Turkish Journal of Veterinary and Animal Sciences

http://journals.tubitak.gov.tr/veterinary/

Research Article

Turk J Vet Anim Sci
(2022) 46: 420-428
© TÜBİTAK
10.55730/1300-0128.4190

Comparison of M, MM and LTS estimators in linear regression in the presence of outlier

Cem TIRINK[1,*], **Hasan ÖNDER**[2]

[1]Department of Animal Science, Faculty of Agriculture, Iğdır University, Iğdır, Turkey
[2]Department of Animal Science, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Turkey

Received: 27.02.2021 • **Accepted/Published Online:** 08.03.2021 • **Final Version:** 13.06.2022

Abstract: In this study, it was aimed to evaluate the performance of different estimators that will be used in regression analysis, which is one of the multivariate statistical methods in the presence of outliers in the data set. Sixth month live weight was estimated with various body measurements for Saanen kids taken from a private farm. In the data set, the use and performance of robust estimators were evaluated because the least squares method did not provide reliable results in the case of outliers. M (for Huber and Tukey bisquare) estimator, MM estimator and LTS estimator were used as robust used in the presence of outliers. MSE, RMSE, rRMSE, MAPE, MAD, R^2, R^2_{adj} and AIC were used as model comparison criteria in the study. As a result of the study, in the case of outlier in the data set, Huber type M estimator can be recommended.

Key words: Least squares, outliers, robust estimator, Saanen

1. Introduction

Goat, which is of primary importance in most civilizations, is a multifaceted small ruminant that has a significant role for development of rural economy and growing healthy human generations. It is an important consideration to reveal the relationship between live weight and body measurements, which are important indirect selection criteria in terms of the studied breed characterization for goat breeding purposes. Prediction of live weight in farm animals is a significant task for feed amount, medicinal dose, and marketing price of an animal, especially under village conditions where weigh bridge is unavailable [1]. The prediction of live weight played a vital role in flock management with the scope of gaining more profit [2]. The sustained attention has been drawn about describing body measurements that positively influence the live weight for reproducing superior offspring that help breeders to increase meat productivity in indirect selection criteria. In this respect, sophistical statistical techniques are still needed to healthy determine the indirect selection criteria.

Many studies are available about the prediction of live weight from body measurements in sheep [3], goat [4], and cattle [5]. There were several published studies on live weight prediction using different statistical techniques i.e. correlation analysis [6–9], simple and multiple linear regression [10–12], use of factor and principal component analysis in multiple linear regression [13] and decision trees and artificial neural network algorithms [1]. However, information reported about M, MM, and LTS estimators is scarce in live weight prediction in the literature for goat species.

Multivariate statistical methods can be used in the determining the relationship between live weight and morphological characteristics in goat breeding. The relationships between explanatory and response variables used in scientific studies will be a source of information for both current studies and future studies. Multivariate statistical methods are needed to determine the estimation equations of linear and nonlinear of the relationships between explanatory and response variables. The method used in modeling the relationship between two and/or more than two variables with cause-effect relationships is called the regression analysis method [14]. Regression analysis has four general uses: modeling between explanatory and response variables, parameter estimation, estimation, and control [15]. Various estimators are used while estimating parameters with the regression analysis. To consider the linear regression model that assumes the linear relationship between the explanatory and the response variable:

$$y = X\beta + \varepsilon$$

where β is the regression coefficients and ε is the error term. The main purpose of using these estimators is to ensure the validity of the predictions to be made. The most

* Correspondence: cem.tirink@igdir.edu.tr

basic and widely used estimator for the simple and linear regression model used in many fields is the least squares (LS) method. LS method aims at minimizing the mean square error of the model [16, 17].

$$\hat{\beta}_{LS} = \operatorname{argmin} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Although the LS estimator has the desired statistical properties among the unbiased linear estimators, it can be affected in the presence of outliers in the data set [18, 19]. Outliers are observations that disrupt the normal distribution of the observations. Outliers are an expression used for observations in the data set that are far from the agglomeration point and appear inconsistent [20]. Outliers can give information about the real status of the data, as well as measurement errors, errors during record keeping and errors made during the transfer of records [21]. The presence of outliers is one of the most likely events for many scientific studies. If there is an outlier in the data set, the reliability of the model estimates to be made with the LS is considered as poor or sensitive [21, 22]. In the presence of outliers, due to the low reliability of the estimates to be made based on the LS, robust statistical methods have been proposed as an alternative to LS.

