C-statistic for evaluating the added predictive ability in the binary risk models: An alternative to the existing approaches

Afrin Sadia Rumana*
Institution of Statistical Research and Training,University of Dhaka, Dhaka, Bangladesh - arumana@isrt.ac.bd

Mohammad Shafiqur Rahman
Institution of Statistical Research and Training,University of Dhaka,Dhaka, Bangladesh - shafiq@isrt.ac.bd

Abstract

Logistic regression models are frequently used in various settings of clinical research to predict the risk of a patient's future health status such as death or illness using his/her clinical and demographic characteristics. Predictions based on these models have an important role in classifying the patients with low-and high-risk and hence in guiding their future courses of treatment. Given their important role in clinical research, it is very essential to evaluate the predictive performance of the model e.g the ability of the model to distinguish between low-and high-risk patients-which is termed as 'discrimination'. Concordance statistic (C-statistic) is frequently used to quantify the discriminatory power of the logistic models. A particular problem of interest in this area is to quantify the added value of a set of influential predictors in the predictive performance of the model. Several proposals of C-statistics has been discussed in the literature, however none of these is able to translate the statistical significance of a new predictor into statistical significant improvement in the predictive performance of the model. More specifically, these C-statistics are not sensitive enough to the inclusion of additional predictors in the model, which leads to misleading conclusion on model's predictive/discriminatory performance. To address this problem, this paper proposed an alternative estimation for C-statistics. The new C-statistic is based on the concordance probability definition proposed by Gonen and Heller for Cox PH models, which quantifies the actual predictive value (or risk difference) added by the new predictor in the model. The method is illustrated by an application to low birth weight data. Further a simulation study was conducted to assess the performance of the new and existing C-statistics and compared the results. The results showed that the new C-statistic was more sensitive to quantify the added predictive value of the model.

Keywords: Logistic regression, C-index, ROC curve, Discrimination, Prediction, Predictive accuracy.

*Presenting author