

## Übungen zur Vorlesung **Multivariate Verfahren**

— Blatt Nr. 7 —

### Aufgabe P14

Betrachten Sie erneut die Korrelationsmatrix aus Aufgabe P12:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_1 & 0 & 0 \\ \rho_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho_2 \\ 0 & 0 & \rho_2 & 1 \end{pmatrix}.$$

Es gelte  $1 \geq \rho_1 > \rho_2 > 0$ .

- Welche sinnvolle Anzahl an Faktoren lässt sich bei dieser Korrelationsmatrix vermuten? Welche möglichst einfache allgemeine Gestalt sollte die Ladungsmatrix  $\mathbf{L}$  haben, um die der Korrelationsmatrix zugrunde liegenden Daten gut zu beschreiben? Schlagen Sie eine Ladungsmatrix  $\mathbf{L}$  in Abhängigkeit möglichst weniger unbekannter Parameter vor.
- Versuchen Sie ein orthogonales Faktorenmodell für  $\mathbf{R}$  aufzustellen, d. h. suchen Sie (nur noch von  $\rho_1$  und  $\rho_2$  abhängige) Matrizen  $\mathbf{L}$  und  $\mathbf{V}$ , für die  $\mathbf{R} = \mathbf{L}\mathbf{L}' + \mathbf{V}$  gilt. Verwenden Sie dabei die in Aufgabenteil a) formulierte Ladungsmatrix  $\mathbf{L}$ . Ist das Modell eindeutig lösbar?
- Faktorenmodelle sind nur bis auf Rotationen festgelegt. Verwenden Sie die Matrix  $\mathbf{A}_\vartheta = \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$ , um eine weitere Darstellung des Faktorenmodells zu bestimmen.
- Bei vier beobachteten Variablen bedeuten zwei Faktoren eine Überparametrisierung des Faktorenmodells (siehe Bemerkung 5.14), die Parameter sind nicht identifizierbar. Geben Sie eine weitere Parametrisierung des Faktorenmodells an, d. h. andere Matrizen  $\mathbf{L}$  und  $\mathbf{V}$ , die nicht durch Rotation der in a) gefundenen Matrizen dargestellt werden können.

### Aufgabe H14 (5/10)

Gegeben sei der Datensatz `AufgabeH14.txt`, der Messungen zu den Merkmalen „Punkte“, „Tore“, „Gegentore“ und „Karten“ bei 18 Mannschaften einer Sportliga enthält. Eine Hauptkomponentenanalyse soll zur Dimensionsreduktion durchgeführt werden.

- Berechnen Sie Koeffizientenvektoren aller vier Hauptkomponenten.
- Bestimmen Sie die Anzahl der Hauptkomponenten nach den drei, in der Vorlesung vorgestellten, Kriterien, wobei beim Kriterium der totalen Variabilität 75% der Varianz erklärt werden soll.

- c) Interpretieren Sie die ersten beiden Hauptkomponenten.
- d) Bestimmen Sie die Kommunalität der Merkmale für die erste und für die ersten beiden Hauptkomponenten.
- e) Erstellen und interpretieren Sie einen Biplot der Individuen und Merkmale.

Zur Lösung dieser Aufgabe sollen keine vordefinierten Befehle zur Berechnung der Hauptkomponentenanalyse verwendet werden (andere Funktionen, z. B. zur Berechnung der Eigenwerte, dürfen aber verwendet werden).

### Aufgabe H15 (Fortsetzung H13) (3/10)

Betrachten Sie erneut die Situation aus Aufgabe H13, d.h. sei

$$\Sigma = \begin{pmatrix} 5 & 2 & 1 & 0 & 0 \\ 2 & 5 & 1 & 0 & 0 \\ 1 & 1 & 4 & 1 & 1 \\ 0 & 0 & 1 & 5 & 2 \\ 0 & 0 & 1 & 2 & 5 \end{pmatrix}.$$

- a) Wie viele Faktoren kann man hier für ein Faktorenmodell für  $\Sigma$  vor dem Hintergrund der Identifizierbarkeit der Parameter maximal sinnvoll wählen?
- b) Bestimmen Sie mittels der Hauptkomponentenmethode die initiale Zerlegung in Kommunalitäten und Restvarianzen für ein Faktorenmodell mit zwei Faktoren

Tipp: Nutzen Sie die Erkenntnisse aus H13.

### Aufgabe H16 (2/10)

Spearman hat in einer Arbeit von 1904 die Prüfungsergebnisse von Kindern in den Fächern *Classics* ( $X_1$ ), *French* ( $X_2$ ) und *English* ( $X_3$ ) untersucht. Dabei ergab sich folgende Korrelationsmatrix:

$$\hat{\mathbf{R}} = \begin{pmatrix} 1 & 0.83 & 0.78 \\ 0.83 & 1 & 0.67 \\ 0.78 & 0.67 & 1 \end{pmatrix}$$

Im Folgenden möchten wir mittels Faktorenanalyse die Dimension  $d = 3$  auf einen Faktor  $p = 1$  reduzieren. Finden Sie eine zu diesem Problem passende Ladungsmatrix  $\mathbf{L}$  und Restvarianzmatrix  $\mathbf{V}$ , sodass  $\hat{\mathbf{R}} = \mathbf{L}\mathbf{L}' + \mathbf{V}$  gilt. Ist  $\mathbf{L}$  eindeutig?

Abgabe der Hausaufgaben bis zum 02.12.2019 12 Uhr. Gruppenarbeit unter Angabe der Gruppenmitglieder möglich, aber pro Person eine eigene handschriftliche Abgabe. Bei Software-Aufgaben schicken Sie bitte **zusätzlich** zum ausgedruckten Blatt (R-Code inkl. Ergebnisse) den zugehörigen Programmcode an ihren Übungsleiter.

Übung	Mailadresse	Briefkasten
Mittwoch 10:15 - 11:45 Uhr	jkoenig@statistik.tu-dortmund.de	131
Donnerstag 08:30 - 10:00 Uhr	nilsjannik.schuessler@tu-dortmund.de	132