Demo Problem 1: Multiple Correspondence Analysis

Install the package `ca`. The data set `tea.txt` contains answers of a questionnaire on tea consumption for 300 individuals. The following questions were asked:

1) What kind of tea do you drink? (black, green, flavored)
2) How do you drink it? (alone, w/milk, w/lemon, other)
3) What kind of presentation do you buy? (tea bags, loose tea, both)
4) Do you add sugar? (yes, no)
5) Where do you buy it? (supermarket, tea shops, both)
6) Do you always drink tea? (always, not always)

Perform correspondence analysis to the data set and interpret the results.

Demo Problem 2: The trace of matrix V

Let V be the matrix defined as in lecture slides 7. Show that

$$\text{Trace}(V) = \frac{K}{P} - 1.$$

Homework Problem 1: Multiple Correspondence Analysis

The data set `attitudes.txt` contains the attitudes of 871 individuals towards science and the environment. Each category contains five possible answers (strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, coded as 1 to 5). The questions are:

A. We believe too often in science, and not enough in feelings and faith.
B. Over all, modern science does more harm than good.
C. Any change humans cause in nature - no matter how scientific - is likely to make things worse.
D. Modern science will solve our environmental problems with little change to our way of life.

In addition, the data set contains three demographic variables (sex, age and education). Variables age and education have 6 categories. For the variable age, 1 indicates that the individual belongs to the youngest age group. Likewise, for the variable education, 1 indicates the lowest level of education. Furthermore, the individuals are categorized as either 1=male or 2=female. Perform MCA using the function `mjca(..., lambda="indicator")`. Provide the requested answers/figures.

(a) Find a combination of two MCA components that explain as much of the variation as possible. What is the combination and how much of the total variation is explained by these two components?

(b) Produce the MCA graph with respect to the components chosen in (a).

(c) What is the relationship between education and strong opinions (strongly agree/strongly disagree) in this data set? Justify!