Multiple regression modelling for mathematics performance: best model selections

ABSTRACT

Initiating mastery of mathematics in primary school is pivotal for successful learning at higher levels. Multiple regression analysis stands as a cornerstone in statistical methods for modelling mathematical achievement. However, despite its prevalence, earlier studies often neglect to disclose the essential assumptions requisite for effective multiple regression modelling. Moreover, the impact of variable selection methods on model generation and subsequent identification of the optimal model remains insufficiently explored. Considering these gaps, this study was undertaken to identify significant factors influencing students' mathematics achievement while ensuring adherence to multiple regression analysis assumptions. Utilizing demographic data, the number of books, home educational resources, student attitudes, and mathematics anxiety as independent variables, two models were derived: Model 1 incorporated all variables without domains, while Model 2 included domainspecific variables and adhered to multiple regression assumptions. The findings revealed that the Model 2 is the best model since it has highest , adjusted , lowest standard error of estimation, lower values in 8 selection criteria which also fulfilled assumptions of multiple regression analysis. In conclusion, key determinants of mathematics achievement were identified as the number of books (101-200), student confidence, and mathematics learning anxiety. The constructed model elucidated 27.6% of the variance in mathematics achievement. This study underscores the importance of meeting regression test assumptions for modelling accuracy and provides actionable insights for schools to design interventions aimed at enhancing mathematics achievement among fifth-year students and the broader elementary school population.