist die empirische Korrelationsmatrix zur Datenmatrix $\underline{\underline{\mathbf{x}}}$, also $\underline{\underline{\mathbf{r}}}_{\underline{\underline{\mathbf{x}}}} = \underline{\underline{\mathbf{r}}}$.

(iii) Hier wird die Datenmatrix im Folgenden weiter mit $\underline{\underline{\mathbf{x}}}$ und die zugehörige empirische Korrelationsmatrix mit $\underline{\underline{\mathbf{r}}}$ bezeichnet.

(iv) Dann sind für festes k eine Matrix der Faktorladungen $\underline{\underline{\boldsymbol{\ell}}}$ und die Diagonalmatrix der spezifischen Varianzen $\underline{\underline{\boldsymbol{\psi}}}$ gesucht, so dass gilt

$$\underline{\underline{\mathbf{r}}} = \underline{\underline{\boldsymbol{\ell}}}\,\underline{\underline{\boldsymbol{\ell}}}^\mathsf{T} + \underline{\underline{\boldsymbol{\psi}}} \qquad \text{bzw.} \qquad \underline{\underline{\mathbf{r}}} - \underline{\underline{\boldsymbol{\psi}}} = \underline{\underline{\boldsymbol{\ell}}}\,\underline{\underline{\boldsymbol{\ell}}}^\mathsf{T}.$$

Dabei sollten die spezifischen Varianzen möglichst klein sein.

Spezialfall 1-Faktor-Modell

Bsp. 6.2.2

- Im Fall eines 1-Faktor-Modells (z.B. mit $p = 3$ und standardisierten Merkmalen) können (falls $r_{ij} \neq 0\,,\ i, j = 1, 2, 3$) die theoretischen Formeln von Beispiel 6.1.8 auf die empirischen Schätzwerte der Korrelationskoeffizienten angewandt werden, um Schätzwerte für die Faktorladungen und Varianzen der spezifischen Faktoren zu erhalten:

$$\hat{\lambda}_1^2 = \ell_1^2 = \frac{r_{12} r_{13}}{r_{23}}\,, \quad \hat{\lambda}_2^2 = \ell_2^2 = \frac{r_{12} r_{23}}{r_{13}}\,, \quad \hat{\lambda}_3^2 = \ell_3^2 = \frac{r_{13} r_{23}}{r_{12}}\,,$$

$$\hat{\psi}_1 = 1 - \hat{\lambda}_1^2\,, \quad \hat{\psi}_2 = 1 - \hat{\lambda}_2^2\,, \quad \hat{\psi}_3 = 1 - \hat{\lambda}_3^2\,.$$

- Gilt dabei $\hat{\lambda}_i^2 < 0$ oder $\hat{\psi}_i < 0$ für ein $i \in \{1, 2, 3\}$, dann kann kein 1-Faktor-Modell gefunden werden. Man nennt dies auch einen HEYWOOD-Fall.

Klassisches Beispiel von Spearman (1904)

Bsp. 6.2.3 (Spearman 1904)

- Untersuchungen von Prüfungsergebnissen von Kindern in Altphilologie ("classics", X_1), Französisch (X_2) und Englisch (X_3). Die Auswertung ergab die folgende Stichproben-Korrelationsmatrix

$$\underline{\underline{\mathbf{r}}} = \begin{pmatrix} 1 & 0.83 & 0.78 \\ & 1 & 0.67 \\ & & 1 \end{pmatrix}.$$

- Diese Matrix hat vollen Rang, besitzt jedoch die Darstellung

$$\begin{aligned} \underline{\underline{\mathbf{r}}} &\approx \begin{pmatrix} 0.983 \\ 0.844 \\ 0.793 \end{pmatrix} (0.983\; 0.844\; 0.793) + \begin{pmatrix} 0.034 & 0 & 0 \\ 0 & 0.287 & 0 \\ 0 & 0 & 0.370 \end{pmatrix} \\ &= \underline{\boldsymbol{\ell}} \cdot \underline{\boldsymbol{\ell}}^{\mathsf{T}} + \underline{\underline{\boldsymbol{\psi}}}. \end{aligned}$$

