Turkish Journal of Agriculture - Food Science and Technology

www.agrifoodscience.com,
Turkish Science and Technology

A Comparative Study of Some Estimation Methods in Simple Linear Regression Model for Different Sample Sizes in Presence of Outliers

Soner Çankaya[1*], Samet Hasan Abacı[2]

[1]*Department of Biostatistics, Faculty of Medicine, Ordu University, 52200 Ordu, Turkey*
[2]*Department of Animal Science, Faculty of Agriculture, Ondokuz Mayıs University, 55139 Samsun, Turkey*

ARTICLE INFO

Article history:
Received 05 January 2015
Accepted 02 February 2015
Available online, ISSN: 2148-127X

Keywords:
Robust regression
Outlier
Karayaka lamb
Chest girth
Body weight

* Corresponding Author:
E-mail: sonercankaya@gmail.com

ABSTRACT

The aim of this study was to compare some estimation methods (LS, M, S, LTS and MM) for estimating the parameters of simple linear regression model in the presence of outlier and different sample size (10, 20, 30, 50 and 100). To compare methods, the effect of chest girth on body weights of Karayaka lambs at weaning period was examined. Chest girth of lambs was used as independent variable and body weight at weaning period was used as dependent variable in the study. Also, it was taken consideration that there were 10-20% outliers of data set for different sample sizes. Mean square error (MSE) and coefficient of determination (R^2) values were used as criteria to evaluate the estimator performance. Research findings showed that LTS estimator is the best models with minimum MSE and maximum R^2 values for different size of sample in the presence of outliers. Thereby, LTS method can be proposed, to predict best-fitted model for relationship between chest girth and body weights of Karayaka lambs at weaning period, to the researches who are studying on small ruminants as an alternative way to estimate the regression parameters in the presence of outliers for different sample size.

Introduction

Regression analysis is an important statistical tool used to fit a model describing quantitative relationship between a response variable (such as body weight) and one or more explanatory variable such as chest girth, chest depth, body length etc. in animal research. To build a regression model, researchers (Benyi, 1997; Atta and El khidir, 2004; Topal and Macit, 2004; Çankaya, 2009; Sarti et al., 2009; Çankaya et al., 2011) have frequently used least squares (LS) method, due to the simplicity of the idea of minimizing the sum of squared residuals and the interpretability of the final model parameter estimates (Pérez et al., 2013). Although LS method achieves optimum results when the underlying error distribution is Gaussian to estimate the weight of live animals, it brings some disadvantages. One of these disadvantages, LS method is sensitive to outliers which can disturb the assumption of normality, one of the most important components of statistical studies. This situation reduces the predictive power of the method (Çankaya et al., 2011). For example, Nsoso et al. (2003) reported that the prediction equation for body weight based on heart girth was very poor (R^2=0.04) during the dry season under extensive management, which is not a true prediction resulted from using an inappropriate method for their study. Successful use of regression requires an appreciation of both the theory and the practical problems that typically arise when the technique is employed with real-world data (Montgomery, 2012).

Outliers may arise for many different reasons such as sampling, human, instrument error etc. and each different reason may require different treatments (Çankaya, 2009). If an outlier arises from a recording or measurement error, in this case elimination of these records may be a good solution. However, if the outliers represent a valid observation, it may point to some significant behavior falling out of range of the model (Zaman et al., 2001). So, robust regression methods such as M-estimation (Huber, 1973) S-estimation (Rousseeuw and Yohai, 1984), LTS (Rousseeuw, 1984) and MM-estimation (Yohai, 1987) are described for the problems. The main propose of robust regression is to provide resistant (stable) results in the presence of outliers (Chen, 2002). Applications of robust regression methods in animal researches began to increase with the availability of related computer packages (Çankaya et al., 2006; Çankaya, 2009; Faustini et al., 2010; Yadav et al., 2011). To our knowledge, there is incomplete knowledge on comparative studies of LS and robust regression methods for different sample size and on comparison of estimation methods for parameters of regression model in animal science. Accordingly, the objectives of the present study were; 1: to estimate the most appropriate mathematical model for defining the relationship between chest girt and body weight for Karayaka lambs in weaning period when the data set was contaminated 10% and 20% with outliers, 2: to compare some estimation methods (LS-, M-, S-, LTS- and MM-estimation) for estimating the parameters of simple linear regression model in the presence of outlier and different sample size (10, 20, 30, 50 and 100).

