Assignment 3
Principle component and factor analysis

Kurskod och namn:	732A97 Multivariate Statistical Methods
Delmomentsansvarig:	Krzysztof Bartoszek
Instruktioner:	This assignment is part of the examination for the Multivariate Statistical Methods course You will work in groups of 2–4. Submit your report as a **.PDF** file Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments. **All code (R) should be included as an appendix into your report.** A typical report should contain 2–4 pages of text plus some amount of figures plus appendix with codes. In the report reference **ALL** consulted sources and disclose **ALL** collaborations. The report should be handed in via LISAM (or alternatively in case of problems e–mailed to krzysztof.bartoszek@liu.se), by **23:59 15 December 2017** at latest. **Late submission may result in an additional penalty assignment.** The report can be written in English or Swedish. Notice there is a final deadline of **23:59 4 February 2018** after which no submissions nor corrections will be considered and you will have to redo the missing labs next year.

Assignment developed by Ann–Charlotte Hallberg and Bertil Wegmann.

Learning objectives

After reading the recommended text and doing the assignment the student shall be able to:

- identify different structures in covariance matrices
- use suitable software for data reduction
- interpret estimated linear components (factors)
- handle a software for factor analysis and interpret the essential parts of the output

Recommended reading

Chapters 8–9 in *Johnson, Wichern*
Chapters 3, 5 in *Everitt, Hothorn*
For R code: the Little Book of R for Multivariate Analysis
(`https://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/`).

If one is not interested in location (i.e. the means) but scales and relationships then the information in a multivariate data set can be condensed in the covariance (or correlation if the scales are not relevant) matrix. For normally distributed data all such information is contained in this matrix. Therefore analyses of covariance matrices are essential in multivariate statistical methods.

Often the variables are strongly correlated and most of the variability (in other words signal) can be explained by some few linear combinations of the variables. And hopefully, these linear combinations can be interpreted in a meaningful (from the application's perspective) way. Sometimes, we have a (factor) model for our data, a model resulting in a special structure of the covariance matrix. Then the goal is to estimate this structure and relate it to the observed covariance matrix.

Question 1: Principal components, including interpretation of them

Solve Exercise 8.18 of *Johnson, Wichern*. The data on the national track records for women, which you have studied earlier, can be found in the file `T1-9.dat`.

Question 2: Factor analysis

Solve Exercise 9.28 of *Johnson, Wichern*, the same data as above. Try both PC and ML as estimation methods. Notice that R's `factanal()` only does ML estimation. For the PC method you can use the `principal()` function of the `psych` package. What does it mean that the parameter `rotation` of `factanal()` is set to `"varimax"` by default (equivalently `rotate` of `principal()`)? Do not forget to check the adequacy of your model
Tip: Read section "A Large Sample Test for the Number of Common Factors".