Fortsetzung Beispiel 6.2.3

- Die Ladungsmatrix ist $\ell = (0.983\ 0.844\ 0.793)^{\mathsf{T}}$. Bei allen drei Merkmalen hat der eine allgemeine Faktor eine hohe Ladung.
- Den einen allgemeinen Faktor F in $\underline{\mathbf{X}} = \underline{\boldsymbol{\lambda}} F + \underline{\mathbf{U}} + \underline{\boldsymbol{\mu}}$ könnte man als „Sprachbegabtheit“ oder “Sprachtalent“ interpretieren.
- Da $\mathbb{V}\mathbf{ar} X_1 = 0.983^2$ und $\psi_1 = 1 - 0.983^2 = 0.034$ ist das Merkmal X_1 („Altphilologie“) fast identisch mit dem allgemeinen Faktor.
- Das dritte Merkmal („Englisch“) lässt sich am wenigsten (aber immerhin zu $\approx 63\,\%$) durch den allgemeinen Faktor „Sprachtalent“ erklären.

Verschiedene Schätzmethoden

Bem. 6.2.4

Es existieren verschiedene Schätzmethoden, z.B.

(i) die Hauptfaktorenanalyse (engl. "principal factor analysis"), dabei wird die Spektralzerlegung symmetrischer Matrizen in einem iterativen Verfahren genutzt;

(ii) die Maximum-Likelihood-Faktor-Analyse unter der Voraussetzung, dass die Daten normalverteilt sind,

$$\underline{\mathbf{X}} \sim \mathrm{N}(\underline{\boldsymbol{\mu}}, \underline{\underline{\boldsymbol{\Sigma}}} = \underline{\underline{\boldsymbol{\lambda}}}\,\underline{\underline{\boldsymbol{\lambda}}}^{\mathsf{T}} + \underline{\underline{\boldsymbol{\psi}}}),$$

$\underline{\underline{\boldsymbol{\lambda}}}$ und $\underline{\underline{\boldsymbol{\psi}}}$ werden in einem iterativen Verfahren geschätzt;

(iii) die Zentroid-Methode (Thurstone, 1931), die relativ einfach, aber auch mit viel Willkür verbunden ist; oder

(iv) die kanonische Faktoranalyse (Rao, 1955), bei der ein Anliegen ist, dass die Korrelation zwischen den p Merkmalen und den zu bestimmenden k allgemeinen Faktoren so groß wie möglich sein soll.

Algorithmus für die Hauptfaktorenanalyse

Algorithmus 6.2.5
(Quelle: Handl, Kuhlenkasper, Multivariate Analysemethoden, Springer 2017, Kap. 9)

1. Mit dem Bestimmtheitsmaß b_i^2 einer Regression von X_i auf die restlichen Variablen sei $\hat{\psi}_i = 1 - b_i^2$ der Schätzer der $i-$ten spezifischen Varianz $(i = 1, \ldots, p)$.
2. Aufstellen von $\underline{\underline{\hat{\psi}}} = \mathrm{diag}(\hat{\psi}_1, \ldots, \hat{\psi}_p)$.
3. Berechnung von $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\psi}}}$.
4. Spektralzerlegung von $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\psi}}}$ zur Bestimmung von $\underline{\underline{\hat{\lambda}}}$.
5. Neue Diagonalmatrix $\underline{\underline{\hat{\psi}}}$ mit den Hauptdiagonalelementen von $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\lambda}}}\,\underline{\underline{\hat{\lambda}}}^{\mathsf{T}}$ auf der Hauptdiagonalen.
6. Wiederholung der Schritte 3., 4. und 5. so lange, bis aufeinander folgende Paare von $\underline{\underline{\hat{\psi}}}$ und $\underline{\underline{\hat{\lambda}}}$ in einer vorgegebenen Genauigkeit identisch sind.