Materials and Methods

Materials

In this study, the data are the measures of body weight (BW) and chest girth (CG) from totally 197 Karayaka lambs at weaning period which were raised at the Research and Application Farm of Agriculture Faculty of Ondokuz Mayis University.

To evaluate the efficiency of the LS-, M-, S-, LTS- and MM- estimation methods, different sample size (10, 20, 30, 50 and 100) and 10-20% outliers of data groups that making random distribution were created with SPSS statistical package program (SPSS, 1999). Standardized residual test was used to determine whether the outlier in each data set. Significance was evaluated at $P<0.05$ for all tests. All statistical analyses were performed by SAS software (SAS, 2002).

Methods

Regression analysis consists of a collection of techniques that are used to explore relationships between variables (Çankaya, 2009). A main objective of regression analysis is to estimate the unknown parameters in the regression model. This process is also called fitting the model to data. Regression models can be either linear or nonlinear. A linear model assumes the relationships between variables are straight-line relationships, while a nonlinear model assumes the relationships between variables are represented by curved lines (Anonymous, 2013). In animal researches, you will often see the relationship between the body weight and the chest girth measured an animal modeled as a linear relationship. A linear regression model with one predictor variables can be expressed with the following equation:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \text{ (i=1, 2, …, n)}$$

Where the intercept β_0, the slope β_1 are unknown parameters and ε_i is a random error component usually assumed to be normally distributed with mean zero and variance σ^2, Y_i is the dependent variable or response, X_i, is independent variable or the predictor. The intercept β_0 gives the value of Y that is expected when X=0. The slope β_1 is interpreted as the expected change in Y_i for a unit change in X_i (Çankaya, 2009). The data from this experiment were used to predict the body weight based on$y_i = b_0 + b_1 x_i$, where y_i is body weight of i^{th} lamb (kg), x_i is the chest girth of i^{th} lamb (cm), b_0 is the constant and b_1 is the regression coefficient.

Here, the methods used for estimation parameters could be introduced as follows.

Least Squares Method: Least squares method is a procedure to determine the best fit line to data in the presence of normally distributed errors and homoscedasticity (constant variances) (Miller, 2006). The concept of "best fit" requires definition of some measure of the error between the data and the line. In other words, LS method attempts to find an estimate b for β which minimizes some criterion function of residuals where the i^{th} residual $r_i = r_i(b) = y_i - \hat{y}_i$), which is defined as difference between the observed response value y_i and the fitted response value $\hat{y}_i$ (Olive and Hawkins, 2003). LS method chooses $\hat{\beta}$ to minimize

$$Q_{(LS)}(b) = \sum_{i=1}^{n} r_i^2$$

This method consists of the minimization of the sum of the squared residuals. However, in spite of the mathematical beauty and computational simplicity of LS method, this estimator is now being criticized more and more for its dramatic lack of robustness. In addition, even there is a single outlier; it can have a large influence on the results of regression equation (Rousseeuw and Leroy, 1987). Outlier is defined as

$$\text{Outlier} = \begin{cases} 0 & \text{if}|r_i| \le k\sigma \\ 1 & \text{otherwise} \end{cases}$$

Where, by default k=3, and scale σ is computed as corrected median of the absolute residuals.

M- Estimation Method: The most common general method is M-estimation in the context of robust regression was first introduced by Huber (1973) as a result of making the least squares approach robust. M-estimators use an iterative calculations process, whereby an estimate is obtained from eachiteration by weighting the observations according to their distance from the core of the data set. Huber's estimator is an M- estimator possessing the characteristics of robustness and efficiency (Çankaya, 2009; Palmer et al., 2006).

