PRACTICAL ASSIGNEMENT

Data Selection

- The data set to be analysed should consist of a group of entities (persons, companies, countries, regions, …) described by a set of variables, at a given point in time – that is, cross-sectional data.
- Do not select time-series data (data observed along time).
- The number of entities should be larger than the number of variables.
- The size of the data array depends in the topic, but : the number of entities should not be larger than 100-200, nor smaller than, say, 30, and the number of variables not larger than 20 nor smaller than 8-10.
- Register your data in an Excel file first.
- If you have counting data, for instance, nb. students in a given region, this should be transformed to relative data, dividing by the total population of the respective region – to avoid a "dimension" effect.

Introduction

Describe briefly the data, and refer its source. Explain the "questions" you have. Explain which methods you will be using to answer your "questions".

Data

Describe the data into more detail, detailing the variables, their meaning and type.

Univariate analysis

Descriptive statistics of ALL variables.

Indicators (position, dispersion, form..), graphic representations, analysis of (possible) outliers.

Position : mean, trimmed mean, median – and comparison between them ; quartiles,…

Dispersion: range and inter-quartile range, variance and standard deviation, coefficient of variation (for variables that do not change sign)

Shape : skewness, kurtosis

Please try to be compact, and not repetitive: use global tables when possible, and then comment on individual results.

Bivariate analysis

Analysis of correlations for numerical variables.

Contingency tables for some pairs (considered of particular interest) of categorical variables.

Statistical tests (e.g. comparison of means/medians between populations) if pertinent.

Multivariate analysis

Multivariate analysis of the data set, using methods learnt in the course (not necessarily all of them !)

Conclusion

Refer the main conclusions drawn from your analysis'.

References