

COMPARISON OF THE PREDICTIVE PERFORMANCE OF LIU-BASED ESTIMATORS.

T. Kayathiri^{1*}, M. Kayanan², and P. Wijekoon³

¹Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka.

²Department of Physical Science, Faculty of Applied Science, University of Vavuniya, Vavuniya, Sri Lanka.

³National Institute of Fundamental Studies, Kandy, Sri Lanka.

*skayathiri1994@gmail.com

Logistic regression is one of the most widely used statistical methods for predicting a binary outcome by analysing the relationship between one or more existing independent variables. Although the maximum likelihood estimation technique is a commonly used method to estimate the parameters, their predictive performances may be affected by a problem called multicollinearity. To reduce the effect of multicollinearity, different biased estimators have been proposed as alternatives to the Maximum Likelihood Estimator (MLE). According to the literature, the superiority of the existing estimators based on Liu estimators was examined using the Mean Square Error Matrix (MSEM) and Scalar Mean Square Error (SMSE) criteria. However, the researchers did not compare the prediction performance of these existing estimators. Therefore, the present study is aimed to compare the prediction performance of Liu-based estimators in logistic regression using balanced accuracy. The prediction performance of the Maximum Likelihood Estimator (MLE), Logistic Liu Estimator (LLE), Almost Unbiased Liu Logistic Estimator (AULLE), and Modified Almost Unbiased Logistic Liu Estimator (MAULLE) are considered for comparison. To evaluate the balanced accuracy of the above estimators, the dataset was split into two so that 70% belongs to the training set and 30% to the test set. The model was trained using the training set, then the testing set was used to evaluate the balanced accuracy. A Monte Carlo simulation study was done to understand the prediction accuracy by setting different levels of correlation among the predictors and sample sizes. Further, a myopia real-world dataset was utilised, and it was observed that the related results tally with the results of the simulation study. Finally, it was noticed that the estimator MAULLE has the best prediction performance when multicollinearity is present, and then LLE performs well. Additionally, the prediction performance of AULLE was significantly better for some selected values of shrinkage parameters. However, the MLE performs well with small sample sizes.

Keywords: Balanced accuracy, Logistic regression, Multicollinearity, Prediction performance.