

# The impact of adjusting turning constant of bisquare psi function on some robust estimators in the presence of outliers

<sup>1</sup>Rasheed, B. A. and <sup>2</sup>Dalah, M. C.

<sup>1</sup>Department of Mathematics and Statistics Gombe State University, Gombe

<sup>2</sup>Department of Mathematics and Statistics Yobe State University, Damaturu

## Paper History

Received: 04<sup>th</sup> Jan, 2024

Accepted: 28<sup>th</sup> May, 2024

Published: May, 2024

## Abstract:

Turning constant value is often used in practice to provide resistance of outlying data in regression model. Outliers which arise from bad data points may have unduly effect on regression estimators. The influence of outliers cannot be removed or reduced by simply transforming the data using known transformation such as logarithmic transformation. This paper, proposed to examine the impact for adjusting a turning constant value of robust S-estimator, robust M- estimator and robust MM estimators in the presence of outliers in regression model using a Bisquare redescendent weighted function. The turning constant value of each estimator will be adjusted by simple modification of increasing or reducing the value to improve the resistant of robust estimators on outlying data. Simulation study was used to compare the performance of the robust methods with adjusting a turning constant value on different sample size. The result of the study showed that increasing or decreasing the turning constant values have significant impact on S- estimators and M estimators on the resistant of outliers. The result also disclosed that, adjusting the turning constant value for MM-estimator will not affect the resistant to outlying data. The study recommends to use smaller value of turning constant for M-estimation to have more resistance against outlying observations in the regression models. Secondly the study also recommended using a value greater than the default K value for the S-estimation against outlying observations.

Corresponding author

Rasheed, B. A.

[arasheedbello@yahoo.com](mailto:arasheedbello@yahoo.com)

**Keywords:** MM-estimation, M-estimation, Outlier, S-Estimation, Turning constant

## 1. Introduction

Regression analysis is conceptually method for determining functional relationships between the dependent and a set of independent variables. The relationship between y and  $x_1, x_2, \dots, x_n$  can be approximated by the regression model [1,2]:

$$y = f(x_1, x_2, \dots, x_n)\beta + e_{ij} \quad (1)$$

Where  $e_{ij}$  is assumed to be a random error representing the discrepancy in the approximation, it accounts for the failure of the model to fit the data exactly. According to Shrestha, [3], to judge how well the estimated regression model fits the data, we can look at the size of the residuals, which is,

$$e_{ij} = y_i - (f(x_1, x_2, \dots, x_n)\beta) \quad (2)$$

A point which lies far from the line and often has a large residual value is called an outlier. However, it is well known that the OLS estimate is extremely sensitive to the outliers. However, regression modeling is guided by a number of assumptions. When the analyst estimates regression models and tests its assumptions, the frequently finds that the assumptions are substantially violated then, the parameter estimated poorly fits the data. Often,

however, a transformation will not eliminate the leverage of influential outliers that bias the prediction and distort the significance of parameter estimates. When such occur, robust regression that is resistant to the influence of outliers may be the reasonable recourse [3, 4, and 5]. The robust methods have been defined to deal with the influential points in regression analysis, where the value of the estimation by using this method is not much affected with outliers. The characterization of robust methods is that they fit the bulk of the data well if the data contain outliers, while if a small proportion of outliers are present the robust method gives approximately the same results as the classical (non-robust) method. As a consequence of fitting the bulk of the data well, robust methods provide a very reliable method of detecting outliers, even in high dimensional multivariate situation [6, 7, 8, 9].

Based on the literature review, we have noticed that most of the previous researchers on this topic stick to the used of default turning constant value C or K of the redescendent weights for the respective robust regression methods despite the fact that the C or K value can be adjusted to evaluated the performance of some robust estimators against OLS for the Measurement of a Topographic Altimetric Network, where they adjusted the constant value for M-estimation method from the default

value 4.685 to 1.345 which led to increases in the break down point of the M estimator. Therefore, this study intends to examine what will happen when there is a change on the default values of the tuning constant. This is where I based my research since the values of the tuning constant for the redescending weight can be set (apart from the pre-set default), this study aimed to verify whether there is an improvement in the high breakdown point modeling when the values are increase or decrease. In this research, a situation when all the estimators are operating under the same redescending weight was studied with adjusting the tuning constants while measuring the performance of such operators and verifying which one outperforms others. The aim of this study is to compare the performance of M, S and MM robust estimators in the presence of outliers using a default and adjusted turning constant value of the redescending weights and also to compare the performance of the robust estimators with default and adjusted turning constant values respectively. The findings of this study will serve as an alternative for researchers to choose between the default and adjusted turning constant value for the robust estimators at different sample sizes.

