HW\_08

**Exercise 1**

Do a correspondence analysis for the car-ratings (file cars.txt). Explain how this table can be considered as a contingency table.  
The data are averaged ratings for 24 car types from a sample of 40 persons. The marks range from 1 (very good) to 6 (very bad).

**Exercise 2**

Write an R function to compute the chi-square statistic of independence. Test the null using for the bachelor data (file bachelors.txt). The data consists of observations of 202,100 bachelors from France and give the frequencies for different sets of modalities classified into regions. The variables (modalities) are: A=Philosophy-Letters, B=Economics and Social Sciences, C=Mathematics and Physics, D=Mathematics and Natural Sciences, E=Mathematics and Techniques, F=Industrial Techniques, G=Economic Techniques, H Computer Techniques.

**Exercise 3**

Do correspondence analysis of the U.S. crime data (file UScrime.txt), and determine the absolute contributions for the first three axes. How can you interpret the third axis? Try to identify the states with one of the four regions to which it belongs. Do you think the four regions have a different behavior with respect to crime? This is a data set consisting of 50 measurements of 7 variables. It contains the number of crimes in the 50 states of the U.S. classified according to 7 categories. Region is 1 for Northeast, 2 for Midwest, 3 for South and 4 for West. Division is 1 for New England, 2 for Mid Atlantic, 3 for E N central, 4 for W N Central, 5 for S Atlantic, 6 for E S Central, 7 for W S Central, 8 for Mountain and 9 for Pacific.

**Exercise 4**

Consider the food data (file food.txt). Given that all of the variables are measured in the same units (dollars), explain how this table can be considered as a contingency table. Perform a correspondence analysis and compare the results to those obtained with the PCA analysis of the correlation matrix. The data set consists of the average expenditures on food for several different types of families (manual workers = MA, employees = EM, managers = CA) with different numbers of children (2,3,4 or 5 children).