Abstract No: 149 ***ICT, Mathematics, and Statistics***

COMPARISON OF REGRESSION METHODS FOR MODELLING LONGITUDINAL DATA

L.D.M.A. Dhaneshwara* **and M.S. Atapattu**

Department of Statistics and Computer Science, Faculty of Science,
University of Peradeniya, Sri Lanka
**s17338@sci.pdn.ac.lk*

This study conducted a detailed comparative analysis of three regression methods suitable for modelling longitudinal data: linear regression, quantile regression, and rank-based regression, utilizing both Monte Carlo simulations and an empirical dataset. The findings from the simulation study yield several key insights. For smaller sample sizes, linear regression, and rank-based regression demonstrate greater accuracy compared to quantile regression. Although the performance of all three methods improves with increasing sample sizes, quantile regression consistently lags behind. Furthermore, the number of independent variables significantly influences regression accuracy in smaller datasets, but this effect diminishes as the sample size grows. The empirical analysis employs the "beach water quality" dataset as a case study, partitioning it into training and test sets to evaluate the performance of the three regression methods. The results indicated that the linear regression model achieves the lowest Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) on both training and test data indicating superior performance. Conversely, the rank-based regression model exhibits the highest MSE and RMSE across both datasets, reflecting comparatively poorer performance. The results suggest that the linear regression model is the most reliable for longitudinal data, consistently outperforming the other two models in the scenarios considered.

Keywords: Linear regression, Longitudinal data, Rank-based regression, Quantile regression