## **1.1 Robust Regression**

Robust regression is a regression method that is used when the distribution of residual is not normal or there are some outliers that affect the model [4]. This method is an important tool for analyzing the data which is affected by outliers so that the resulting models are stout against outliers [5]. A robust regression is an iterative procedure that is designed to overcome the problem of outliers and influential observations in the data and minimize their impact over the regression coefficients [5]. The main objective of robust estimation is to obtain reliable estimates/inferences for unknown parameters in the presence of outliers. The robust procedure replaces the sum of squared residuals of the OLS with some other function that is being less influenced by the unusual observations. These procedures first fit a regression to the data and then identify the outliers as those observations having large residuals. Robust techniques have three desiring properties, namely, efficiency, breakdown point, and bounded influence. The breakdown point is the smallest fraction of the unusual observations that an estimator can tolerate before giving an incorrect result. It is always a value between zero and 0.5. It measures the degree of robustness, the robustness of an estimator increases as the breakdown point increases. For example, OLS has a breakdown point of 0% which represents that even a single outlier is sufficient to distort the OLS estimators. The robust techniques have 50% of breakdown point which is considered as the highest breakdown point. The property of bounded influence measures the resistance of the estimator against bad observations. It encounters the tendency of the least squares to allow the leverage points to exhibit greater influence.

### **1.1.1 Outliers**

Outliers are unusual data values that occur almost in all research projects involving data collection. Outliers are

observations that have extreme value relations. The term outlier is defined as the data which are far away from the bulk of the data, or more generally, from the pattern set by the majority of the data [8, 9]. According to Justo [6], an outlier is a data point that is far outside the norm for a variable or population. An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism [10, 11, 12, and 13]. Beside this it is better to point out some definition about terms which is going hand by hand with outlier as follows. Leverage is a measure of how an independent variable deviates from its mean. An observation with an extreme value on a predictor variable is a point with high leverage. Influence an observation is said to be influential if removing that observation substantially changes the estimation of the coefficients. A datum that is "influential" is one for which the regression estimate changes considerably if it is removed. Rejection Point is the point beyond which the influence function becomes zero. That is the contribution of the points beyond the rejection point to the final estimate is comparatively legible. According to Das [14], the presence of outliers can lead to inflated error rates and substantial distortions of parameter and statistic estimates when using either parametric or nonparametric tests. Casual observation of the literature suggests that researchers rarely report checking for outliers of any sort. Outliers can have deleterious effects on statistical analyses. First, they generally serve to increase error variance and reduce the power of statistical tests. Second, if none randomly distributed they can decrease normality (and in multivariate analyses, violate as assumptions of sphericity and multivariate normality), altering the odds of making both Type I and Type II errors. Third, they can seriously bias or influence estimates that may be of substantive interest [15, 16].

### **1.1.2 Causes of Outliers**

Outliers from incorrect recording data: Outliers are often caused by human error, such as errors in data collection, recording or entry. Data from an interview can be recorded incorrectly, or mistaken upon data entry. Errors of this nature can often be corrected by returning to the original documents or even the subjects if necessary and possible and entering the correct value. Outliers from intentional or motivated mis-reporting: There are times when participants purposefully report incorrect data to experimenters or surveyors [17, 18]. A participant may make conscious effort to sabotage the research or may be acting from other motives. Social desirability and self-presentation motives can be powerful. This can also happen for obvious reasons when data are sensitive. Outliers from sampling error: Another cause of outliers is sampling. It is possible that a few members of a sample were inadvertently drawn from a different population than the rest of the sample. For example: in education, in advert entry sampling academically gifted or mentally retorted students is a possibility and (depending on the goal of the study) might provide undesirable outliers. These cases should be removed as they do not reflect the target population. Outliers as legitimate cases sampled from the correct population: It

is possible that an outlier can come from the population being sampled legitimately through random chance, it is important to note that sample size plays a role in the probability of outlying values, [18].

