Multivariate Statistik, Übung 10

Henry Haustein

Aufgabe 1

(a) Die Kommunalitäten sind die durch die Faktoren erklärte Varianz einer Variablen.

(b) Mit dem Fundamentaltheorem der FA stellen die Faktoren ein statisches Modell der Variablen dar.

$$\begin{aligned}
R = \mathrm{Cov}(\mathcal{Z}) &= \mathbb{E}(\mathcal{Z}\mathcal{Z}') \\
&= \mathbb{E}((L\mathcal{F}+\mathcal{U})(L\mathcal{F}+\mathcal{U})') \\
&= \mathbb{E}(L\mathcal{F}\mathcal{F}'L) + \mathbb{E}(L\mathcal{F}\mathcal{U}') + \mathbb{E}(\mathcal{U}\mathcal{F}'L') + \mathbb{E}(\mathcal{U}\mathcal{U}') \\
&= L\mathbb{E}(\mathcal{F}\mathcal{F}')L' + L\mathbb{E}(\mathcal{F}\mathcal{U}') + \mathbb{E}(\mathcal{U}\mathcal{F}')L' + \mathbb{E}(\mathcal{U}\mathcal{U}') \\
&= LL' + \mathrm{Cov}(\mathcal{U})
\end{aligned}$$

(c) Bei einer Einfachstruktur der Faktoren gilt:

- Jede Zeile von L soll mindestens eine Null enthalten (Variable wird durch höchstens $k-1$ Faktoren beschrieben)
- Jede Spalte enthält mindestens q Nullladungen (Faktor beschreibt höchstens $k-q$ Variablen)
- Für jedes Spaltenpaar in L sollten nur wenige Variablen in beiden Spalten hohe Ladungen haben. (Im Beispiel ist dies Variable X_3)

(d) Es wird metrisches Skalenniveau benötigt, da wir auch die Differenzen zwischen Bewertungen interpretieren müssen.

Aufgabe 2

Ich werde hier immer den auf 2 Nachkommastellen gerundeten Wert angegeben, aber zusätzlich auch den exakten Wert, da sich dieser zum Teil deutlich von dem gerundeten Wert unterscheidet.

Zuerst testen wir mittels Bartletts-Test, ob wir überhaupt eine Faktorenanalyse machen sollten. Die Teststatistik ergibt sich zu

$$\begin{aligned}
\chi^2_{err} &= -\left(n - \frac{2\cdot k + 11}{6}\right)\cdot \ln\left(\prod_{i=1}^{4}\lambda_i\right) \\
&= -\left(6 - \frac{2\cdot 4 + 11}{6}\right)\cdot \ln(2.91\cdot 0.82\cdot 0.27\cdot 0.01 \quad \color{red}{2.906162898\cdot 0.815503170\cdot 0.271634856\cdot 0.006699077}) \\
&= -\left(6 - \frac{2\cdot 4 + 11}{6}\right)\cdot \ln(0.01 \quad \color{red}{0.004312668}) \\
&= 13.05 \quad \color{red}{15.4309}
\end{aligned}$$

Der kritische Wert ist $\chi^2_{1-\alpha;\frac{k(k-1)}{2}} = \chi^2_{0.95;6} = 12.59$ 12.5916, also wird die Nullhypothese abgelehnt und eine Faktorenanalyse ist sinnvoll.

Wir schätzen nun die Kommunalitäten mittels

$$h_j^2 = 1 - \frac{1}{r^{jj}}$$

Es ergibt sich

	Kommunalitäten (h_j^2)	**Einzelrestvarianzen** (u_j^2)
Preis	0.99 0.9893378	0.01 0.01066222
Nützlichkeit	0.25 0.25	0.75 0.75
Aussehen	0.97 0.9706669	0.03 0.02933313
Haltbarkeit	0.96 0.9604938	0.04 0.03950617

Da genau ein Eigenwert der reduzierten Korrelationsmatrix größer als 1 ist, wählen wir genau einen Faktor. Der Eigenwert ist $\lambda_1 = 2.82$ 2.820235871 und der zugehörige Eigenvektor ist $v_1 = (0.59, -0.25, 0.56, 0.52)$ $(0.5943829, -0.2517280, 0.5558792, 0.5237750)$. Damit ergibt sich die Ladungsmatrix zu $L = \sqrt{\lambda_1} \cdot v_1$:

$$L = \begin{pmatrix} 0.99 \\ -0.42 \\ 0.94 \\ 0.87 \end{pmatrix} \quad \begin{pmatrix} 0.9981804 \\ -0.4227409 \\ 0.9335190 \\ 0.8796046 \end{pmatrix}$$

Unser Faktor lädt also auf die Variablen Preis, Aussehen und Haltbarkeit hoch.¸