In this study, we aimed to estimate the optimum model by comparing the performance of the model with the use of robust estimators such as M (for Huber and Tukey bisquare) estimator, MM estimator and LTS estimator that proposed as alternative estimators to LS in the presence of outliers in the data set in linear regression.

2. Material and methods

2.1. Material

The data set used as a material in this study was taken from Saanen kids in a private business in Bafra district of Samsun province. For this purpose, various body measurements (wither height (WH), body length (BL), chest depth (CD) and rump height(RH)) and live weights (LW) (up to 6 months of age including birth) taken from 82 head Saanen kids for 6th month were used. Sixth month live weight (kg) was used as a response variable (y).

For the descriptive statistics of the data performed by using "psych" package with R software [23]. Analyzes were made in RStudio program using packages containing the solution of related estimators [24]. The related packages for MM and S estimator were used "robustbase", for LTS estimator "robustHD" were used [25, 26]. Also for Huber and bisquare M estimator were used MATLAB program. In the calculation of model comparison criteria, the RStudio program was used by using the "ehaGoF" package written by Eyduran [27].

2.2. Methods

Linear regression is an approach to explain and model the relationship between explanatory and response variables. The matrix notation of the function of the multiple regression model with more than one independent variable is given as follows:

$$y = X\beta + \varepsilon$$

where β is the regression coefficients, y is $n \times 1$ dimensional vector as response variable, X is $n \times (p + 1)$ dimensional matrix as an explanatory variable and ε is $n \times 1$ dimensional vector for the error term.

The prediction equation obtained as a matrix notation is

$$\hat{y} = X\hat{\beta}.$$

Many methods have been proposed to find the estimated parameters in the regression model represented in matrix format. The most widely used method for linear regression is the LS method. Although LS is shown as the most appropriate method in model estimation, if model assumptions i.e. constant variance, linearity, normality, and multicollinearity, etc., cannot be achieved, the use of this method may not be reliable [28, 29]. Due to the low performance of the model estimation in the case of outliers in the data set with the LS method, new methods that are not affected by the outliers and make more efficient model estimation with the adding of a weighting function have been developed with the term "Robustness" first proposed by [29, 30].

2.2.1. Least squares method

The LS is the method used to determine the best-fit line for a data set, providing a visual representation of the relationship between data points [4]. The main purpose of LS is to minimize the sum of the squares of error terms of the estimated parameters (β) in the model of the regression model. Estimation parameters (β) with the LS method are calculated as given below [31].

$$\hat{\beta}_{LS} = (X'X)^{-1}X'Y$$

The prediction of β_{LS} is the one with the smallest variance of all the unbiased estimators of the parameter β. In case of outliers in the data set, the reliability of the regression model will be adversely affected [16, 32]. The normal distribution of the error terms, which is one of the assumptions of LS, occurs when there are outliers in the data set.

As an alternative to LS, many methods have been proposed to eliminate the negative effect of outliers.

2.2.2. M estimator

For the model estimates to be made with LS, if there is an outlier in the data set, the reliability of the model to be obtained decreases. The M estimator was proposed by [33] and eliminates the negative effect of the outlier in the model estimation. The M estimator is an extension of the maximum likelihood (ML) method. M estimator is a widely used method when there are leverage points between observations [34]. In the LS method, it is aimed to minimize the errors, but the presence of an outlier in the data will make it difficult to minimize the error and cause unreliable results [21]. For this reason, the M estimator, which is developed, tries to minimize the $p(e_i)$ function in which the errors are found, instead of minimizing the mean square error.

$$\hat{\beta}_M = \operatorname{argmin} \sum_{i=0}^{n} p(e_i),$$

where $p(e_i)$ is a differentiable function. This function when derivatived according to the β parameters, the weighted function is obtained as given below.

$$\sum_{i=1}^{n} \psi(e_i)x_i = 0$$

There are different functions such as Huber and Tukey bisquare which are used in the calculation of the M estimator.

