

# Homework 2

2024-05-17

## Correspondence Analysis and Cluster Analysis

**The homework should not exceed 5 pages. Preferably, you can use R Markdown. Please submit both the script and the pdf file.**

Import the data set “food-consumption.csv” that is related to consumption of certain food items in 16 different European and Scandinavian countries.

The data set consists of 20 variables. The numbers represent the percentage of the population consuming that food type.

The objectives of the study are:

- (i) to analyze relationships between countries and consumed food types.
- (ii) to cluster countries based on food consumption types.

### 1. Application of Correspondence Analysis.

- a) Remove observations with missing values using function `na.omit()`. (1p)
- b) Apply correspondence analysis by using `CA()` function in `FactoMineR`. (2p)
- c) How many dimensions would you need to extract? Why? (5p)
- d) Plot the row and column profiles on the extracted dimensions.(5p)
- e) Interpret the plot(s) considering (12p)
  - (i) similarities between countries,
  - (ii) similarities between food categories.
  - (iii) the relationship between food categories and countries.

### 2. Application of Cluster Analysis.

- a) Compute a distance matrix. (2p)
- b) Apply hierarchical cluster analysis to group countries according to their food consumption and plot dendrogram. (Try different methods but in the report just show the one that you prefer. Explain why you have chosen that method.) (3p)
- c) How many clusters do you think there are? (Considering the dendrogram plot) (2p)
- d) Draw an elbow plot showing within sum of squares per dimension. Interpret it. (3p)
- e) Apply k-means clustering according to the number of groups you have chosen. (2p)
- f) Compute the centroids (means) of the clusters and show them in a table. (3p)
- g) Interpret the most important characteristics of each cluster according to the summary table given in section (e). (10p)