

Project requirements

The task is to find a proper data set. The data set should contain at least 2 qualitative and 2 quantitative variables and minimum 100 observations. You will clean and prepare data for analysis (missing data issue – imputation; outlier problem etc.). The following contents are required:

- Short description about variables, type of variables, scales of measurement
- Missing values, patterns – method of imputation
- Descriptive statistics for numeric variables: mean, median, quartiles, deviation, skewness, kurtosis – informative figures about it. If stratifying is informative and useful, please calculate some statistics and depict graphs for strata.
- Ratios, frequencies and figures for qualitative variables.
- Confidence interval estimation for variables.
- Analysis of relation between variables (crosstabs, Anova, covariance and correlation)
- At least 3 rational hypothesis testing for variables (different group means, ratios etc.)

Appearance

- Font size: 12
- Times New Roman
- Justified
- Nice figures with titles and all the information which are needed. (not pixelated)
- Structure: Introduction, Dataset description, Analysis, Satisfying conclusion (not 3 lines...!)

Required length of the essay is at least 20000 characters. We ask a short presentation (10 minutes) at last week.

In hidden Teams channels you have to report about the process week by week.

Deadline: 2021.10.29. 23:55