2.2.2.1. Huber M estimator

The function for Huber M estimator as given below.

$$\rho(e) = \begin{cases} \frac{e^2}{2}, & |e| \le k \\ k|e| - \frac{k^2}{2}, & |e| > k \end{cases}$$

Taking the derivative of the p(e) function according to the errors, the weighted function of the Huber M estimator [35] and also the value of the k parameter of the Huber M estimator is 1.345 [36].

$$\psi(e) = \begin{cases} e, & |e| \le k \\ k/|e|, & |e| > k \end{cases}$$

2.2.2.2. Tukey bisquare M estimator

The function for Tukey bisquare M estimator as given below.

$$\rho(e) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - \left(\frac{e}{k}\right)^2\right]^3\right\}, \\ \frac{k^2}{6}, \end{cases}$$

Taking the derivative of the p(e) function of the Tukey bisquare M estimator and also the value of the k parameter of the Tukey bisquare M estimator is 4.685 [36].

2.2.3. MM estimator

The MM estimator was proposed by [37] as a method with a high breakdown point. The MM estimator makes the regression parameter estimation using the S estimator that minimizes the scale of the M estimator [38]. MM estimator allows data to be generated by looking at the distances between subsets and the method based on defining data away from the majority subset as an outlier [39]. MM estimation aims to obtain estimates with high breakdown point and more reliable results. The MM estimator obtains the errors using the S estimator. In the second step, it calculates the scale parameter of the model obtained. Then, with the calculated scale parameters, the scaled errors can be calculated as given below:

$$u_i = e_i/\hat{\sigma}_i$$

The weight function is calculated given below using the scaled errors.

$$w_i = \begin{cases} \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2, & |u_i| \le 4.685 \\ 0, & |u_i| > 4.685 \end{cases}$$

Further, estimate β_{MM} using the weighted least square (WLS) with obtained weights from the weighted function. This process continues until convergence is complete. If there is no convergence, the MM estimator returns to the first stage and the iterative process starts again [38].

2.2.4. Least trimmed squares (LTS) estimator

LTS estimator, which is one of the Robust regression methods, is proposed by [40]. The objective function of the LTS estimator is the smallest trimmed of squared residuals as given below.

$$\hat{\beta}_{LTS} = \operatorname{argmin} \sum_{i=1}^{h} e_{i:n}^2$$

The $e_{i:n}^2$ represents the ith order statistics from a series of e_i^2. The LTS estimator aims to minimize the sum of squares of h least error by ordering the squares of the error values in ascending order. During the process, it is important that the determination of the optimum number of h. The number of h is calculated as given below.

$$h = \left(\frac{n}{2}\right) + \left(\frac{p+1}{2}\right)$$

With the number of h, LTS estimator achieves the possible value for the breakdown point [41, 42].

2.3. Model comparison criteria

As the model comparison criteria for the models obtained from various estimators, mean square error (MSE), root mean square error (RMSE), relative root mean square error (rRMSE), mean absolute percentage error (MAPE), mean absolute deviation (MAD), determination of coefficients (R^2), adjusted determination coefficients (R^2_{adj}) and Akakike information criteria (AIC) formulas given below were used.

Model comparisons were made according to the lowest RMSE, rRMSE, MAPE, MAD and AIC values and the highest R^2 and R^2_{adj} values [43].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|$$

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$R^2 = \left[1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right]$$

$$R^2_{adj} = \left[1 - \frac{\frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - \bar{y})^2}\right]$$

$$\text{AIC} = n * \ln\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right] + 2p$$

3. Results and discussion

The Kolmogorov–Smirnov test was used to check whether the data was suitable for normal distribution in the study in which the 6th month live weight of Saanen races was wanted to be estimated with different estimators and to evaluate their performance. Descriptive statistics and normality test results as a p-value about variables used as result and explanatory variable are given in Table 1. The explanatory variables such as 1st month live weight, 6th month withers height, 3th month body length, birth chest depth, 1st month chest depth, 2nd month chest depth, 4th month chest depth, 6th month chest depth, 3rd rump height, and 6th rump height were determined that they do not show normal distribution ($p < 0.05$). The data set was divided as 70% train and 30% test set.

3.1. Results of least squares method

The model obtained from the LS method by using various body measurement values of the 6th month live weight is given in Table 2.

In Table 2, constant, 5th month body weight and 6th month breast depth variables have a statistically significant contribution in the model for training set ($p < 0.05$). The MSE, R^2 and R^2_{adj} values of the model are 2.148, 84.3 and 70.7, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.918.

3.2. The result of M estimator

3.2.1. Huber M estimator

The model obtained from the Huber M estimator method by using various body measurement values of the 6th month live weight is given in Table 3.

