



The Influence of Multicollinearity in Binary Logistic Regression and Additive Models

Mustafa Cavus*

Department of Statistics Science Faculty Anadolu University, Eskisehir, Turkey –
mustafacavus@anadolu.edu.tr

Betul Kan Kilinc

Department of Statistics Science Faculty Anadolu University, Eskisehir, Turkey –
bkan@anadolu.edu.tr

Abstract

The generalized linear modeling has appeared very useful in researches to create regression models with any distribution of the dependent variable. It can be count, binary or ordinal. The binary logistic regression is a generalized linear model having a logit link function. In the present study we examine and compare from a statistical point of view the binary logistic models and additive logistic models in the presence of multicollinearity. As the collinearity inflates the variances of the parameter estimates and therefore it results incorrect inferences about the relationships between explanatory and response variables, etc. It also leads to instable estimated coefficients in generalized additive models. For this purpose, highly correlated explanatory variables with a binary response variable are fitted by generalized linear model and generalized additive logistic models (GALM) in a conducted simulation study for moderate to larger sample sizes. The results are compared by the existing two approaches using simulated datasets. Finally, their advantages and drawbacks are pointed out in conclusion.

Keywords: multicollinearity; logistic; additive.