EnQuireR: exploration of questionnaires with R

J. Bouche[1], M. Cadoret[1,*], G. Fournier[1], O. Fournier[1], S. Lê[1], F. Le Poder[1]

1. AGROCAMPUS OUEST, 65 rue de Saint-Brieuc, CS 84215, F-35042 Rennes cedex

* Contact author: marine.cadoret@agrocampus-ouest.fr

Keywords: categorical variables, univariate data analysis, multivariate data analysis, clustering, semantic markup

The use of categorical variables is commonplace in many fields (consumer market studies, politics, health, food science and so on), and in particular in making surveys. As surveys are becoming increasingly popular, there is a growing need for statistical methods including categorical variables.

The main objective of the **EnQuireR** package is to automate the survey process. This package will perform univariate and multivariate data analyses. Those two levels of analysis provide the user a range of functions to improve decision-making aid. Until now, multivariate analysis of categorical variables was performed for instance by the R package **FactoMineR**. Unlike **FactoMineR**, the **EnQuireR** package focuses on one type of applications, *i.e.* the statistical analysis of questionnaires and hence on categorical variables mainly, and allows:

- a faster way to perform the survey process, or any dataset including categorical variables;
- the display of many different outputs including both numerical results and graphs which are precious tools for decision-making aid;
- an easier view of the results by the automatic generation of a *.pdf* report and of a *Beamer* type presentation via the use of **Sweave**.

This package targets a wide range of users from students to scientists and is designed to be accessible to anyone with a basic knowledge of statistics. During the talk we will first present the univariate analysis methods then some methodologies dedicated to multivariate analysis.

The **EnQuireR** package contains the following functions:

- Bar plots: the function **ENbarplot()** can be used to obtain bar plots either sorted by alphabetical order or by bar sizes for each variable with the percentage of missing values. The function **XvsYbarplot()** allows to obtain a bar plot for a variable depending on another variable.
- Distance between variables: the function **chisq.desc()** can be used to measure the statistical relationship between categorical variables. The χ^2 test is used to measure this relationship.
- Multiple Correspondence Analysis (MCA): **missmca()** performs MCA with missing values, **ENlisib()** can be used to improve the graph readability by suppressing the display of objects with a poor quality of representation and **ENMCA()** is used to do cluster analysis following MCA.
- **ENellipse()** draws confidence ellipses around the categories of a variable of interest.
- The function **ENmark()** performs a semantic markup with one, two or three levels.

References

S. Lê, J. Josse and F. Husson (2008). FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, 25 (1). pp. 1-18.

F. Leisch (2002). Sweave, Part I: Mixing R and LaTeX: A short introduction to the Sweave file format and corresponding R functions. *R News*, 2 (3). pp. 28-31.