

# Project 1

---

## Data Analysis and Statistical Modelling

(ADME, 1<sup>st</sup> Semester, 2023/2024)

Handed out on 02 Dec, 2023.

To be handed back on 22 Dec, 12pm, 2023.

Consider for the **BudgetItaly** data frame, available in R library **Ecdat**, the subset with observations for **year 73** of all variables except the variable **pfood**.

1. Explore/describe the data applying methods taught in this course, in particular using plots and summary statistics (e.g. mean, median, trimmed and winsorized mean, variance, mad, covariance, generalized/total variance and Mahalanobis distances) and discuss what you have learned from this preliminary analysis.
2. With the aim of dimension reduction:
  - (a) Apply principal components analysis (PCA):
    - i. Considering the variables in original scale and the classical sample covariance estimate;
    - ii. Considering the standardized variables.
  - (b) Which of the two previous analyses do you recommend? Justify your choice based on the percentage of total variance explained by both sample principal components. Interpret the sample principal components you have chosen to retain.
3. Introduce five potential outliers in the data set, replacing the first five observations by 0.01 times the original values. Apply to the new data set (without standardization):
  - (a) the classical PCA;
  - (b) the robust PCA based on the MCD estimate.

Discuss the effect of atypical observations in both analysis.

### About the report:

The report should not exceed 10 pages (with Annexes). Do not forget to include: introduction, the dataset in study, objectives of study, decisions, conclusions and bibliography. The R code and the report must be upload in fenix.