Robust logistic diagnostic for the identification of high leverage points in logistic regression model

ABSTRACT

High leverage points are observations that have outlying values in covariate space. In logistic regression model, the identification of high leverage points becomes essential due to their gross effects on the parameter estimates. Currently, the distance from the mean diagnostic method is used to identify the high leverage points. The main limitation of the distance from the mean diagnostic method is that it tends to swamp some low leverage points even though it can identify the high leverage points correctly. In this study, we propose a new diagnostic method for the identification of high leverage points. Robust approach is firstly used to identify suspected high leverage points by computing the robust mahalanobis distance based on minimum volume ellipsoid or minimum covariance determinant estimators. For confirmation, the diagnostic procedure is used by computing the group deleted potential. We called this proposed diagnostic method the robust logistic diagnostic. The performance of the proposed diagnostic method is then investigated through real examples and monte carlo simulation study. The result of this study indicates that the proposed diagnostic method ensures only correct high leverage points are identified and free from swamping and masking effects.