Robustness in practice

P. Filzmoser[1]*

[1] *Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria; P.Filzmoser@tuwien.ac.at*
[*] *Presenting author*

Keywords. *Robust statistics; Applications; R; robustbase.*

1 Abstract

During the last decades, a lot of effort has been made in the development of robust statistical methods. Nowadays, many robust counterparts to traditional non-robust methods are available, and there is a lot of knowledge about the theoretical properties of these methods (**?**). Moreover, robust methods have been widely implemented in software packages, for example in the statistical software environment R (**?**). Thus, the availability of those methods could even lead to the hypothesis that robust tools will at some point replace the traditional tools. However, so far we cannot see this tendency.

Thinking about regression analysis, several robust counterparts to least-squares regression have been developed, and they have been implemented for example in the R package `robustbase` (**?**). We usually teach our students that if no outliers are present in the data (and if the usual data requirements are met), robust regression leads to about the same answer as least-squares regression; however, in presence of outliers, least-squares regression results could be heavily spoiled. This is often demonstrated with simulations and with "real" data, which are frequently "toy" data sets, or data sets which served as test data also in other scientific papers. One could thus ask if real "real" data show the same features. Is there still a striking difference in the analysis when using non-robust and robust methods? Is it even possible to use the code implementation of the robust methods in real applications without any difficulties? Do the parameters of the routines require special adaptations or tuning? Is the runtime of these algorithms comparable to the runtime of algorithms for traditional methods?

We will present some use-cases and focus on different robust methods for outlier

detection, regression and discrimination. Available routines in R are tested, and results (if any) are compared to non-robust counterparts. Recommendations for practical use are given, and possible directions for future developments are provided.

References

Maronna, R., Martin, D. & Yohai, V. (2006). Robust Statistics: Theory and Methods. John Wiley & Sons, Chichester.

R Core Team (2015). R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. Vienna, Austria. `https://www.R-project.org/`.

Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M. & Maechler, M. (2015). robustbase: Basic Robust Statistics. R package version 0.92-5, `http://CRAN.R-project.org/package=robustbase`.