Exercise 5

Problem 1: Bivariate Correspondence Analysis

The data set SCIENCEDOCTORATES.txt contains the number of doctors graduated from different fields of science. The data is from USA between the years 1960-1975.

Find the attraction repulsion matrix related to bivariate correspondence analysis and interpret the results.

Problem 2: Some Basic Results

a) Let x and y be continuous random variables with probability density functions f_x and f_y respectively. Furthermore, let x and y be independent. Show that for any functions g and h:

$$\mathbb{E}\left[g(x)h(y)\right] = \mathbb{E}\left[g(x)\right]\mathbb{E}\left[h(y)\right],$$

assuming that the expected values exists.

b) Let X be a p variate random variable with a cumulative distribution function F_x. Show that $\text{Cov}\left(x\right)$ is positive semidefinite.

Home Exercise 5: Bivariate Correspondence Analysis

The data SMOKING1.txt contains a 2-dimensional frequency table, where the employees of a firm have been categorized according to their position (5 categories: SM = Senior Managers, JM= Junior Managers, SE = Senior Employees, JE = Junior Employees, SC = Secretaries). Furthermore, the smoking of the employees have 4 categories (None, Light, Medium, Heavy).

a) Form the theoretical frequencies under independence.

b) Calculate the attraction repulsion matrix.

c) Interpret the results. Is smoking more frequent among a specific group in this company?