Instead of minimizing a sum of squares of the residuals, a Huber-type M estimator $\hat{\beta}_M$ or b_M of β minimizes a sum of less rapidly increasing functions of the residuals:

$$Q_{(M)}(b) = \min_{\hat{\beta}} \sum_{i=1}^{n} p\left(\frac{r_i}{\sigma}\right)$$

where $r_i = r_i(b) = y_i - \hat{y}_i$. If σ is known, by taking derivatives with respect to β is also a solution of system of p equations:

$$\sum_{i=1}^{n} \psi\left(\frac{r_i}{\sigma}\right) x_{ij} = 0 \qquad j = 1, \ldots, p$$

where, ψ = p'. If p is convex, b_M is the unique solution (Chen, 2002). For the LS estimate, p is the quadratic function,

$$p(r) = \begin{cases} \frac{1}{2} r^2 & for|r| \le k, \\ k|r| - \frac{1}{2} k^2 & for|r| > k, \end{cases}$$

Where k=1.345 for Huber estimator.

The robust version of R^2 for the M estimate is defined as (SAS Institute Inc, 2009);

$$R^2 = \frac{\sum p\left(\frac{y_i - \hat{\mu}}{\hat{s}}\right) - \sum p\left(\frac{y_i - x_i^T\hat{\theta}}{\hat{s}}\right)}{\sum p\left(\frac{y_i - \hat{\mu}}{\hat{s}}\right)}$$

S- Estimation Method: The S estimate proposed by Rousseeuw and Yohai (1984) is defined as the p-vector

$$\hat{\theta}_S = arg \underset{\theta}{min}\, S(\theta)$$

Where the dispersion S(θ) is the solution of

$$\frac{1}{n-p}\sum_{i=1}^{n}\chi\left(\frac{y_i - x_i^T\theta}{S}\right) = \beta$$

β is set to $\int \chi(s) d\,\phi(s)$ such that $\hat{\theta}_S$ and S($\hat{\theta}_S$) are asymptotically consistent estimates of θ and σ for the Gaussian regression model. The breakdown value of the S estimate is

$$\frac{\beta}{sup_s \chi(s)}$$

The robust version of R^2 for the S estimate is defined as

$$R_S^2 = 1 - \frac{(n-p)S_p^2}{(n-1)S_\mu^2}$$

for the model with the intercept term and

$$R_S^2 = 1 - \frac{(n-p)S_p^2}{nS_0^2}$$

for the model without the intercept term, where *Sp* is the S estimate of the scale in the full model, S_μ is the estimate of the scale in the regression model with only the intercept term, and S_0 is the S estimate of the scale without any regressor (SAS Institute Inc, 2009).

LTS Estimation Method: The least trimmed squares (LTS) estimate proposed by Rousseeuw (1984) is defined as the p-vector

$$\hat{\theta}_{LTS} = \arg \underset{\theta}{min} Q_{LTS}(\theta)$$

where

$$Q_{LTS}(\theta) = \sum_{i=1}^{h} r_{(i)}^2$$

$r_{(1)}^2 \le r_{(2)}^2 \le \cdots \le r_{(n)}^2$ are the ordered squared residuals $r_i^2 = (y_i - x_i^T\theta)^2$, i=1,…,n, and h is defined in the range $\frac{n}{2} + 1 + \le h \le \frac{3n+p+1}{4}$.

By default, $h = \left[\frac{3n+p+1}{4}\right]$. The breakdown value is $\frac{n-h}{n}$ for the LTS estimate (Chen, 2002).

The robust version of R^2 for the LTS estimate is defined as

$$R_{LTS}^2 = 1 - \frac{s_{LTS}^2(X,\ y)}{s_{LTS}^2(1, y)}$$

For models with the intercept term and as

$$R_{LTS}^2 = 1 - \frac{s_{LTS}^2(X,\ y)}{s_{LTS}^2(0, y)}$$

For models without the intercept term, where

$$s_{LTS}^2(X, \quad y) = d_{h,n}\sqrt{\frac{1}{h}\sum_{i=1}^{h} r_{(i)}^2}$$

S_{LTS} is a preliminary estimate of the parameter σ in the distribution function $L(\cdot/\sigma)$.

Here $d_{h,n}$is chosen to make s_{LTS} consistent, assuming a Gaussian model (SAS Institute Inc, 2009).

*MM- Estimation Method:*MM estimation is a combination of high breakdown value estimation and efficient estimation, which was introduced by Yohai (1987). MM-estimation method has three stage procedures (Stromberg, 1993; Alma, 2011).

- The first stage is calculating an S-estimation method with influence function

$$p(x) = \begin{cases} 3\left(\frac{x}{c}\right)^2 - 3\left(\frac{x}{c}\right)^4 + 3\left(\frac{x}{c}\right)^6, & if|x| \le c \\ 1 & otherwise \end{cases}$$

The value of tuning constant, c, is selected as 1,548.

