Exercise 2

Problem 1: Principal Component Analysis

Download the file DECATHLON.txt into your R-workspace. The file contains the results of 48 decathletes from 1973. Familiarize yourself with the data and perform the covariance matrix based PCA transformation.

a) Familiarize yourself with the function *princomp* from the package *stats*.

b) How much of the variation of the original data is explained by k principal components, where $k = 1, 2, ..., 10$.

c) Choose a sufficient amount of principal components and try to interpret them.

d) Calculate the sample mean and covariance matrix from the principal components.

Problem 2: Eigendecomposition of a symmetric matrix

Let A be a symmetric matrix with distinct eigenvalues. Show that the eigenvector matrix of A is orthogonal.

Home Exercise 2: PCA for simulated data

Simulate 2000 observations from bivariate normal distribution with parameters:

$$\mu = \left(4, 7\right), \qquad \Sigma = \begin{pmatrix} 10 & 6 \\ 6 & 8 \end{pmatrix}.$$

(a) Plot the data.

(b) Apply the covariance based PCA transformation to the data set.

(c) Plot the transformed data. Use the same scale as in (a).

(d) Compare the plots and describe the differences.

(e) Plot the direction of the first principal component to the original data.