

MAINT.DATA: Modeling and Analysing Interval Data in R

A. Pedro Duarte Silva^{1,*}, Paula Brito²

1. Faculdade de Economia e Gestão & CEGE, Universidade Católica Portuguesa at Porto, Porto, Portugal

2. Faculdade de Economia & LIAAD-INESC Porto LA, Universidade do Porto, Porto, Portugal

*Contact author: psilva@porto.ucp.pt

Keywords: Symbolic data, Interval data, Parametric modelling of interval data, Statistical tests for interval data, Skew-Normal distribution

In the classical model of multivariate data analysis, data is represented in a $n \times p$ data-array where n “individuals” (usually in rows) take exactly one value for each variable (usually in columns). Symbolic Data Analysis (Diday and Noirhomme-Fraiture (2008), Noirhomme-Fraiture and Brito (2011)) provided a framework where new variable types allow to take directly into account variability and/or uncertainty associated to each single “individual”, by allowing multiple, possibly weighted, values for each variable. New variable types - interval, categorical multi-valued and modal variables - have been introduced. We focus on the analysis of interval data, i.e., where elements are described by variables whose values are intervals of \mathbb{R} . Parametric inference methodologies based on probabilistic models for interval variables are developed in Brito and Duarte Silva (2011) where each interval is represented by its midpoint and log-range, for which Normal and Skew-Normal (Azzalini and Dalla Valle (1996)) distributions are assumed. The intrinsic nature of the interval variables leads to special structures of the variance-covariance matrix, which are represented by five different possible configurations.

In this work, we introduce the package MAINT.DATA, which implements the proposed methodologies in R. It introduces a data class for representing interval data. MAINT.DATA includes functions for modeling and analysing interval data, in particular maximum likelihood estimation and statistical tests for the different considered configurations. Methods for (M)ANOVA and Linear and Quadratic Discriminant Analysis of this data class are also provided.

References

- Azzalini, A. and Dalla Valle, A. (1996). The multivariate Skew-Normal distribution, *Biometrika* 83 (4), pp. 715–726.
- Brito, P. and Duarte Silva, A.P. (2011). Modelling interval data with Normal and Skew-Normal distributions. *Journal of Applied Statistics*, (in press).
- Noirhomme-Fraiture, M. and Brito, P. (2011). Far Beyond the Classical Data Models: Symbolic Data Analysis. *Statistical Analysis and Data Mining* (in press).
- Diday, E. and Noirhomme-Fraiture, M. (2008). *Symbolic Data Analysis and the SODAS Software*, Wiley, Chichester.