Beispiel für die Hauptfaktorenanalyse

Bsp. 6.2.6 (Quelle: Handl, Kuhlenkasper, Multivariate Analysemethoden, Springer 2017, Kap. 9, Bsp. 40)

- Empirische Korrelationsmatrix und deren Inverse.

$$\underline{\underline{\mathbf{r}}} = \begin{pmatrix} 1.000 & 0.223 & 0.133 & 0.625 & 0.506 & 0.500 \\ 0.223 & 1.000 & 0.544 & 0.365 & 0.320 & 0.361 \\ 0.133 & 0.544 & 1.000 & 0.248 & 0.179 & 0.288 \\ 0.625 & 0.365 & 0.248 & 1.000 & 0.624 & 0.630 \\ 0.506 & 0.320 & 0.179 & 0.624 & 1.000 & 0.625 \\ 0.500 & 0.361 & 0.288 & 0.630 & 0.625 & 1.000 \end{pmatrix}$$

$$\underline{\underline{\mathbf{r}}}^{-1} = \begin{pmatrix} 1.731 & 0.040 & 0.060 & -0.810 & -0.250 & -0.231 \\ 0.040 & 1.578 & -0.740 & -0.234 & -0.170 & -0.123 \\ 0.060 & -0.740 & 1.451 & -0.070 & 0.124 & -0.215 \\ -0.810 & -0.234 & -0.070 & 2.365 & -0.600 & -0.605 \\ -0.250 & -0.170 & 0.124 & -0.600 & 1.974 & -0.705 \\ -0.231 & -0.123 & -0.215 & -0.605 & -0.705 & 2.043 \end{pmatrix}$$

Bsp. 6.2.6 Schätzung der Kommunalitäten

- Es gilt $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\psi}}} = \underline{\underline{\hat{\boldsymbol{\lambda}}}}\,\underline{\underline{\hat{\boldsymbol{\lambda}}}}^{\mathsf{T}}$. Die Matrizen $\underline{\underline{\mathbf{r}}}$ und $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\psi}}}$ unterscheiden sich nur auf der Hauptdiagonalen, auf der rechten Seite stehen dort die Schätzungen der Kommunalitäten $\hat{h}_i^2, i = 1, \ldots, p$. Diesen Schätzwert erhält man, indem man die $i-$te Variable auf die anderen Variablen regressiert, der Wert ist das Bestimmtheitsmaß b_i^2, für welches gilt $b_i^2 = 1 - \dfrac{1}{r^{ii}}$, wobei r^{ii} das $i-$te Element der Hauptdiagonale von $\underline{\underline{\mathbf{r}}}^{-1}$ ist.
- So erhält man die Schätzungen

$$\hat{h}_1^2 = 0.422\,, \quad \hat{h}_2^2 = 0.366\,, \quad \hat{h}_3^2 = 0.311\,,$$
$$\hat{h}_4^2 = 0.577\,, \quad \hat{h}_5^2 = 0.493\,, \quad \hat{h}_6^2 = 0.511\,.$$

Bsp. 6.2.6 Spektralzerlegung

- Mit diesen Schätzern gilt

$$\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\psi}}} = \begin{pmatrix} 0.422 & 0.223 & 0.133 & 0.625 & 0.506 & 0.500 \\ 0.223 & 0.366 & 0.544 & 0.365 & 0.320 & 0.361 \\ 0.133 & 0.544 & 0.311 & 0.248 & 0.179 & 0.288 \\ 0.625 & 0.365 & 0.248 & 0.577 & 0.624 & 0.630 \\ 0.506 & 0.320 & 0.179 & 0.624 & 0.493 & 0.625 \\ 0.500 & 0.361 & 0.288 & 0.630 & 0.625 & 0.511 \end{pmatrix}$$

- Die Spektralzerlegung ergibt

$$\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\psi}}} = \underline{\underline{\mathbf{v}}}\,\underline{\underline{\mathbf{d}}}\,\underline{\underline{\mathbf{v}}}^{\mathsf{T}}$$

mit einer Diagonalmatrix $\underline{\underline{\mathbf{d}}}$, deren Diagonalelemente im Allgemeinen nicht alle nichtnegativ sind.

