Exercise 8

Problem 1: Canonical Correlation Analysis

In today's exercise we have a data set related to cars. We have 24 international car manufacturers and they have been evaluated in eight different categories. The scores are from 1 (very good) to 6 (very bad). Perform the canonical correlation analysis to find relationships between the groups X =(Price, Value) and Y =(Economy, Service, Design, Sport, Safety, Easy h.).

a) How many pairs of canonical variables can we obtain?

b) Compute the canonical vectors (α_k and β_k) with corrected scaling. Give canonical variables u_1, v_1, u_2 and v_2.

c) Compute the score vectors corresponding to the canonical variables. Examine the sample correlation structure related to the vector $(u_1, u_2, v_1, v_2)^T$.

d) Give an interpretation of the first pair of canonical variables and plot the corresponding scores.

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

f) Consider the following subset of variables (Price) and (Economy, Easy handling). Compare the strength of the relationship between the canonical variables with the ones obtained in (a) - (e).

g) Compare the result of (f) to classical linear regression (L^2).

Problem 2: Course assignment

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

Home Exercise 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) and X = (HEIGHT, WEIGHT). Repeat the steps (a) - (e) from Problem 1.