- The second stage calculates the parameters that provide the minimum value of $\sum_{i=1}^{n} p\left(\frac{y_{i-}x_i'\hat{\beta}_{MM}}{\hat{\sigma}_0}\right)$ where p(x) is the influence function used in the first stage with tuning constant 4,687 and $\hat{\sigma}_0$ is the estimate of scale form the first step (standard deviation of the residuals).
- The final step computes the MM estimate of scale as the solution to

$$\frac{1}{n-p}\sum_{i=1}^{n} p\left(\frac{y_i - x_i'\hat{\beta}}{s}\right) = 0{,}5$$

The robust version of R^2 for the MM estimate is defined as (SAS Institute Inc, 2009);

$$R^2 = \frac{\sum p\left(\frac{y_i - \hat{\mu}}{\hat{s}}\right) - \sum p\left(\frac{y_i - x_i'\hat{\beta}}{\hat{s}}\right)}{\sum p\left(\frac{y_i - \hat{\mu}}{\hat{s}}\right)}$$

Results

The data obtained from this study was initially examined for being compatible with Shapiro Wilk (n=10, 20 and 30) or Kolmogorov Smirnov test (n=50 and 100) for normal distribution. Descriptive statistics (means, standard deviations, coefficients of variation) and significant values of normality test for body weight and chest girth of the Karayaka lambs at the weaning period with different sample size (n=10, 20, 30, 50 and 100) and 10% outliers are given in Table 1. The average body weight of the Karayaka lambs is between 17.18 - 17.88 kg while the chest girth measurements are ranged from 58.90 to 60.47 cm for the different sample size. In addition, it can be said that the data in terms of variation is a homogeneous structure because the coefficients of variation by BW and CG are generally small than 30%.

In this study, four robust regression (M, MM, LTS and S-estimation) methods were comparatively evaluated against LS regression method in the presence of outlier (10 and 20% outliers in BW variable) and different sample size (10, 20, 30, 50 and 100). Table 2 presents the results of regression analysis in which five estimation methods were used to predict best-fitted model for relationship between CG and BW of Karayaka lambs with different sample size and 10% outliers.

As seen Table 2, if 10% of the data set (BW variable) to be outliers, the model estimated by LTS estimation method was the best model for body weight due to maximum R^2 and minimum MSE values for different sample size in this study. Moreover, the results for 10% outlier showed that LS method's performance was generally increasing while the sample size was increasing.

Descriptive statistics (means, standard deviations, coefficients of variation) and significant values of normality test for body weight and chest girth of the Karayaka lambs at the weaning period with different sample size and 20% outliers are given in Table 3. The average body weight of the Karayaka lambs is between 15.08 - 17.72 kg while the chest girth measurements are ranged from 56.36 to 59.85 cm for the different sample size.

Table 4 presents the results of regression analysis in which five estimation methods were used to predict best-fitted model for relationship between CG and BW of Karayaka lambs with different sample size and 20% outliers.

If 20% of the data set (BW variable) to be outliers, the model estimated by LTS estimation method was the best model for body weight due to maximum R^2 and minimum MSE values for different sample size in this study (Table 4).

Discussion

Sensitivity of established studies in the field of animal science is very important. Because, the pre-values were obtained, it could be to help breeding work in the future. So, the data to be reliable and records must be careful. But in animal science, depending on the care and feeding conditions excepted from measurement or recording errors can be outliers between the data sets. Previous studies indicated that the method of LS is not resistant to outliers. So, simple linear regression model was estimated with LS method. For the whole situation of outliers, the LS method has low explanatory power. The results obtained from this study were similar to other studies. So, robust regression techniques (M-, S-, LTS- and MM-estimation methods) which are more resistant to outliers were used for the simple linear regression estimation. According to the findings of the study, LTS-estimation method which has maximum R^2 and minimum MSE values is the best method from the others. Therefore LTS-estimation method has helped to estimate the best fitted model for different sample size and in the presence of 10-20% outliers. The LTS estimation method is the best model which can explain relationship between CG and BW data. In second place, it was showed that the S-estimation method was the best performance. Also, in previous studies based on regression analysis for estimating response variable from measurement/s or simulated data including outliers, M-estimation (McKean et al., 1993; Çankaya, 2009; Alma, 2011), S-estimation (Çankaya et al., 2011; Alma, 2011), LTS-estimation (Schumacker et al., 2002; Çankaya et al., 2006; Alma, 2011) and MM estimation (Schumacker et al., 2002; Alma, 2011; Çankaya et al., 2011) methods have preferred to least squares method. Alma (2011) were similar findings for LTS estimation method in the multiple regression presence of 10% outlier 5% leverage points and 15% outlier and 5% leverage points. But all results for the research showed that S and M estimation methods perform better than LTS and MM estimation methods.

