Election of an Optimal Trimming Level for Trimmed Means Family

J.Fco. Ortega[1]

[1] Universidad de Castilla-La Mancha. Facultad de CC. Económicas y Empresariales de Albacete. Plaza de la Universidad 1. 02071 Albacete, Spain.

Keywords: Estimators of Location, Trimmed Means, Efficiency and Robustness.

1 Abstract

In most of the procedures proposed in Robust Statistics, not only in the definition of Robust Estimators but also in the Methods of Outliers Detection, Estimators of Location are used as elements in their construction, so that the good qualities and the behaviors of such procedures depend on the behavior of these estimators. In the bibliography there has different families of Estimators of Location such as the known as M-Estimators, L-Estimators and R-Estimators and some other (which cannot be included within these big families) that have been carefully studied, although very little used, in general.

One of the quoted families, the L-Estimators, contains a subfamily known as Trimmed Means whose elements use in their definition a parameter (α) so that the Trimmed Means of parameter α results from the mean of the sample observations eliminating the α% of the biggest observations and the α% of the smallest. That is, the Trimmed Means use a part of the sample, eliminating, in a symmetrical way, a certain amount of extreme observations.

This parameter α, that we will term Trimming Level, conditions the behavior of the element of the family. Besides, it defines this element under certain desirable proprieties. In a more concrete way, its election determines; on the one hand, the behavior of the estimator that generates the advisable conditions for the researches carried in presence of Outliers (Robustness); on the other hand, thatch parameter also determines the behavior before the Efficiency of the estimator of the Trimmed Means family, to which it defines. That is, the elements of the Trimmed Means family go through the possible categories of both proprieties, where the extreme cases are formed by the Mean, in which the only observation of the sample is able to have influence, in such a way, on its calculation, that can lead it to any desirable value, so it has the lowest possible levels of Robustness, whereas its Efficiency is the optimal, and regarding the Median, whose resistance or sensibility towards possible Outliers is the maximum, that is, this estimator is not affected if there is a set too small compared to the whole of Outliers, being its Efficiency (Efficiency related to the Mean) of the 64%.

Definitively, the election of a suitable Trimming Level, in any particular situation in which it is advisable to use the Trimmed Means, is of high interest, because this one will determine the behavior of the estimator that has originated under Efficiency and Robustness.

In this paper, we have realizad a detailed study of the family of Estimators of Location, Trimmed Means, showing its main characteristics and proposing a procedure for the election, in some sense, of the best Trimming Level in every possible situation. Therefore, the research starts with the exposition of the Trimmed Means family and the description of its main statistical characteristics. In the next section, we focus on the proposal of an optimal Trimming Level, starting with the establishment of the possible values of this parameter and lately we have dealt with the election of the most suitable one, under some hypothesis. Finally, we show a set of examples used in previous researches, for the study of Estimators of Location in contaminated samples, to which we have applied the previously introduced theories, and some final comments, as a kind of summary.

References

D.F. Andrews et al. (1972). *Robust Estimates of Location*. Princeton University Press.

V. Barnett and T. Lewis (1994). *Outliers in Statistical Data*. Wiley & Sons.

F.R. Hampel et al. (1986). *Robust Statistics: The approach based on influence functions*. Wiley & Sons.

D.C. Hoaglin et al. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley & Sons.

J.Fco. Ortega (2000). *Nuevas Familias de Estimadores Robustos y Detección de Observaciones Atípicas en Modelos Lineales*. Tesis Doctoral. Univ. de Castilla-La Mancha (Spain).

P.J. Rousseeuw and C. Croux (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, No. 424.

P.J. Rousseeuw and A.M. Leroy (1987). *Robust Regression and Outlier Detection*. Wiley & Sons.

S.M. Stigler (1977). Do robust estimators work with real data. *The Annals of Statistics*, Vol.5, No.6.

Please fill in this form and mail it together with your abstract.

My abstract fits best to topic number: 2; 8; 17.

List of Topics:

1. Algorithms
2. Applications
3. Biostatistics
4. Computing and graphics
5. Data analysis
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression
21. Time series analysis
22. Wavelets
23. Other (please specify)