Bsp. 6.2.6 Approximation von $\underline{\underline{\hat{\boldsymbol{\lambda}}}}$

- Sei $\underline{\underline{\mathbf{v}}}_1$ die Matrix mit den normierten Eigenvektoren zu den k größten Eigenwerten, wobei vorausgesetzt wird, dass diese positiv sind, $\underline{\underline{\mathbf{d}}}_1$ sei die Diagonalmatrix mit diesen Eigenwerten.
- Eine neue Approximation von $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\boldsymbol{\psi}}}}$ ist

$$\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\boldsymbol{\psi}}}} \approx \underline{\underline{\mathbf{v}}}_1 \, \underline{\underline{\mathbf{d}}}_1^{1/2} \left(\underline{\underline{\mathbf{v}}}_1 \, \underline{\underline{\mathbf{d}}}_1^{1/2} \right)^{\mathsf{T}} =: \underline{\underline{\hat{\boldsymbol{\lambda}}}}_1 \, \underline{\underline{\hat{\boldsymbol{\lambda}}}}_1^{\mathsf{T}} \, .$$

Diese Schätzung kann in die Ausgangsgleichung in 5. eingesetzt und damit die Iterationen fortgesetzt werden.

Fortsetzung Beispiel 6.2.6

- Im Beispiel gibt es zwei positive Eigenwerte $d_1 = 2.606\,, d_2 = 0.573$ mit zugehörigen Eigenvektoren

$$\underline{\mathbf{v}}_1 = (0.403\,,\, 0.322\,,\, 0.248\,,\, 0.497\,,\, 0.452\,,\, 0.470)^{\mathsf{T}}$$

$$\underline{\mathbf{v}}_2 = (0.316\,,\, -0.602\,,\, -0.669\,,\, 0.191\,,\, 0.219\,,\, 0.081)^{\mathsf{T}}$$

- Durch Multiplikation von $\underline{\mathbf{v}}_1$ mit $\sqrt{d_1} = 1.614$ und von $\underline{\mathbf{v}}_2$ mit $\sqrt{d_2} = 0.757$ erhalten wir die Spalten der Matrix $\underline{\underline{\hat{\boldsymbol{\lambda}}}}_1$

$$\underline{\underline{\hat{\boldsymbol{\lambda}}}}_1 = \begin{pmatrix} 0.650 & 0.519 & 0.400 & 0.803 & 0.730 & 0.759 \\ 0.239 & -0.455 & -0.506 & 0.145 & 0.165 & 0.061 \end{pmatrix}^{\mathsf{T}}.$$

Fortsetzung Beispiel 6.2.6

- Man erhält

$$\underline{\underline{\hat{\boldsymbol{\lambda}}}}_1 \, \underline{\underline{\hat{\boldsymbol{\lambda}}}}_1^{\mathsf{T}} = \begin{pmatrix} 0.480 & 0.229 & 0.139 & 0.557 & 0.514 & 0.508 \\ 0.229 & 0.477 & 0.438 & 0.351 & 0.304 & 0.366 \\ 0.139 & 0.438 & 0.416 & 0.248 & 0.208 & 0.272 \\ 0.557 & 0.351 & 0.248 & 0.666 & 0.610 & 0.618 \\ 0.514 & 0.304 & 0.208 & 0.610 & 0.560 & 0.564 \\ 0.508 & 0.366 & 0.272 & 0.618 & 0.564 & 0.580 \end{pmatrix} .$$

- Also gilt

$$\hat{\psi}_1 = 0.520\,, \quad \hat{\psi}_2 = 0.523\,, \quad \hat{\psi}_3 = 0.584\,,$$
$$\hat{\psi}_4 = 0.334\,, \quad \hat{\psi}_5 = 0.440\,, \quad \hat{\psi}_6 = 0.420\,.$$

Mit dieser neuen Matrix $\underline{\underline{\hat{\boldsymbol{\psi}}}}$ kann ein neuer Schätzer $\underline{\underline{\hat{\boldsymbol{\lambda}}}}$ bestimmt werden .

Anzahl der allgemeinen Faktoren

Eine praktische Bestimmung der Anzahl k der Faktoren kann z.B. erfolgen als

- die Anzahl der Eigenwerte der empirischen Korrelationsmatrix, die größer als 1 sind;
- die Anzahl der positiven Eigenwerte der Matrix $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\boldsymbol{\psi}}}}$ im Algorithmus 6.2.5 5.;
- die kleinste Anzahl k, bei der die Summe der k größten Eigenwerte von $\underline{\underline{\mathbf{r}}} - \underline{\underline{\hat{\boldsymbol{\psi}}}}$ im Algorithmus 6.2.5 5. die Summe aller Eigenwerte übertrifft.
- Daneben kann man sich auch an den Ergebnissen einer Hauptkomponentenanalyse der (skalierten) Daten orientieren.