In Table 3, the intercept and the 5th month body weight variable has a statistically significant contribution to the model for the training set ($p < 0.05$). The MSE, R^2 and R^2_{adj} values of the model are 2.705, 80.2% and 63%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.903.

For the testing set, the MSE, R^2 and R^2_{adj} values of the model are 6.119, 61.8% and 58.3%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.792.

3.2.2. Tukey bisquare M estimator

The model obtained from the Tukey bisquare M estimator method by using various body measurement values of the 6th month live weight is given in Table 4.

In Table 4, 5th month body weight variable has a statistically significant contribution to the model for the training set ($p < 0.05$). The MSE, R^2 and R^2_{adj} values of the model are 3.241, 76.3% and 75.4%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.886.

For the testing set, the MSE, R^2 and R^2_{adj} values of the model are 6.589, 58.8% and 55.1%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.774.

3.3. Results of MM estimator

The model obtained from the MM estimator method by using various body measurement values of the 6th month live weight is given in Table 5.

In the training set, the best model estimation was performed under conditions 44th iteration and 0.2571 scale parameter. In Table 5, intercept, 1st month live weight, 3rd month live weight, 5th month live weight, 1st mont withers height, 2nd month withers height, 6th month withers height, birth body length, 2nd month body length, 3rd month body length, birth chest depth, 1st month chest depth, 6th month chest depth, 5th month rump height and 6th month rump height variable has a statistically significant contribution to the model for the training set ($p < 0.05$). The MSE, R^2 and R^2_{adj} values of the model are 2.543, 81.4% and 80.7%, respectively. The Pearson correlation coefficient between the estimated

Table 1. Descriptive statistics and normality test results.

	Mean	Standard deviation	Minimum	Maximum	p-value*
LW6	30.79	3.85	19.90	44.40	0.076
BW	3.77	0.81	2.00	5.70	0.200
LW1	9.38	1.37	7.00	12.70	0.004
LW3	17.54	2.24	12.60	22.20	0.200
LW5	26.17	2.58	17.20	31.20	0.200
BWH	39.73	1.61	35.00	43.00	0.083
WH1	42.78	1.48	39.00	46.00	0.200
WH2	46.71	1.63	42.50	50.00	0.043
WH4	53.74	2.06	49.00	59.00	0.005
WH6	59.85	2.57	54.50	67.50	0.001
BBL	38.34	1.65	34.50	42.00	0.200
BL1	42.12	1.68	37.50	45.50	0.165
BL2	46.19	1.69	42.50	50.00	0.200
BL3	49.01	2.22	44.00	55.00	0.004
BL5	53.93	2.75	47.50	60.00	0.046
BCD	12.12	0.91	10.00	14.50	0.003
CD1	15.63	1.16	13.00	19.00	0.018
CD2	17.36	1.24	14.50	21.00	0.007
CD4	20.63	1.42	18.00	24.00	0.004
CD5	22.55	1.35	19.50	26.50	0.096
CD6	25.13	1.10	22.50	27.50	0.006
RH2	45.47	2.31	39.50	49.50	0.044
RH3	48.73	2.08	43.00	53.00	0.001
RH4	51.77	1.87	47.00	57.00	0.182
RH5	55.22	2.17	51.00	61.00	0.161
RH6	59.59	2.82	52.00	67.50	0.038

* p-value for Kolmogorov–Smirnov test.

6th month live weight values and the actual values was determined as 0.904.

For the testing set, the MSE, R^2 and R^2_{adj} values of the model are 6.131, 61.7% and 58.2%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.793.

3.4. Results of LTS estimator

In the training set, the best model estimation from the LTS estimator was performed under conditions 0.5022 scale parameter. The estimated intercept and 5th month live weight have a statistically significant contribution to the model for the training set ($p < 0.05$) and coefficients of significant variables were detected 4.8894359 and 0.9771741, respectively. The MSE, R^2 and R^2_{adj} values of the model are 4.507, 61.7% and 65.9%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.859.

For the testing set, the MSE, R^2 and R^2_{adj} values of the model are 6.725, 58.0% and 54.2%, respectively. The Pearson correlation coefficient between the estimated 6th month live weight values and the actual values was determined as 0.768.

3.5. Model comparison

In the study where alternative estimators were used in cases where outliers were found in the data set, MAD, MAPE, RMSE, MSE, R^2, R^2_{adj}, rRMSE and AIC were used as model comparison criteria.

