ABSTRACT

One of the most important statistical tools is a linear regression analysis for many fields. A problem that we often encounter in the application of regression is the presence of an outlier or outliers in the data. They make serious effects of statistical modeling/inference. Even one outlying observation can destroy least squares estimation, resulting in parameter estimates that do not provide useful information for the majority of the data. Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is, and whether this should be thrown out. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. M estimator (Huber, Hampel and Bi-square), MM estimator, LAD, LMS, LTS and S estimator are being used for this research. The properties of efficiency, breakdown point, and bounded influence are used to define the measure of robust technique performance in a theoretical sense. Efficiency can tell us how well a robust technique performs relative to least squares on clean data (without outliers). High efficiency is mostly desired on estimation. The breakdown point is a measure for stability of the estimator when the sample contains a large fraction of outliers. The comparison of the finite sample performance of these estimators is carried out by a bootstrapping procedure with used repetition 1000 for analysis. The results show that the influential points are the major cause of disturbance, the regression line estimated, under OLS is more affected. On the other hand the further regression line are least effected and when we compare the robust regression lines, the more precise estimators are M-estimators by Bi-square by using efficiency analysis and LMS and LTS are also giving more suited regression lines due to not effected by outliers. MSE's of particular model is the smaller then the variance of the beta for this respective model is also smaller for a given sample size. We have analyzed that the M-estimator by bi-square is the most efficient with respect to variances of beta's and MSE's of the data. When n=50 the results are more precise because there is not contain any outliers. When we increase sample size then it contain two outliers in the data and comparing the robust estimates efficiency with respect to OLS, they are more precise then OLS in the presence of outlier.