6.3 Rotation der Faktorladungsmatrix

Bem. 6.3.1

(i) Liegen Schätzwerte für die Faktorladungsmatrix und die Varianzen der spezifischen Faktoren vor, beginnt das Problem der Deutung des Ergebnisses. Dabei kann die Nicht-Eindeutigkeit der Faktorladungen (siehe Bem. 6.1.6) genutzt werden.

(ii) Eine Interpretation von Faktorladungen ist einfacher, wenn
- jedes Merkmal eine hohe Ladung wenigstens eines Faktors hat,
- jede Faktorladung entweder hoch oder nahe Null ist, d.h. es möglichst wenige „mittlere" Ladungen gibt. Man spricht auch von Faktorladungen mit einer sogenannten Einfachstruktur.

(iii) Eine Rotation muss nicht zur besseren Interpretation führen.

(iv) Zur Bestimmung einer geeigneten orthogonalen Matrix $\underline{\underline{\mathbf{g}}}$ können verschiedene Prinzipien genutzt werden, hier sollen kurz die Varimax-Rotation und die Quartimax-Rotation erläutert werden. Dabei wird in den Formeln auf die Kennzeichnung der Schätzungen durch $\hat{\cdot}$ verzichtet.

Varimax-Rotation

Bem. 6.3.2

(i) Gesucht wird eine transformierte Matrix der Faktorladungen $\underline{\underline{\boldsymbol{\Delta}}} = \underline{\underline{\boldsymbol{\lambda}}}\,\underline{\underline{\mathbf{g}}}$, die möglichst viele kleine oder große Ladungen hat.

(ii) Es sei $\underline{\underline{\boldsymbol{\Delta}}} = (\delta_{ij})_{i=1,\ldots,p;j=1,\ldots,k}$ und $d_{ij} = \dfrac{\delta_{ij}}{h_i}$, hier sind h_i, $i = 1, \ldots, p$, wieder die Kommunalitäten.

(iii) Bei der Varimax-Rotation wird die Matrix $\underline{\underline{\boldsymbol{\Delta}}}$ so bestimmt, dass die quadrierten Ladungen maximal streuen (die Varianzen der Ladungsquadrate eines Faktors werden maximiert), d.h. durch das Extremwertproblem

$$\sum_{j=1}^{k}\sum_{i=1}^{p}\left(d_{ij}^2 - \overline{(d_{\bullet j}^2)}\right)^2 \stackrel{!}{=} \max \quad \text{mit} \quad \overline{(d_{\bullet j}^2)} := \frac{1}{p}\sum_{i=1}^{p} d_{ij}^2 \,.$$

(iv) Die numerische Lösung der Optimierungsaufgabe erfolgt sukzessive für je zwei Faktoren.

Varimax-Rotation: Erster Schritt Optimierungsaufgabe

(v) Man beginnt mit $j = 1$ und $j = 2$.
Seien $\underline{\underline{\mathbf{g}}}_{(1\ 2)} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ die Rotationsmatrix mit dem Rotationswinkel θ und $\underline{\underline{\boldsymbol{\lambda}}}_{(1\ 2)}$ die auf die ersten zwei Spalten reduzierte Faktorladungsmatrix.

(vi) Dann ergeben sich die Elemente der ersten zwei Spalten der normierten und rotierten Faktorladungsmatrix zu $(i = 1, \ldots, p)$

$$d_{i1} = \frac{1}{h_i}(\lambda_{i1}\cos\theta - \lambda_{i2}\sin\theta)\,, \quad d_{i2} = \frac{1}{h_i}(\lambda_{i1}\sin\theta + \lambda_{i2}\cos\theta)\,.$$

(vii) Mit diesen d_{i1} und d_{i2} löst man die Optimierungsaufgabe

$$\sum_{i=1}^{p}\left(d_{i1}^2 - \overline{(d_{\bullet 1}^2)}\right)^2 + \sum_{i=1}^{p}\left(d_{i2}^2 - \overline{(d_{\bullet 2}^2)}\right)^2 \overset{!}{=} \max_{\theta}\,.$$

Varimax-Rotation: Fortsetzung Optimierungsaufgabe

(viii) Anschließend werden die dritte, vierte, ..., und letzte Spalte der unrotierten Faktorladungsmatrix $\underline{\underline{\boldsymbol{\lambda}}}$ mit der rotierten ersten Spalte auf obige Weise rotiert.