Model comparisons were made according to the lowest RMSE, rRMSE, MAPE, MAD and AIC values and the highest R^2 and R^2_{adj} values [43].

Table 2. Results of least squares.

	Estimate	Std. error	t value	Sig.
(Intercept)	–43.1123	19.1871	–2.247	0.0319
BW	–2.0637	1.7964	–1.149	0.2594
LW1	1.7787	1.5932	1.116	0.2728
LW3	–1.3767	0.8092	–1.701	0.0989
LW5	1.9060	0.4119	4.627	6.24e-05
BWH	0.5136	0.5710	0.900	0.3753
WH1	0.6600	0.5305	1.244	0.2228
WH2	0.8968	0.8096	1.108	0.2765
WH4	–1.1275	0.7504	–1.503	0.1431
WH6	0.9015	0.6238	1.445	0.1584
BBL	–0.8471	0.5690	–1.489	0.1467
BL1	–0.5404	0.9696	–0.557	0.5813
BL2	–1.0051	0.6535	–1.538	0.1342
BL3	0.3733	0.6090	0.613	0.5444
BL5	0.1979	0.4373	0.453	0.6540
BCD	1.2968	1.0451	1.241	0.2240
CD1	–0.3957	1.3880	–0.285	0.7775
CD2	–0.4001	0.9451	–0.423	0.6750
CD4	0.3501	0.7274	0.481	0.6337
CD5	–1.6402	1.0583	–1.550	0.1313
CD6	1.8093	0.8586	2.107	0.0433
RH2	–0.3051	0.5152	–0.592	0.5581
RH3	0.2942	0.6031	0.488	0.6291
RH4	0.2150	0.6915	0.311	0.7580
RH5	0.2736	0.4553	0.601	0.5523
RH6	–0.2855	0.4162	–0.686	0.4979

Table 3. Results of Huber M estimator.

	Estimate	Std. error	t value	Sig.
(Intercept)	–-14.136	18.544	–0.762	0.452
BW	–0.522	1.736	–0.301	0.765
LW1	0.612	1.540	0.398	0.694
LW3	–0.736	0.782	-0.941	0.354
LW5	1.488	0.398	3.737	0.001
BWH	0.283	0.552	0.513	0.612
WH1	0.206	0.513	0.402	0.691
WH2	0.638	0.782	0.816	0.421
WH4	–0.367	0.725	–0.505	0.617
WH6	0.134	0.603	0.222	0.826
BBL	–0.675	0.550	–1.227	0.229
BL1	0.071	0.937	0.076	0.940
BL2	–0.563	0.632	-0.892	0.379
BL3	0.405	0.589	0.688	0.496
BL5	–0.027	0.423	–0.063	0.950
BCD	0.326	1.010	0.323	0.749
CD1	0.016	1.342	0.012	0.990
CD2	–0.280	0.913	–0.306	0.761
CD4	0.082	0.703	0.117	0.908
CD5	–0.093	1.023	–0.091	0.928
CD6	0.397	0.830	0.478	0.636
RH2	–0.086	0.498	–0.173	0.864
RH3	0.080	0.583	0.137	0.892
RH4	–0.114	0.668	–0.170	0.866
RH5	0.108	0.440	0.245	0.808
RH6	–0.030	0.402	–0.075	0.941

According Table 6, it was determined that the highest R^2 and R^2_{adj} value is for the M-Huber estimator. Moreover, when the other model selection criteria are examined, it was seen that the MM estimator has the lowest value in terms of MAPE and MAD, M-Huber estimator has the lowest value in terms of MSE, RMSE, rRMSE and AIC.

Under these conditions, it is recommended to use the M-Huber estimator to estimate the 6th month live weight of Saanen kids.

4. Conclusion

In the present study, we aimed to determine the effective performances of M, MM and LTS estimator according to LS method in the presence of outliers in the data set. For this purpose, we first made the model predictions in the data set with outliers and we concluded that the model performance of the M-Huber estimator is more reliable in the presence of outliers in the data set. It supports the result we have achieved in the studies.

[44] aimed to compare the model performances in which LS method, the M estimator, Theil estimator, and the least absolute deviation (LAD) estimator and stated that the M estimator gave the best result in terms of model performance. [45] aimed to compare of performance the M estimator, which is one of the robust estimators, as an alternative to the LS in the presence of outliers, stated that the M estimator is more useful than the LS. [46] emphasized that the best three methods were M, MM-S and MM estimators in their study where they compared EKK, Huber M estimator, bisquare M estimator, MM estimator, S estimator, and MM (-S) estimators.

