Assignment 4
Canonical correlation 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 19 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:

- formulate the association concept between two variable sets and the simplification of this due to the canonical correlations concept
- use suitable R software for canonical correlation analysis, identify and implement missing functionality
- use the output to interpret the canonical correlations and canonical variates
- validate the results from the output

Recommended reading
Chapter 10 in *Johnson, Wichern*
Chapter 3.13 in *Everitt, Hothorn*
`https://cran.r-project.org/web/packages/yacca/yacca.pdf`
`https://cran.r-project.org/web/packages/CCA/CCA.pdf`
`?cancor`
`https://cran.r-project.org/web/packages/vegan/vegan.pdf`, parts on canonical correlation analysis
`https://stats.idre.ucla.edu/r/dae/canonical-correlation-analysis/`

Question: Canonical correlation analysis by utilizing suitable software

Look at the data described in Exercise 10.16 of *Johnson, Wichern*. You may find it in the file `P10-16.DAT`. The data for 46 patients are summarized in a covariance matrix, which will be analyzed in R. Read through the description of the different R packages and functions so you may chose the must suitable one for the analysis. Supplement with own code where necessary.

a) Test at the 5% level if there is any association between the groups of variables.

b) How many pairs of canonical variates are significant?

c) Interpret the "significant" squared canonical correlations.
Tip: Read section "Canonical Correlations as Generalizations of Other Correlation Coefficients".

d) Interpret the canonical variates by using the coefficients and suitable correlations.

e) Are the "significant" canonical variates good summary measures of the respective data sets?
Tip: Read section "Proportions of Explained Sample Variance".

f) Give your opinion on the success of this canonical correlation analysis.