(ix) Daran anschließend wird die rotierte zweite Spalte mit der rotierten dritten, vierten, ..., rotierten letzten Spalte rotiert.

(x) Dieses Verfahren wird fortgesetzt, bis insgesamt $k(k-1)/2$ Rotationen durchgeführt sind.

(xi) Mitunter wird in (ii) auch die Skalierung durch die Kommunalitäten nicht durchgeführt.

Quartimax-Rotation

Bem. 6.3.3

- Bei der Quartimax-Rotation möchte man möglichst jedes Merkmal durch **wenige** allgemeinen Faktor erklären.
- Praktisch sucht man die Rotation, bei der die Summe der Varianzen der gewichteten Ladungsquadrate einer Variablen

$$\sum_{i=1}^{p}\sum_{j=1}^{k}\sum_{i=1}^{p}\left(\delta_{ij}^2 - \overline{(\delta_{i\bullet}^2)}\right)^2 \stackrel{!}{=} \max \quad \text{mit} \quad \overline{(\delta_{i\bullet})} := \frac{1}{k}\sum_{i=1}^{p}\delta_{ij}^2 .$$

maximiert wird.

6.4 Schätzung der allgemeinen Faktoren für einen Merkmalsvektor

- Bisher wurden für eine Datenmatrix $\underline{\underline{\mathbf{x}}}$ die Schätzungen $\underline{\underline{\ell}}$ für die Faktorladungsmatrix, $\underline{\underline{\psi}}$ für die Matrix der spezifischen Varianzen und $\underline{\overline{\mathbf{x}}}$ für die Erwartungswerte betrachtet.
- In bestimmten Situationen kann es aber auch von großem Interesse sein, für einen gegebenen konkreten Merkmalsvektor $\underline{\mathbf{x}}$ die zugehörigen Ausprägungen der allgemeinen Faktoren zu schätzen, d.h. einen $k-$dimesionalen Vektor $\underline{\hat{\mathbf{f}}}$, so dass gilt

$$\underline{\mathbf{x}} = \underline{\underline{\ell}}\,\underline{\hat{\mathbf{f}}} + \underline{\hat{\mathbf{u}}} + \underline{\overline{\mathbf{x}}}\,.$$

- Genutzt werden bei den Schätzungen oft der zentrierte Merkmalsvektor $\underline{\mathbf{y}} = \underline{\mathbf{x}} - \underline{\overline{\mathbf{x}}}$ oder der standardisierte Merkmalsvektor mit den Komponenten $y_i = \dfrac{x_i - \overline{x_i}}{s_{ii}}$.

Bartlett- und Thompson-Faktor-Schätzung

- Es gibt verschiedene Schätzmethoden für die Faktorwerte, z.B. basierend auf einer Regression oder die Schätzung nach Anderson-Rubin.
- Die Bartlett-Faktor-Schätzung nutzt die Formel

$$\underline{\hat{\mathbf{f}}} = \left(\underline{\underline{\ell}}^{\mathsf{T}}\underline{\underline{\boldsymbol{\psi}}}^{-1}\underline{\underline{\ell}}\right)^{-1}\underline{\underline{\ell}}^{\mathsf{T}}\underline{\underline{\boldsymbol{\psi}}}^{-1}\underline{\mathbf{y}}\,.$$

- Die Thompson-Faktor-Schätzung nutzt die Formel

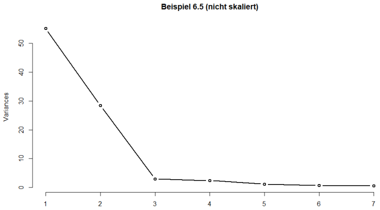
$$\underline{\hat{\mathbf{f}}} = \left(\underline{\underline{\ell}}^{\mathsf{T}}\underline{\underline{\boldsymbol{\psi}}}^{-1}\underline{\underline{\ell}} + \underline{\underline{\mathbb{I}}}_k\right)^{-1}\underline{\underline{\ell}}^{\mathsf{T}}\underline{\underline{\boldsymbol{\psi}}}^{-1}\underline{\mathbf{y}}\,.$$

