
Praktikumsaufgaben 7

WiMa-Praktikum

Aufgabe 1 Der Datensatz `usair` im Paket `gam1ss` enthält Informationen zur Luftverschmutzung mit SO₂ sowie zum Klima und Bevölkerung verschiedener amerikanischer Städte.

- Verschaffen Sie sich einen Überblick über die Daten indem Sie Scatterplots und Korrelationen aller Kombinationen der Variablen betrachten. Gibt es Ausreißer?
- Ignorieren Sie vorerst die SO₂ Variable und führen Sie "per Hand", d.h. ohne eine der in R vorhandenen Hauptkomponentenanalyse-Funktionen zu verwenden, eine Hauptkomponentenanalyse für den restlichen Datensatz durch. Die erste Hauptkomponente wurde von einem kreativen Autor als ein Index für die Lebensqualität bezeichnet. Können Sie Gründe für diese Interpretation finden? Gibt es eine anschauliche Interpretationen für die zweite Hauptkomponente?
- Wie viele Hauptkomponenten sind nötig um 90% der Gesamtvarianz zu erklären und wie viele schlägt der Scree Test vor?
- Plotten Sie die Werte der ersten und zweiten Hauptkomponente für die einzelnen Datensätze gegeneinander und erstellen Sie Boxplots der Hauptkomponenten. Lassen sich dadurch Ausreißer identifizieren?
- Untersuchen Sie mittels Regression wie gut sich die Luftverschmutzung durch die Hauptkomponenten vorhersagen lässt.

Aufgabe 2 Der `possum` Datensatz im Paket `DAAG` enthält morphometrische Messungen von Opossums.

- Führen Sie eine standardisierte Hauptkomponentenanalyse der Spalten 6-14 für den gesamten `possum` Datensatz, sowie für beide Geschlechter getrennt aus.
- Plotten Sie die Varianzen der Hauptkomponenten und führen Sie einen Scree Test für alle drei Szenarien durch.
- Plotten Sie die beiden ersten Hauptkomponenten für die weiblichen bzw. männlichen Opossums gegen die entsprechende Hauptkomponente für den Gesamtdatensatz. Gibt es auffällige Unterschiede?

Aufgabe 3 Es sei $N = (N_1, N_2)^T$ eine $\mathbf{B}(1, p, q)$ -verteilte Zufallsvariable, $p + q = 1$. Bestimmen Sie die Hauptkomponenten.