Regression analysis, which is one of the multivariate statistical methods used in the interpretation of the data obtained as a result of scientific studies, provides the

Table 4. Results of Tukey bisquare M estimator.

	Estimate	Std. error	t value	Sig.
(Intercept)	–1.959	18.546	–0.106	0.917
BW	0.132	1.736	0.076	0.940
LW1	0.168	1.540	0.109	0.914
LW3	–0.513	0.782	–0.656	0.517
LW5	1.383	0.398	3.473	0.002
BWH	0.286	0.552	0.519	0.608
WH1	0.113	0.513	0.221	0.826
WH2	0.572	0.783	0.730	0.471
WH4	–0.180	0.725	–0.249	0.805
WH6	–0.100	0.603	–0.166	0.869
BBL	–0.524	0.550	–0.953	0.348
BL1	0.232	0.937	0.248	0.806
BL2	–0.459	0.632	–0.726	0.473
BL3	0.367	0.589	0.624	0.537
BL5	–0.105	0.423	–0.249	0.805
BCD	0.074	1.010	0.073	0.942
CD1	0.472	1.342	0.352	0.727
CD2	–0.419	0.914	–0.459	0.650
CD4	0.036	0.703	0.051	0.959
CD5	0.184	1.023	0.180	0.858
CD6	–0.033	0.830	–0.039	0.969
RH2	–0.112	0.498	–0.224	0.824
RH3	0.018	0.583	0.031	0.975
RH4	–0.236	0.668	–0.353	0.726
RH5	0.022	0.440	0.051	0.960
RH6	0.087	0.402	0.216	0.830

Table 5. Results of MM estimator.

	Estimate	Std. error	t value	Sig.
(Intercept)	–25.9431	3.69507	–7.021	6.99E-08
BW	–1.51388	0.20641	–7.334	2.96E-08
LW1	1.34885	0.1928	6.996	7.48E-08
LW3	–1.08014	0.40673	–2.656	0.01239
LW5	1.57157	0.31239	5.031	1.97E-05
BWH	0.1992	0.11472	1.736	0.09243
WH1	0.13241	0.05729	2.311	0.02765
WH2	0.62715	0.22672	2.766	0.00947
WH4	–0.34124	0.18465	–1.848	0.07417
WH6	0.21114	0.0715	2.953	0.00595
BBL	–0.88609	0.15566	–5.692	2.95E-06
BL1	0.08818	0.358	0.246	0.80707
BL2	–0.66909	0.20562	–3.254	0.00275
BL3	0.55256	0.091	6.072	1.00E-06
BL5	0.03478	0.04362	0.797	0.4313
BCD	0.5132	0.09615	5.338	8.16E-06
CD1	–0.63247	0.20423	–3.097	0.00413
CD2	–0.01596	0.11028	–0.145	0.88586
CD4	0.01685	0.08261	0.204	0.83975
CD5	–0.07721	0.15163	–0.509	0.6142
CD6	0.67918	0.11537	5.887	1.69E-06
RH2	0.04523	0.17222	0.263	0.79457
RH3	0.04833	0.09352	0.517	0.60897
RH4	0.00979	0.07623	0.128	0.89863
RH5	0.28268	0.04131	6.844	1.14E-07
RH6	–0.16482	0.07647	–2.155	0.03902

Table 6. Comparisons of the model performances.

	LS	M-Huber	M-bisquare	MM	LTS
MSE	5.528	6.119	6.589	6.131	6.725
RMSE	2.351	2.474	2.567	2.476	2.593
rRMSE	7.817	8.225	8.535	8.233	8.622
MAPE	3.970	2.675	2.755	2.164	2.560
MAD	1.264	0.928	0.962	0.776	0.875
R^2	0.655	0.618	0.588	0.617	0.58
R^2_{adj}	0.624	0.583	0.551	0.582	0.542
AIC	46.745	49.286	51.137	49.334	51.645

opportunity to model the studies and make predictions. Using the correct estimator is important for the reliability of the results for the regression analysis. Especially in biological studies, the accuracy of the analysis method used as well as the sensitivity of the experiments established is important for the reliability of the results to be obtained.

In the current study, it was aimed to estimate the 6th month live weights from various body measurements taken from Saanen kids in the presence of outliers in the data set, Huber type M estimator, one of the robust estimating methods, was proposed.

