



Estimation and Analysis of Factors Affecting Diabetes Using Full and Stepwise Logistic Regression Models: A Comparative Study

Elnazeer Mohamed Elnoor

IBN SINA University

Corresponding Author : ac.aff.isu@gmail.com

Abstract:

This study aims to evaluate the performance of a predictive model of diabetes using a classification table; a complete logistic regression model (8 variables) and a graduated logistic regression model (5 variables) were studied. The performance of the two models was analyzed using various metrics including weighting logarithm, determination coefficients (Nagelkerke and Cox & Snell), Hosmer and Lemeshow test, Omnibus test, and rating accuracy. The results showed a slight superiority of the complete model in the overall classification accuracy (78.3% versus 77.5% for the graduated model). The full model also outperformed the following: Data relevance: Lower weighting logarithm (723.445 vs. 728.560). Explanatory power: higher determination coefficients (0.408 and 0.296 vs. 0.402 and 0.292). Statistical significance: Chi-square value is higher in the Omnibus test (270.039 vs. 264.924). While the graduated model showed slight superiority in the Hosmer and Lemeshow test (higher p-value: 0.421 vs. 0.403), this did not compensate for the full model's models, but lower sensitivity, which means it is difficult to identify the actual infected. The study concluded that the complete model is generally better, emphasizing the importance of including all possible variables. However, both models still need to be improved, especially in sensitivity, by collecting additional data or using sophisticated modeling techniques. The study also recommends the use of additional assessment measures such as precision and the F1 coefficient for a more comprehensive assessment. In short, the complete model offers better performance, but it needs improvements, especially in identifying actual casualties to reduce negative errors and ensure accurate diagnosis.

Keywords: *diabetes, prediction, classification model, classification table, specific sensitivity, negative error.*