

Computer Assignment 1

Examining multivariate data

Course code and name:	732A97 (2021) Multivariate Statistical Methods
Examiner:	Maryna Prus
Instructions:	Submit your report as a PDF file. All R-codes should be included into your report. The report should be written in English. The report should be handed in via LISAM or alternatively in case of problems emailed to maryna.prus@liu.se This assignment is a part of the examination.
Submission deadline:	23:59 on 7-th of February 2021

Recommended literature

Chapters 1, 3 in Johnson, Wichern, Int. Ed. (Chapters 1, 2 in 6-th Ed.)

Chapters 1, 2 in Everitt, Hothorn

For R codes: The Little Book of R for Multivariate Analysis

(<https://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/>).

Data Set

For all problems consider the data set in the file T1-9.dat, which includes the national track records for women. For 55 countries national records for 7 variables (x_1 : 100 m; x_2 : 200 m; x_3 : 400 m; x_4 : 800 m; x_5 : 1500 m; x_6 : 3000 m and x_7 : marathon) are given. Use R to perform the following data analyses.

Problem 1

Describe the individual variables in the following way:

- Compute the means, the variances and the standard deviations for all variables.
- Illustrate the variables using box-plots and histograms. For the graphs use suitable labels, front sizes, etc. Do the variables look normally distributed? Justify your answer.

Problem 2

Investigate the relationships between the variables as follows:

- Compute the covariance and correlation matrices for the 7 variables.
- Illustrate the relations between variables for the 6 pairs: $(x_1; x_2)$, $(x_2; x_3)$, $(x_3; x_4)$, $(x_4; x_5)$, $(x_5; x_6)$ and $(x_6; x_7)$, using scatterplots. Do you observe some extreme values?
- Which other plotting possibilities for multivariate data you know? Present at least one of them for the given data set. Why did you choose this graph?

Problem 3

In problem 2, b) you observed some extreme values. Which countries look the most extreme? One of the possibilities to answer this question is to compute a distance between an observation

and the sample mean vector (to look how far an observation is from the average). Compute the Euclidean distances of observations from the sample mean for all countries. Which 3 countries are the most extreme?