Population, it is more probable that a given data point will be drawn from the most densely concentrated area of the distribution, rather than one of the tails. Draper, [10] have said as a researcher casts a wider net and the data set becomes larger, the more the sample resembles the population from which it was drawn and thus the likelihood of outlying values become greater. In other words, there is only about one percentage chance you will get an outlying data point from a normally distributed population, this means that, on the average, about one percentage of your subjects should be three standard deviations from the mean [19].

### 1.1.3 Outliers Management.

Outliers being inconsistent observations and largely deviated from the majority of the observations in data need proper handling as they pose serious threat to the regression model and its estimated coefficients and, as a result, give misleading outcomes. Two types of outliers can happen in the regression dataset. One with extremely large values in the response is referred to as vertical outliers, whereas observations with extremely large values in the explanatory variable are called leverage points. Outliers being influential to the classical regression require methods that are insensitive to it. Two approaches are commonly used to cope with this problem, popularly known as diagnostic approach and robust procedures [20]. The diagnostic approaches try to identify the unusual observations through diagnostic statistics and remove it from the data and then classical procedures, for example, least square estimation procedure is then applied to the remaining clean dataset. This approach is suitable for simple data or when there are one or two outliers, but it becomes inefficient to detect outliers in case of multiple outliers present in a multidimensional dataset. Therefore, an appropriate procedure to deal with outliers is robust procedures that not only detect multiple outliers in complex data but also give efficient results [21].

## 2. Methodology

### 2.1 M Estimator

One of the robust regression estimation methods is the M estimation. The letter M indicates that M estimation is an estimation of the maximum likelihood type. If estimator at M estimation is  $\hat{\beta}_M = \beta_n(x_1, x_2, \dots, x_n)$  then:

$$E[\hat{\beta} = \beta_n(x_1, x_2, \dots, x_n)] = \beta \quad (3)$$

This implies that equation (1) is an unbiased estimator with its corresponding variance given as:

$$Var(\hat{\beta}) = \frac{[\hat{\beta}]^2}{n.E\left[\frac{d}{d\beta} \ln f(x_i; \beta)\right]^2} \quad (4)$$

Where  $\tilde{\beta}$  is the linear unbiased estimator for  $\beta$ . M estimation is an extension of the maximum likelihood estimate method and a robust estimation [21 and 22].

### 2.1.1 Algorithm 1: M - Estimator

- Estimate regression coefficients on the data using OLS.
- Test assumptions of the regression model
- Detect the presence of outliers in the data.
- Calculate estimated parameter  $\beta_0$  with OLS.
- Calculate residual value  $e_i = y_i - \hat{y}_i$
- Calculate value  $\hat{\sigma}_i = 1.4826 MAD$
- Calculate value  $u_i = e_i / \hat{\sigma}_i$
- Calculate the weighted value using:
 
$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2 & ; |u_i| \leq 4.685 \\ 0 & ; |u_i| > 4.685 \end{cases}$$
- Calculate  $\hat{\beta}_M$  using weighted least squares (WLS) method with weighted  $w_i$
- Repeat steps 5-8 to obtain a convergent value of  $\hat{\beta}_M$  and test to determine whether independent variables have significant effect on the dependent variable [22].

### 2.2 S. Estimator:

The weakness of M-estimation is the lack of consideration on the data distribution and not a function of the overall data because only using the median as the weighted value. This method uses the residual standard deviation to overcome the weaknesses of M- estimator. According to Rambli [23], the S-estimator is defined by  $\beta_S = \text{Min}_{\beta} \hat{\sigma}_s(e_1, e_2, \dots, e_n)$  with determining minimum robust scale estimator.

### 2.2.1 Algorithm 2: S - Estimator

- Estimate regression coefficients on the data using OLS.
- Test assumptions of the regression model
- Detect the presence of outliers in the data.
- Calculate estimated parameter  $\beta_0$  with OLS.
- Calculate residual value  $e_i = y_i - \hat{y}_i$
- Calculate value  $\hat{\sigma}_i$ ;
 
$$\hat{\sigma}_i = \begin{cases} \frac{\text{Median}(e_i) - \text{Median}(e_i)}{0.6745}, & \text{Iteration} = 1 \\ \frac{1}{nK} \sum w_i e_i^2, & \text{Iteration} > 1 \end{cases}$$
- Calculate the value  $u_i = \frac{e_i}{\hat{\sigma}_i}$
- Calculate the weighted value,  $w_i$ 

