

The data collected through business surveys is as a rule plagued with outliers, it is incomplete (missing values and non-response), usually skewed, and semi-continuous distributions occur often. The multivariate aspect of the data collected in business surveys makes the task of outlier identification particularly challenging. The outliers can be completely hidden in one or two dimensional views of the data.

Nevertheless, not many studies of outlier detection methods for business survey data based on robust methods are found in the literature, Todorov, Templ and Filzmoser (2011) and Bill and Hulliger (2015) are two that we can mention. These two studies present two R packages (rrcovNA and modi) implementing several reliable methods for outlier detection and present comparisons of these methods.

In this presentation new methods, added to the R package rrcovNA will be presented and their performance will be investigated using the real data set included in the 'modi' package as well as a data set based on the African Investment Survey (2011) carried by UNIDO.