Acknowledgment
This research was produced from a PhD thesis prepared by first author.

Conflict of interest
The authors declare no conflict of interest for the present study.

References

1. Eyduran E, Zaborski D, Waheed A, Celik S, Karadas K et al. Comparison of the predictive capabilities of several data mining algorithms and multiple linear regression in the prediction of body weight by means of body measurements in the indigenous Beetal goat of Pakistan. Pakistan Journal of Zoology 2017; 49 (1): 257-265.
2. Birteeb PT, Ozoje MO. Prediction of live body weight from linear body measurement of West African long-legged and West African dwarf sheep in Northern Ghana. Online Journal of Animal and Feed Research 2012; 2: 425-434.
3. Eyduran E, Akin M, Eyduran SP. Application of Multivariate Adaptive Regression Splines through R Software. Ankara, Turkey: Nobel Academic Publishing; 2019.
4. Yilmaz F, Bayyurt L, Abaci SH, Tahtali Y. Comparison of Least Squares and Some Bias Estimators in Multicollinearity. Turkish Journal of Agriculture - Food Science and Technology 2020; 8 (3): 793-799.
5. Aytekin I, Eyduran E, Karadas K, Akşahan R, Keskin I. Prediction of fattening final live weight from some body measurements and fattening period in young bulls of crossbred and exotic breeds using MARS data mining algorithm. Pakistan Journal of Zoology 2018; 50 (1): 189-195.
6. Yakubu A. Fixing collinearity instability ın the estimation of body weight from morpho-biometrical traits of West African dwarf goats. Trakia Journal of Sciences 2009; 7 (2): 61-66.
7. Tadesse A, Gebremariam T, Gangwar SK. Application of linear body measurements for predicting body weight of Abergelle goat breed in Tigray Region, Northern Ethiopia. Global Joural of Bio-Science & Biotechnology 2012; 1 (2): 314-319.
8. Dakhlan A, Saputra A, Hamdani MDI, Sulastri S. Regression models and correlation analysis for predicting body weight of female ettawa grade goat using its body measurements. Advances in Animal and Veterinary Sciences 2020; 8 (11): 1142-1146.
9. Gore DLM, Muasya TK. Okeno TO, Mburu JN. Comparative reproductive performance of Saanen and Toggenburg bucks raised under tropical environment. Tropical Animal Health and Production 2020; 52: 2653-2658.
10. Moaeen-ud-Din M, Ahmad N, Iqbal A, Abdullah M. Evaluation of different formulas for weight estimation in Beetal, Teddi and crossbred (Beetal x Teddi) goats. Journal of Animal and Plant Sciences 2006; 16 (3-4): 70-74.
11. Parés PM, Mwaanga ES, Caballero M, Sabate J, Valenzuela S. Live weight estimation of Gwembe goat (*Capra hircus*) from measurement of thoracic girth. Journal of Veterinary Anatomy 2012; 5 (2): 9-14.
12. Tyasi TL, Mathapo MC, Mokoena K, Maluleke D, Rashijane LT et al. Assessment of relationship between body weight and morphological traits of South African non-descript indigenous goats. Journal of Animal Health and Production 2020; 8 (1): 32-39.
13. Eyduran E, Waheed A, Tariq MM, Iqbal F, Ahmad S. Prediction of live weight from morphological characteristics of commercial goat in Pakistan using factor and principal component scores in multiple linear regression. The Journal of Animal & Plant Sciences 2013; 23 (6): 1532-1540.
14. Ari A, Onder H. Regression models used for different data structures. Anadolu Journal of Agricultural Sciences 2013, 28 (3): 168-174 (in Turkish with an abstract in English).
15. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis 3rd ed. New York, NY,USA: John Wiley & Sons, Inc.; 2001.
16. Cankaya S, Kayaalp GT, Sangun L, Tahtali Y, Akar M. A comparative study of estimation methods for parameters in multiple linear regression model. Journal of Applied Animal Research 2006, 29 (1): 43-47.
17. Gurunlu Alma O,Vupa O. The comparison of Least Squares and Least Median Squares Estimation methods which are used in linear regression analysis. Süleyman Demirel University Faculty of Arts and Science Journal of Science 2008, 3 (2): 219-229 (in Turkish with an abstract in English).
18. Maronna RA, Martin RD, Yohai VJ, Salibian-Barrera M. 2nd ed. Robust Statistics: Theory and Methods (with R). Oxford, UK: John Wiley & Sons; 2019.
19. Filzmoser P. Kurnaz FS. A robust Liu regression estimator. Communications in Statistics - Simulation and Computation 2018; 47 (2): 432-443.
20. Walfish S. A review of statistical outlier methods. Pharmaceutical Technology 2006, 30 (11): 82-86.
21. Abacı SH. A comparative study of some estimation methods in simple linear regression model for different sample sizes in presence of outliers. MSc, Ondokuz Mayis University, Samsun, Turkey, 2013.

