Hauptkomponentenanalyse

Aufgabe 1:

Sei $\mathbf{x} = (x_1, x_2)^T$ ein Zufallsvektor mit Kovarianz

$$\Sigma_{\mathbf{x}} = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

- Bestimmen Sie die Eigenwerte λ_1 und λ_2 und die (normierten) Eigenvektoren der Matrix $\Sigma_{\mathbf{x}}$.
- Berechnen Sie mit Hilfe von a) einen Zufallsvektor $\mathbf{y} = (y_1, y_2)^T$, dessen Komponenten y_1 und y_2 lineare Kombinationen von x_1 und x_2 sind und der folgende Kovarianz hat:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Aufgabe 2:

Der Datensatz `europa.txt` enthält für 24 europäische Länder die Variablen Oberfläche (`ober`), Einwohnerzahl (`einw`), Bruttonozialprodukt (`brut`) und Arbeitslosigkeit (`arbl`) (vgl. Blatt 7)

- Bestimmen Sie Erwartungswertvektor, Kovarianz- und Korrelationsmatrix der Daten.
- Warum ist es nicht sinnvoll, die Hauptkomponentenanalyse ohne Skalierung der vier Variablen, d.h. auf der Basis der Kovarianzmatrix durchzuführen?
- Machen Sie sich mit der Funktion `princomp()` vertraut und führen Sie damit eine Hauptkomponentenanalyse bzgl. der Korrelationsmatrix durch.
- Stellen Sie die Varianzen graphisch mit Hilfe der Funktionen `screeplot()` dar.
- Interpretieren Sie die vier Hauptkomponenten (Funktion `loadings()`) mit Hilfe von vier Balkendiagrammen, mit denen Sie das Gewicht jeder ursprünglichen Variable in jeder Hauptkomponente darstellen.
- Die R-Funktion `eigen()` berechnet die Eigenwerte und Eigenvektoren einer Matrix. Für die Korrelationsmatrix liefert sie folgendes Ergebnis:

```
$values [1] 1.9465661 1.4235518 0.3731246 0.2567576
```

```
$vectors
```

```
      [,1]      [,2]      [,3]      [,4]  
[1,] -0.4959586 -0.4927052  0.6173002 -0.3608423  
[2,] -0.4842364 -0.4975888 -0.7066954  0.1360225  
[3,]  0.4361680 -0.5963492  0.2420757  0.6289074  
[4,] -0.5738445  0.3924470  0.2468229  0.6751046
```

Wie groß ist der Anteil der Gesamtstreuung, der durch die einzelnen Hauptkomponenten erklärt wird?

- Wieviele Hauptkomponenten halten Sie für notwendig?
- Wie werden für jedes Land die Werte, die ihm durch die ersten beiden Hauptkomponenten zugewiesen werden, berechnet? Berechnen Sie diese für die Schweiz und Bulgarien. Interpretieren Sie die Ergebnisse.