The FSDA MATLAB toolbox: an integrated framework to assess and apply robust methods to complex datasets

Francesca Torti

[1] *European Commission, Joint Research Centre, Ispra, Italy; francesca.torti@ec.europa.eu*

1 Summary of the contribution

Flexible Statistics for Data Analysis toolbox (FSDA, http://rosa.unipr.it/fsda.html, see Riani et al., 2012) is an open source statistical library providing a rich variety of robust and computationally efficient methods for the analysis of complex data, possibly affected by outliers and other sources of heterogeneity. It offers interactive tools for exploratory data analysis. It provides several state of the art robust estimation methods in:

- *Regression*: Least Trimmed Squares (LTS), Least Median of Squares (LMS), M, MM, S (see Maronna, et al., 2006), the Forward Search by Atkinson, et al. (2000);
- *Multivariate analysis*: Minimum Covariance Determinant (MCD), Minimum Volume Ellipsoid (MVE), MM, S (see Maronna, et al., 2006), the Forward Search by Atkinson, et al. (2004);
- *Clustering*: Trimmed KMEANS (TKMEANS), TCLUST and Robust Linear Grouping Analysis (RLGA) by Garcìa-Escudero et al. (2010).
- *Time Series*: Least Trimmed Squares Time Series (LTSts) and Forward Search Time Series (FSRts) by Rousseeuw, et al. (2018).

It offers also principled routines for the evaluation and calibration of clustering methods, based on the generation of synthetic data under controlled settings (see Riani et al., 2015), using the MIXSIM framework of Maitra and Melnykov (2010).

FSDA is the results of the joint work of many researchers from different institutions, among which the University of Parma and the Joint Reseach Centre of the European Commission.

References

Atkinson, A. C. and Riani, M. (2000). Robust Diagnostic Regression Analysis. Springer-Verlag, New York.

Atkinson, A. C., Riani, M. and Cerioli, A. (2004). Exploring Multivariate Data with the Forward Search. Springer-Verlag, New York.

Maronna R. A., Martin R.D., Yohai V.J. (2006). *Robust statistics: theory and methods*. Wiley, Chichister.

Melnykov V., Maitra R. (2010). Finite mixture models and model-based clustering. *Statist. Surv.* 4, 80–116.

Riani M., Perrotta D., Torti F. (2012). FSDA: A MATLAB toolbox for robust analysis and interactive data exploration. *Chemometrics and intelligent laboratory systems*, 116, p. 17-32.

Riani M., Cerioli A., Perrotta D., Torti F. (2015). Simulating mixtures of multivariate data with fixed cluster overlap in FSDA. *Advances in data analysis and classification*. Volume 9, Issue 4, pp 461-481.

Rousseeuw P.J., Perrotta D., Riani M., Hubert M. (2018). Robust Monitoring of Many Time Series with Application to Fraud Detection. *Econometrics and Statistics*.