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{1.547}\right)^2\right]^2, & |u_i| \leq 1.547 \\ 0, & |u_i| > 1.547 \end{cases} \quad \text{Iteration} = 1$$

$$\frac{\rho(u)}{u^2} \quad \text{Iteration} > 1$$
- Calculate  $\hat{\beta}_S$  with WLS method with weighted  $w_i$
- Repeat steps 5 – 8 to obtain a convergent value of  $\hat{\beta}_S$  and test to determine whether independent variables have significant effect on the dependent variable [24].

### 2.3 MM Estimator:

MM estimation provide a high breakdown value and more efficient than M and S estimators. Breakdown value is a common measure of the proportion of outliers that can be addressed before these observations which affect the model [23]. According to Begashaw [25] MM-estimator is the solution of the following equation given as

$$\sum_{i=1}^n \rho_1(u_i) X_{ij} = 0 \text{ or } \sum_{i=1}^n \rho_1 \left( \frac{y_i - \sum_{j=0}^k X_{ij} \beta_j}{S_{MM}} \right) X_{ij} = 0 \quad (5)$$

Where  $S_{MM}$  is the standard deviation of S estimation and  $\rho$  is the Turkey's biweight function defined by:

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^2}, & -c \leq u_i \leq c \\ \frac{c^2}{6}, & u_i < -c \text{ or } u_i > c \end{cases} \quad (6)$$

### 2.3.1 Algorithm 3: MM Estimator

- Estimate regression coefficients on the data using OLS.
- Test assumptions of the regression model
- Detect the presence of outliers in the data.
- Calculate residual value  $e_i = y_i - \hat{y}_i$  of S estimate
- Calculate  $\hat{\sigma}_i = \hat{\sigma}_{sn}$
- Calculate  $u_i = \frac{e_i}{\hat{\sigma}_i}$
- Calculate the weighted value  $w_i$ :
$$w_i = \begin{cases} \left[ 1 - \left( \frac{u_i}{4.685} \right)^2 \right]^2, & |u_i| \leq 4.685 \\ 0, & |u_i| > 4.685 \end{cases}$$
- Calculate  $\hat{\beta}_{MM}$  using weighted least square method with weighted  $w_i$
- Repeat step 5 – 8 to obtain a convergent value of  $\hat{\beta}_{MM}$  and test whether the independent variables have a significant effect on the dependent variable [24, 25].

### 2.4 Simulation Study: Generating Data Set

Simulation methodology was adopted are as follows: Let  $n$  be sample size, and  $p$  the number of predictor variables; then  $ncases$  will be generated randomly in order to test the performances of the different estimators considered [26]

Table 1: Factors and levels for the simulated data sets

Number of Independent Variables 'p'	Sample Size 'n'		
10	50	150	300

## 3. Results and discussions

Tables Analysis task provides simulation results of Robust M-estimation for different sample size. The Robust MM-estimation including the Robust S-estimation. The root mean square errors can be computed for the sample size of when  $n = 50$ ,  $n = 150$  and  $n = 3000$  respectively. In addition, the Tables Analysis task also provides performances of all the estimators under study for different level of turning constant such as when  $C = 4.685$ ,  $C = 2.685$  and when  $C = 5.685$ . The Tables also contain the results for different estimators when the default value of our turning constant is either increase or decreasing to a certain number.

Table 2: Result of M-, S- and MM-estimations on simulated data with 10 levels on sample size 50.

Methods	RMSE		
	C=4.685	C=3.685	C=5.685
M-estimation	0.3414	0.3294	0.3442
MM-estimation	0.2711	0.2711	0.2711
S-estimation	C=1.547	C=0.547	C=2.547
	0.2712	0.7673	0.1647

Table 3: Result of M-, S- and MM-estimations on simulated data with 10 levels on sample size 150

Methods	RMSE		
	C=4.685	C=2.685	C=5.685
M-estimation	0.3569	0.3544	0.3593
MM-estimation	0.3329	0.3391	0.3329
S-estimation	C=1.547	C=0.547	C=2.547
	0.3489	0.9878	0.2116

Table 4: Result of M-, S- and MM-estimation on simulated data with 10 levels on sample size 300