Table 1 Mean body weight and chest girth of Karayaka lambs in the different sample and 10% outliers.

Traits	n	Means	Std. Deviation	CV (%)	**P***
BW	10	17.25	3.338	19.345	0.828
CG	10	59.02	6.728	11.399	0.077
BW	20	17.39	5.989	34.439	0.560
CG	20	60.47	9.566	15.818	0.703
BW	30	17.18	3.592	20.908	0.090
CG	30	60.17	7.136	11.860	0.263
BW	50	17.64	5.265	29.847	0.085
CG	50	58.90	9.755	16.562	0.200
BW	100	17.88	4.561	25.508	<0.001
CG	100	60.12	8.615	14.329	0.182

*: Sig. Values for Shapiro Wilk and Kolmogorow Smirnov normality test

Table 2 The results of regression analysis for different sample size and 10% outliers

Methods	n		Coef.	Lower Bound	Upper Bound	Sig. Levels (P values)	MSE	R2
LS	10	Constant	2.250	-18.204	23.345	0.783	9.380	0.140
		CG	0.249	-0.101	0.599	0.140		
S		Constant	-14.980	-26.744	-3.217	0.013	2.313	0.650
		CG	0.527	0.333	0.722	<0.001		
M		Constant	-0.485	-18.204	17.234	0.957	3.447	0.277
		CG	0.298	-0.0006	0.569	0.050		
LTS		Constant	-14.698	-25.437	-3.959	0.007	1.474	0.849
		CG	0.524	0.347	0.702	<0.001		
MM		Constant	-14.899	-26.536	-3.263	0.012	2.355	0.511
		CG	0.526	0.334	0.719	<0.001		
LS	20	Constant	-13.018	-24.323	-1.713	0.026	13.445	0.645
		CG	0.503	0.318	0.688	<0.001		
S		Constant	-22.621	-25.863	-19.378	<0.001	1.333	0.957
		CG	0.655	0.602	0.707	<0.001		
M		Constant	-22.570	-25.655	-19.486	<0.001	0.956	0.722
		CG	0.654	0.603	0.704	<0.001		
LTS		Constant	-22.439	-25.256	-19.624	<0.001	0.901	0.971
		CG	0.652	0.606	0.698	<0.001		
MM		Constant	-22.549	-25.682	-19.417	<0.001	1.404	0.718
		CG	0.654	0.603	0.705	<0.001		
LS	30	Constant	-4.195	-12.560	4.169	0.313	6.710	0.498
		CG	0.355	0.217	0.493	<0.001		
S		Constant	-7.104	-18.253	4.045	0.212	2.734	0.489
		CG	0.402	0.219	0.584	<0.001		
M		Constant	-4.621	-13.049	3.808	0.282	3.201	0.465
		CG	0.362	0.223	0.501	<0.001		
LTS		Constant	-14.340	-22.373	-6.307	<0.001	2.042	0.625
		CG	0.513	0.384	0.643	<0.001		
MM		Constant	-5.308	-14.682	4.065	0.267	3.017	0.422
		CG	0.373	0.219	0.527	<0.001		
LS	50	Constant	-5.662	-12.022	0.699	0.080	13.096	0.537
		CG	0.396	0.289	0.502	<0.001		
S		Constant	-12.078	-18.949	-5.207	<0.001	3.558	0.566
		CG	0.486	0.373	0.599	<0.001		
M		Constant	-6.231	-12.613	0.152	0.056	3.484	0.469
		CG	0.400	0.293	0.507	<0.001		
LTS		Constant	-14.296	-19.696	-8.897	<0.001	2.543	0.698
		CG	0.526	0.437	0.615	<0.001		
MM		Constant	-7.282	-14.159	-0.405	0.038	3.707	0.422
		CG	0.414	0.299	0.528	<0.001		
LS	100	Constant	-8.023	-11.767	-4.278	<0.001	7.093	0.663
		CG	0.431	0.369	0.493	<0.001		
S		Constant	-17.890	-21.502	-14.278	<0.001	2.343	0.765
		CG	0.582	0.524	0.641	<0.001		
M		Constant	-11.021	-14.459	-7.582	<0.001	2.636	0.634
		CG	0.476	0.420	0.533	<0.001		
LTS		Constant	-18.579	-21.574	-15.584	<0.001	1.675	0.829
		CG	0.593	0.545	0.642	<0.001		
MM		Constant	-15.476	-19.122	-11.831	<0.001	2.409	0.577
		CG	0.545	0.486	0.604	<0.001		

