

Hauptkomponentenanalyse

Aufgabe 1:

In einer Erhebung wurden folgende persönliche Charakteristika (Datensatz: `individuum`) von 239 Teilnehmern erhoben:

- Groesse: Körpergröße in cm
- Schuh: Die Schuhgröße von 36 bis 52
- Gewicht: Körpergewicht in kg

Wir analysieren diesen Datensatz durch eine Eigenwertzerlegung der Kovarianzmatrix:

```
> eigen(cov(individuum))

$values
[1] 266.322975 28.256195 2.009739
$vectors
      [,1]      [,2]      [,3]
[1,] -0.5226376  0.8312910  0.1892224
[2,] -0.1569001  0.1243717 -0.9797520
[3,] -0.8379930 -0.5417443  0.0654283
```

und durch eine Hauptkomponentenanalyse der Korrelationsmatrix:

```
> prcomp(individuum, scale=TRUE)

Standard deviations:
[1] 1.6154090 0.4909864 0.3865048

Rotation:
      PC1      PC2      PC3
Groesse 0.5809216 -0.4715234  0.66347255
Schuh    0.5846446 -0.3254109 -0.74316787
Gewicht  0.5663222  0.8196179  0.08663539
```

- Beschreiben Sie die vorliegenden Outputs und gehen Sie dabei auf Gemeinsamkeiten und Unterschiede der beiden Ergebnisse ein.
- Welcher Anteil der Varianz wird jeweils von den ersten beiden Komponenten erklärt?
- Zeichnen sie einen Screeplot zur zweiten Analyse und entscheiden Sie wieviele Komponenten Sie auswählen würden.

Aufgabe 2:

Laden Sie den Datensatz `miete03.asc` von der Homepage herunter und importieren Sie diesen in R. Betrachten Sie im Folgenden lediglich die Variablen `nm`, `wfl`, `nmqm` und `bj`.

- a) Bestimmen Sie die Eigenwerte und Eigenvektoren der Kovarianzmatrix mit der Funktion `eigen()` und führen Sie eine Hauptkomponentenanalyse (PCA) mit der Funktion `prcomp()` durch. Vergleichen Sie dann die Outputs.
- b) Plotten sie einen Screeplot und visualisieren Sie die Loadings. Warum ist die in a) durchgeführte PCA nicht sinnvoll?
- c) Wie würden Sie Teilaufgabe a) sinnvoll lösen?