Methods	RMSE		
	C=4.685	C=2.685	C=5.685
M-estimation	0.3863	0.3746	0.3871
MM-estimation	0.3536	0.3568	0.3529
S-estimation	K=1.547	K=0.547	K=2.547
	0.3512	0.9925	0.2129

Table 5 Comparing the performance of robust methods with 10 levels when  $n=50$ , 150 and 300

Sample Size	Methods	Root Mean Square Error (RMSE)		
n=50	M- Estimator.	0.3414	0.3294	0.3442
	MM- Estimator.	0.2711	0.2711	0.2711
	S- Estimator.	0.2712	0.7673	0.1657
n=150	M- Estimator.	0.3569	0.3544	0.3593
	MM- Estimator.	0.3329	0.3391	0.3329
	S- Estimator.	0.3489	0.9878	0.2116
n=300	M- Estimator.	0.3863	0.3746	0.3871
	MM- Estimator.	0.3536	0.3568	0.3529
	S- Estimator.	0.3512	0.9925	0.2129

## 4. Discussion of Results

Table 2 to 4 showed the result of M, MM and S estimators methods of robust regression with ten (10) independent variables on sample sizes 50, 150 and 300 respectively. Table 5 shows a comparison between the methods when the turning constant increases and decreases respectively. When  $n=50$  decreasing the  $c$  value improves the performance of M-estimator by reducing the RSE, while increasing the  $c$  value also increases the RSE. For MM-estimator increasing and decreasing the turning constant did not make any impact to the model and for the S-estimator increasing the  $k$  value to (2.547) gives a better result than the default value (1.547). For  $n=150$ , the case is same as that of  $n=50$  except for the MM-estimator which reducing the turning constant gives a RSE (0.3391) greater than that of the default value (0.3329) meaning that the default value is the best and when it increases gives a better result than reducing the  $c$  value unlike in the case of M-estimator. While for  $n=300$ , the result is same as that of  $n=150$  except for MM-estimator which shows that increasing the default value gives better result with RSE (0.3529) followed by the default value with RSE (0.3536) then when  $c$  is increased with RSE (0.3568) respectively.

## 5. Conclusion

Based on the findings of this study, we can say that adjusting the turning constant value for M, MM and S estimators have a significant impact on the resistibility of the estimators against outlying observations. From the result of the study, we can conclude that decreasing the default



turning constant value for M-estimation from (4.685) to (2.685) will increase the resistibility of the estimator against outlying observations on all sample sizes. The study also concludes that maintaining the default  $c$  value for MM-estimator gives a better result. The findings of the study also conclude that increasing the K constant value of S-estimator from (1.547) to (2.547) increases the resistibility of the estimator by having smaller RSE on all the sample sizes 50, 150 and 300 respectively. This study recommends that analyst should use smaller value than the default turning constant for M-estimation to have more resistibility against outlying observations in the regression models. Because the smaller the turning constant values the more resistant the M-estimator will be. Secondly the study recommended that analyst should use a value greater than the default K value for the S-estimation, because the higher the turning constant the more resistant the S-estimator against outlying observations.