Table 3 Mean body weight and chest girth of Karayaka lambs in the different sample and 20% outliers.

Traits	n	Means	Std. Deviation	CV (%)	P*
BW	10	15.08	4.494	29.801	0.583
CG	10	59.85	7.764	12.972	0.584
BW	20	17.72	4.648	26.230	0.195
CG	20	56.75	9.889	17.425	0.398
BW	30	17.07	4.917	28.805	0.023
CG	30	56.36	12.605	22.360	0.472
BW	50	17.48	7.027	40.200	0.200
CG	50	57.50	9.528	16.570	0.200
BW	100	16.50	7.910	47.939	0.093
CG	100	56.43	13.460	23.852	0.145

*: Sig. Values for Shapiro Wilk and Kolmogorow Smirnov normality test

Table 4. The results of regression analysis for different sample size and 20% outliers

Methods	n		Coef.	Lower Bound	Upper Bound	Sig. Levels (P values)	MSE	R2
LS	10	Constant	-8.544	-29.360	12.272	0.072	12.157	0.465
		CG	0.395	0.050	0.740	0.030		
S		Constant	-11.303	-29.592	6.987	0.226	3.598	0.513
		CG	0.453	0.146	0.760	0.004		
M		Constant	-14.993	-25.547	-4.438	0.005	1.751	0.471
		CG	0.527	0.352	0.702	<0.001		
LTS		Constant	-14.886	-21.660	-8.113	<0.001	1.223	0.847
		CG	0.526	0.412	0.641	<0.001		
MM		Constant	-9.740	-28.182	8.701	0.301	3.772	0.438
		CG	0.421	0.113	0.729	0.007		
LS	20	Constant	2.089	-8.767	12.945	0.691	14.975	0.344
		CG	0.276	0.087	0.464	0.007		
S		Constant	-15.505	-23.781	-7.229	<0.001	3.139	0.590
		CG	0.557	0.419	0.695	<0.001		
M		Constant	-13.708	-20.140	-7.277	<0.001	2.479	0.404
		CG	0.529	0.417	0.641	<0.001		
LTS		Constant	-22.243	-27.181	-17.305	<0.001	1.174	0.835
		CG	0.663	0.582	0.744	<0.001		
MM		Constant	-13.724	-21.874	-5.574	0.001	3.363	0.392
		CG	0.529	0.393	0.666	<0.001		
LS	30	Constant	2.822	-3.814	9.457	0.391	14.521	0.420
		CG	0.253	0.138	0.368	<0.001		
S		Constant	-2.211	-9.372	4.950	0.545	3.504	0.555
		CG	0.332	0.211	0.453	<0.001		
M		Constant	-0.264	-6.132	5.604	0.929	3.719	0.437
		CG	0.302	0.200	0.403	<0.001		
LTS		Constant	-18.056	-23.071	-13.042	<0.001	1.532	0.817
		CG	0.579	0.498	0.661	<0.001		
MM		Constant	-1.286	-7.517	4.946	0.686	4.343	0.444
		CG	0.317	0.212	0.423	<0.001		
LS	50	Constant	5.912	-6.086	17.910	0.327	46.657	0.074
		CG	0.201	-0.005	0.407	0.055		
S		Constant	-2.274	-14.742	10.194	0.721	6.977	0.079
		CG	0.369	0.147	0.591	0.001		
M		Constant	4.643	-8.192	17.478	0.478	5.467	0.062
		CG	0.227	0.007	0.447	0.043		
LTS		Constant	-14.720	-20.794	-8.646	<0.001	3.009	0.646
		CG	0.622	0.513	0.731	<0.001		
MM		Constant	4.585	-8.223	17.393	0.483	7.381	0.060
		CG	0.228	0.004	0.453	0.045		
LS	100	Constant	1.920	-4.212	8.051	0.536	50.957	0.194
		CG	0.259	0.153	0.364	<0.001		
S		Constant	-2.293	-5.957	1.372	0.220	6.398	0.422
		CG	0.375	0.309	0.439	<0.001		
M		Constant	-0.663	-6.057	4.731	0.809	5.334	0.251
		CG	0.326	0.233	0.418	<0.001		
LTS		Constant	-2.493	-4.872	-0.115	0.039	2.716	0.787
		CG	0.388	0.346	0.430	<0.001		
MM		Constant	-1.447	-6.086	3.191	0.541	6.474	0.276
		CG	0.347	0.266	0.429	<0.001		

