Exercise 8

Problem 1: Canonical Correlation Analysis

Upload CAR.txt into your R-workspace. It contains several international car manufacturers that have been evaluated in eight different categories. The scores are from 1 (very good) to 6 (very bad). Perform the canonical correlation analysis and use the following groups:

$X =$ **(Price, Value)**,
$Y =$ **(Economy, Service, Design, Sport, Safety, Easy h.)**.

Note that cars with score 6 for the variable **Value**, lose their value the fastest over time.

(a) Compute the sample canonical vectors ($\hat{\alpha}_k$ and $\hat{\beta}_k$) with corrected scaling.

(b) Compute the score vectors corresponding to the canonical vectors and the sample canonical correlations.

(c) Interpret the first pair of canonical variables and plot the corresponding scores.

(d) Repeat (c) for the second pair of canonical variables.

Problem 2: Course assignment

There will be no proof in todays exercise session. Instead, we will speak about the course assignment.

Homework Assignment 8: Canonical Correlation Analysis

Perform the canonical correlation analysis to the data DECATHLON.txt. Find the relationships between the groups

$Y =$ **(R100, LONG_JUMP, SHOT_PUT, HIGH_JUMP, R400, H110, DISCUS_THROW, JAVELIN, POLE_VAULT, R1500M)**,
$X =$ **(HEIGHT, WEIGHT)**.

Repeat the steps (a) - (c) from Problem 1. In (a) give the correctly scaled canonical vectors and in (b) the sample canonical correlations explicitly, not just the code. In addition, provide an answer to the following.

(e) How many pairs of canonical variables can we obtain?

(f) What happens to the results if we instead choose the groups as

$X =$ **(R100, LONG_JUMP, SHOT_PUT, HIGH_JUMP, R400, H110, DISCUS_THROW, JAVELIN, POLE_VAULT, R1500M)**,
$Y =$ **(HEIGHT, WEIGHT)**?