Data Analysis

Research Project

The analysis should be done based on the data collected by students or downloaded from open sources. The dataset should contain not less than 10 variables with different scales of measurement and not less than 100 cases. If needed, preliminary data transformation could be done using recoding, calculation of new variables based on existing variables, filters, aggregation, ranking, etc.

The students should conduct a research project applying all the knowledge and skills acquired during the course:

- Formulate the goal, objectives, and hypotheses of the research (at least 10 hypotheses).
- Perform a descriptive data analysis (frequency analysis, descriptive statistics, graphs).
- Calculate new variables based on the existing ones (if applicable).
- Explore relationships between variables (contingency tables, Chi-squared tests, correlations, t-tests, nonparametric tests).
- Conduct ANOVA (analysis of variance), PCA and cluster analysis.
- Build and interpret regression models (preferably several, but at least one). Each regression model must include at least 5 predictors.
- Formulate general conclusions based on the results of the study.

If certain data analysis methods are not applicable to the selected dataset, evidence of their inapplicability should be provided, along with an explanation of why these methods are not appropriate for the research.

The work can be done individually or in a group of two students. The evaluation criteria are identical in both cases. The main results of analysis should be presented on the slides (not less than 10 slides).

By **23:55 December 3** any group member should upload the following files to Smart LMS:

1. Jupyter Notebook containing the code used for the analysis.
2. PowerPoint presentation summarizing the main results of the data analysis.
3. The data file used in the analysis.

In December each group will present the report. Each group member must present a part of the report and be ready to answer the questions.