

The identification of outliers in generalized linear models

Nihan Acar and Aydin Erar

Mimar Sinan Fine Arts University, Turkey

Abstract

Generalized linear models are used in the analysis of exponential distribution families such as Normal, Binomial, Poisson and Gamma distributions where a simple linear relationship need not be found between response variable and explanatory variables. In the analysis of data including observations that differ from main part of the data, identification of outliers is a necessary step to obtain valid results. The outlier identification problem is actually the problem of identifying those observations that lie in an area called the outlier region. There are several ways of identifying outliers. Outlier detection methods go into division depending upon number of outliers and reliability of results. According to number of outliers one step or stepwise methods can be used. To avoid masking and swamping effects robust methods are preferred rather than non-robust methods. In this study, we focus on two types of outlier identification rules and compare obtained results.

Keywords

Generalized linear models, Outliers, Outlier region.

References

- Asikgil, B. and Erar, A. (2009). Examination of multiple outliers in generalized linear models. *The Announcement Book of the 6th Statistics Days Symposium*, 16–26.
- Dobson, J.A. (1990). *An Introduction to Generalized Linear Models*. Chapman and Hall.
- Davies, L. and Gather, U. (1993). The identification of multiple outliers. *J. Amer. Statist. Assoc.* 88, 782–792.

Kuhnt, S. and Pawlitschko J. (2005). Outlier identification rules for generalized linear models. In: Baier, D., Wernecke, D. (Eds.), *Innovations in Classification, Data Science and Information Systems* (pp. 165–172). Heidelberg: Springer.

Nelder, J.A. and Wedderburn W.M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A.* 134, 370–384.