Multivariate Data Analysis
Prof. Dr. Christina Andersson

Exercise Sheet 5: Canonical Correlation Analysis

Computer Problems:

1. A researcher has collected data on three psychological variables, four academic variables (standardized test scores) and gender for 600 college freshman. She is interested in how the set of psychological variables relates to the academic variables and gender. In particular, the researcher is interested in how many dimensions (canonical variables) are necessary to understand the association between the two sets of variables.

   We have a data file, *mmreg*, with 600 observations on eight variables. The psychological variables are locus_of_control, self_concept and motivation. The academic variables are standardized tests in reading (read), writing (write), math (math) and science (science). Additionally, the variable female is a zero-one indicator variable with the one indicating a female student.

   (a) Confirm the dimensions of *mmreg*.

   (b) The psychological variables are locus of control, self-concept and motivation. The academic variables are standardized tests in reading, writing, math and science. Additionally, the variable female is a zero-one indicator variable with the one indicating a female student.

   (c) Apply some descriptive statistics to the data.

   (d) Perform a canonical correlation analysis. Display the canonical correlations.

   (e) Display the raw canonical correlations.

   (f) Interpret the raw canonical correlations.

   (g) Display the loadings of the variables on the canonical dimensions (variates).

   (h) In general, the number of canonical dimensions is equal to the number of variables in the smaller set; however, the number of significant dimensions may be even smaller. Find the number of significant dimensions!