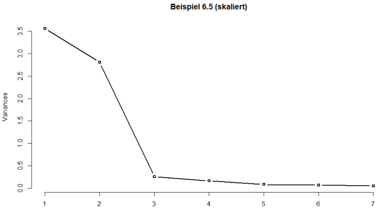
6.5 Beispiel Autokaufentscheidung

- Quelle: Horst Rinne, Statistische Analyse multivariater Daten : Einführung, R.Oldenbourg Verlag, München, Wien, 2000; Beispiel 5/1; siehe auch Vorlesungsskript Prof. Dr. Fred Böker, Georg-August-Universität Göttingen, Analyse mehrdimensionaler Daten, Kap. 5 Faktorenanalyse, `https://www.uni-goettingen.de/de/304527.html`.
- Vergebene Punktzahlen (zwischen 0 und 20) von 25 Käufern eines neuen Autos in einem Autohaus hinsichtlich der Wichtigkeit der Merkmale (in dieser Reihenfoge)
  - Anschaffungspreis,
  - Betriebskosten,
  - Umfang der Serienausstattung,
  - Styling der Karosserie,
  - Prestige der Marke,
  - Fahrkomfort,
  - Raumangebot.

Fortsetzung Bsp. 6.5

- Es gibt zwei Eigenwerte der Korrelationsmatrix größer als 1.
- Eine Hauptkomponentenanalyse der nicht skalierten und der skalierten Daten ergibt jeweils zwei Hauptkomponenten, die zusammen mehr als 91 % der Gesamtstreuung erklären, bei 3 Hauptkomponenten sind es jeweils mehr als 94 % der Gesamtstreuung.
- Die Scree-Plots ergeben auch die Nutzung von 2 oder 3 Faktoren, hier werden 2 gewählt.





2. Fortsetzung Bsp. 6.5

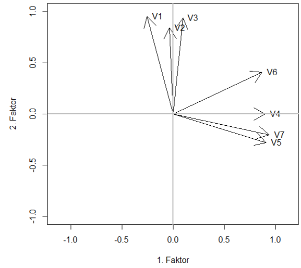
- In R kann eine Faktoranalyse z.B. mit dem Befehl `factanal()` durchgeführt werden. Dabei wird eine Normalverteilung der Daten angenommen und die Maximum-Likelihood-Faktor-Analyse mit Varimax-Rotation berechnet.
- Man erhält für die Daten die folgende Faktorladungsmatrix und ein entsprechendes Schema für die Faktorladungen betragsmäßig größer als 0.5 :

$$\underline{\underline{\ell}} = \begin{pmatrix} -0.2522 & 0.9539 \\ -0.0322 & 0.8431 \\ 0.0964 & 0.9383 \\ 0.8995 & -0.0007 \\ 0.9093 & -0.2814 \\ 0.8713 & 0.4080 \\ 0.9415 & -0.2049 \end{pmatrix}, \qquad \begin{array}{|c|c|} \hline & \text{x} \\ \hline & \text{x} \\ \hline & \text{x} \\ \hline \text{x} & \\ \hline \text{x} & \\ \hline \text{x} & \\ \hline \text{x} & \\ \hline \end{array} .$$

- Den ersten Faktor könnte man als „Produktdesign“, den zweiten als „Wirtschaftlichkeit“ bezeichnen.

3. Fortsetzung Bsp. 6.5

- Die Variablen können in Faktorraum wie folgt dargestellt werden:



- Eine Berechnung der THOMPSON-Faktorschätzwerte erfolgt z.B. durch `factanal(...,scores=''regression'')`.
- Eine Varimax-Rotation der ersten beiden Hauptkomponenten ergibt eine Ladungsmatrix, die sich nur wenig von der Ladungsmatrix bei der Maximum-Likelihood-Faktoranalyse mit Varimax-Rotation unterscheidet.