22. Agullo J, Croux C, Van Aelst S. The multivariate least trimmed squares estimator. Journal of Multivariate Analysis 2008, 99 (3): 311-318.

23. Revelle W. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, 2020.

24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020.

25. Maechler M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A et al. robustbase: Basic Robust Statistics R package version 0.93-6; 2020.

26. Alfons A. robustHD: Robust Methods for High-Dimensional Data. R package version 0.6.1; 2019.

27. Eyduran E. ehaGoF: Calculates Goodness of Fit Statistics. R package version 0.1.0; 2019.

28. Mendes M, Akkartal E. Regression tree analysis for predicting slaughter weight in broilers. Italian Journal of Animal Science 2009, 8: 615-624.

29. Gokce A. Biased regression analysis based on some robust estimators. MSc, Eskişehir Osmangazi University, Eskişehir, Turkey, 2019.

30. Box GE. Non-normality and tests on variances. Biometrika 1953; 40 (3/4): 318-335.

31. Alpar R. Uygulamalı çok değişkenli istatistiksel yöntemler. 4. Baskı. Ankara, Turkey: Detay Yayıncılık; 2013.

32. Cankaya S, Eker S, Abaci SH. Comparison of Least Squares, Ridge Regression and Principal Component approaches in the presence of multicollinearity in regression analysis. Turkish Journal of Agriculture-Food Science and Technology 2019; 7 (8): 1166-1172.

33. Huber PJ. Robust Regression: Asymptotics, Conjectures, and Monte Carlo. Annals of Statistics 1973; 1: 799-821.

34. SAS Institute Inc. 2018. SAS/STAT® 15.1 User's Guide. Chapter 4. Cary, NC: SAS Institute Inc.

35. Huber PJ. Robust Statistics. New York , USA: John Wiley & Sons; 1981.

36. Alamgir, Ali A., Khan SH, Khan DM, Khalil U. A New Efficient Redescending M- Estimator: Alamgir Redescending M-estimator. Research Journal of Recent Sciences 2013; 2 (8): 79-91.

37. Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. The Annals of Statistics 1987; 15: 642-656.

38. Susanti Y, Pratiwi H, Sulistijowati HS, Liana T. M estimation, S estimation, and MM estimation in robust regression. International Journal of Pure and Applied Mathematics 2014; 91 (3): 349-360.

39. Arslan O, Billor N. Robust Liu estimator for regression based on an M-estimator. Journal of Applied Statistics 2000; 27 (1): 39-47.

40. Rousseeuw PJ. Least median of squares regression. Journal of American Statistical Association 1984; 79 (388): 871-880.

41. Rousseeuw PJ, Leroy AM. Robust Regression and Outlier Detection. New York, NY, USA: John Wiley & Sons; 1987.

42. Kan B, Alpu O, Yazici B. Robust ridge and robust Liu estimator for regression based on the LTS estimator, Journal of Applied Statistics 2013; 40 (3): 644-655.

43. Tatliyer A. The effects of raising type on performances of some data mining algorithms in lambs. Journal Of Agriculture and Nature 2020; 23 (3): 772-780.

44. Cankaya S. A Comparative study of some estimation methods for parameters and effects of outliers in simple regression model for research on small ruminants. Tropical Animal Health and Production 2009; 41 (1): 35-41.

45. Coskuntuncel O. Eğitimle İlgili Sapan Değer İçeren Veri Kümelerinde En Küçük Kareler ve Robust M Tahmin Edicilerin Karşılaştırılması. Mersin University Journal of the Faculty of Education 2009; 5 (2): 251-262 (in Turkish with an abstract in English).

46. Almetwally EM, Almongy, HM. Comparison between M-estimation, S-estimation, and MM estimation methods of robust estimation with application and simulation. International Journal of Mathematical Archive 2018; 9 (11): 55-63.