



Contribution ID: 59

Type: **not specified**

## A Model-Robust Subsampling Approach in Presence of Outliers

*Monday, 11 September 2023 13:50 (20 minutes)*

### Abstract

In the era of big data, several sampling approaches are proposed to reduce costs (and time) and to help in informed decision making. Most of these proposals require the specification of a model for the big data. This model assumption, as well as the possible presence of outliers in the big dataset, represent a limitation for the most commonly applied subsampling criterions.

The task of avoiding outliers in a subsample of data was addressed by Deldossi et al. (2023), who introduced non-informative and informative exchange algorithms to select “nearly”D-optimal subsets without outliers in a linear regression model. In this study, we extend their proposal to account for model uncertainty. More precisely, we propose a model robust approach where a set of candidate models is considered; the optimal subset is obtained by merging the subsamples that would be selected by applying the approach of Deldossi et al. (2023) if each model was considered as the true generating process.

The approach is applied in a simulation study and some comparisons with other subsampling procedures are provided.

### References

Deldossi, L., Pesce, E., Tommasi, C. (2023) Accounting for outliers in optimal subsampling methods, Statistical Papers, <https://doi.org/10.1007/s00362-023-01422-3>.

### Keywords

Active learning, D-optimality, Subsampling

### Classification

Both methodology and application

**Primary authors:** Prof. DELDOSSI, Laura (Università Cattolica del Sacro Cuore); Prof. TOMMASI, Chiara (University of Milan)

**Presenter:** Prof. DELDOSSI, Laura (Università Cattolica del Sacro Cuore)

**Session Classification:** CONTRIBUTED Design of Experiments 1

**Track Classification:** Design and analysis of experiments