

## **The Performance of Robust Estimator on Linear Regression Model Having both Continuous and Categorical Variables with Heteroscedastic Errors**

**Habshah Midi and Bashar A. Talib**

*Laboratory of Applied and Computational Statistics*

*Universiti Putra Malaysia*

*43400 Serdang, Selangor, Malaysia*

*E-mail: habshahmidi@hotmail.com*

### **ABSTRACT**

The ordinary least squares (OLS) technique is often used in practice to estimate the parameters of a multiple linear regression model with both continuous and categorical variables. It has been the most popular technique due to its optimal properties and ease of computation. Nevertheless, in the presence of outliers, the OLS can result in very poor estimates. Outliers which arise from bad data points may have undue effect on the OLS estimates. The problem is further complicated when both outliers and heteroscedasticity or non-constant error variances are present in the data. The influence of outliers and heteroscedasticity cannot be removed or reduced by simply transforming the data using known transformation such as logarithmic transformation. In this paper, we proposed a robust technique to deal with these two problems simultaneously. A robust estimate of scales for each level of categorical variables are first estimated by using robust distance S and M (RDSM) estimates. Then we determine the weighting scheme for each level of the categorical variables and transform the model. The reweighted least squares based on RDSM (RLSRDSM) is then applied to the transformed model. The empirical evidence shows that the proposed method has reduced the heteroscedastic effect to a greater extent.

**Keywords:** outliers, heteroscedasticity, robust Distance, RDLI, S/M estimates, RDSM.