## Reference

- [1]. Kumari, K. and Yadav, S., (2018), Linear Regression Analysis. *Journal of the Practice of Cardiovascular Sciences*, 4(1), 33 – 36.
- [2]. Parmar, T., Mori, H., Poriya, D., Vanani, V., Chauhan, S. and Bhagat, Y., (2018). Identification of Methodology Used in Real Estate Property Valuation, *International Research Journal of Engineering and Technology*, 5(2), 170-173
- [3]. Shrestha, N. (2020). Detecting multi-collinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42.
- [4]. Yuliana, S., Hasih, P., Sri, S. H. and Twenty, L., (2014). M-estimation, S-estimation, and MM-estimation in robust regression. *International Journal of Pure and Applied Mathematics*, 91, 349-360.
- [5]. Almetwally, E. M. and Almongy, H. M., (2018). Comparison between M-estimation, S-estimation, and MM-estimation methods of robust estimation with application and simulation. *International Journal of Mathematical Archive*. 9(11), 1-9
- [6]. Justo, C. E. and Calandra, M. V., (2021). Evaluation of robust linear regression methods for the measurement of a topographic altimetric network. *Journal of Scientific and Engineering Research*, 8(4), 30-39
- [7]. Enejoh, S. Y. and Tsauni, A. M., (2017). An analytical study of the impact of inflation on economic growth in Nigeria (1970– 2016). *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 7(4), 110-120.
- [8]. Muhammad, K., Anum, Y., Seema, Z., Muhammad A. K., Zubair, A. and Osama, A. A., (2021). Applications of robust regression techniques: An econometric approach. *Hindawi Journal of Mathematical Problems in Engineering*, 9(1), 1-9.
- [9]. Yuliana Susanti, Pratiwi H., Liana T., (2014). M Estimation, S Estimation, and MM Estimation in Robust Regression. *International Journal of Pure and Applied Mathematics*, Volume 91 No. 3, 349-360.
- [10]. Das, S., Dey, A., Pal, A., and Roy, N. (2015). Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *International Journal of Computer Applications*, 115(9), 31-41.
- [11]. Khaleelur Rahman, S. M. A., Mohamed, M., Sathik, K. and Kannan, S., (2012). Multiple linear regression models in outlier detection, *International Journal of Research in Computer Science*, 2(2), 23-28
- [12]. Ibrahim, S., Rambli, A., Hussin, A.G. and Mohamed, I., (2013). Outlier detection in a circular regression model using COVRATIO statistic. *Communications in Statistics Simulation and Computation*, 42(10), 2272-2280.
- [13]. Jha, J., Biswas, A. and Cheng, T. C., (2022). Trimmed estimator for circular–circular regression: Breakdown properties and an exact algorithm for computation. *Statistics*, 56(2), 375-395.
- [14]. Rasheed, B. A., Adnan, R. and Lukunti, S., (2022). Comparisons of Hampel and Andrews sin PSI function for heteroscedasticity model in the presence of outliers and multi-collinearity. *International Journal of Statistics and Applied Mathematics*, 7(3), 17-27.
- [15]. Begashaw, G. B. and Yohannes, Y. B., (2020). Review of Outlier Detection and Identifying Using Robust Regression Model. *International Journal of Systems Science and Applied Mathematics*, 5(1), 2020, pp. 4-11.
- [16]. Ranjit, K. P., (2014), Some Methods of Detection of Outliers in Linear Regression Model, Iasri, *Library Avenue*, New Delhi-110012.
- [17]. Salman, M. J. and Salah, R. A., (2022), Comparison of some artificial neural networks for graduate students, *Periodicals of Engineering and Natural Sciences Original Research*, 10(3), 187-196.
- [18]. Mokhtar, N. A., Zubairi, Y. Z., Hussin, A. G. and Moslim, N. H., (2019). An outlier detection method for circular linear functional relationship model using covratio statistics. *Malaysian Journal of Science* 38(Special Issue 2), 46-54.
- [19]. Rambli, A., Yunus, R. M., Mohamed, I. and Hussin, A. G., (2015). Outlier detection in a circular regression model. *Sains Malaysiana* 44(7), 1027-1032.
- [20]. Lukman A. F., Ayinde K., Ajiboye A. S., (2017). Monte Carlo study of some classification-based ridge parameter estimators. *Journal of Modern Applied Statistical Methods*, 16(1), 428–451.
- [21]. Begashaw, A. B. and Yohannes, Y. B., (2020). Review of outlier detection and identifying using robust regression. *International Journal of Systems Science and applied mathematics*. 5(1), 4-11.
- [22]. Yohai, V. J., (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 5, 642-656.
- [23]. Liang, Y. and Sullivan, V. A., (2004). Robust Estimation Using SAS Software, *Department of Biostatistics Merrell Dow Research Institute*

- Cincinnati, Ohio 45215 Mathsoft, Inc. Seattle, WA, 255-298.
- [24]. Victor, B. Alberto, J. and Justo P., (2022). Robust optimal classification trees under noisy labels, *Advances in Data Analysis and Classification* 16, 155–179.
- [25]. Alita, D., Putra, A. D. and Darwis, D., (2021). Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation, *Indonesian Journal of. Computing and Cybernetics. Systems*, 15(3), 295- 308.
- [26]. Salibian, M. and Yohai, V. J., (2006). A Fast Algorithm for S-Regression Estimates. *Journal of Computational and Graphical Statistics*, 15(2), 414-427.