Çankaya et al. (2011) recommend that S estimation method was the best performance than the others (MM-Theil estimation). These findings were similar for our study except LTS estimation method. Schumacher et al. (2002) compared OLS, LTS and MM estimation methods. They recommended MM estimation method for the multiple regression but others not recommended the use. Çankaya (2009) different robust techniques (Theil, M, LAD) compared for different sample sizes (n=10, 20, 30 and 50) but LTS- estimation method did not evaluated in his study. M- estimation method was proposed method against outliers. But in our study established that S estimation method is more resistant other methods except LTS- estimation method. MM estimation method is the worst performance in robust techniques. So, studies determined that a linear relationship between the variables, not recommended for use.

In conclusion, observation values obtained from research should be re-examined, in the presence of outliers. However, if there is a real outlier representing information, the use of robust techniques will be useful. Among these techniques, LTS- estimation method, as an alternative to the method of least squares, may increase the degree of accuracy of the model estimates. Furthermore, studies might compare the performance of LTS- estimation method against other robust methods unused in our study under both simple and multiple linear regressions.

Acknowledgements

This study was supported by OYP 1919-001 Project of Ondokuz Mayis University.

References

Alma ÖG. 2011. Comparison of robust regression methods in linear regression. Int. J. Contemp. Math. Sciences, 6 (9), 409-421.

Anonymous. 2013. http://ci.columbia.edu/ci/premba_test/c0331/s7/s7_6.html, Accessed on: 30 July 2013.

Atta M, El khidir OA. 2004. Use of heart girth, wither height and scapuloischial length for prediction of liveweight of Nilotic sheep. Small Ruminant Res, 55 (1): 233-237.

Benyi K. 1997. Estimation of live weight from chest girth in pure and crossbred West African goats. Trop Anim Health Pro, 29 (2): 124-128.

Çankaya S, Eker S, Tahtali Y, Ceyhan A. 2011. Comparison of some estimation methods for parameters of simple regression model in the presence of outliers. 7th National Animal Science Congress, 14-16 September, 2011, Adana, Turkey, p:136-141.

Çankaya S, Kayaalp GT, Sangun L, Tahtali Y, Akar M. 2006. A comparative study of estimation methods for parameters in multiple linear regression model. J Appl Anim Res, 29 (1), 43-47.

Çankaya S. 2009. A comparative study of some estimation methods for parameters and effects of outliers in simple regression model for research on small ruminants. Trop Anim Health Pro, 41 (1): 35-41.

Chen C. 2002. Robust regression and outlier detection with the robustreg procedure. SUGI Paper, 265-27, 2002.

Faustini M, Colombani C, Vigo D, Communod R, Russo V, Chlapanidas T, Munari E, Morandotti A, Torre ML. 2010. Dimensional analysis of milk fat globules in sow milk: effects of the lactation stage and fat content and comparison with vaccine milk. Vet Res Comm, 34 (1), 29-32.

Huber PJ. 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Stat., 1 (5), 799-821.

McKean JW, Sheather SJ, Hettmansperger TP. 1993. The use and interpretation of residuals based on robust estimation. J Am Stat Assoc, 88 (424), 1254-1263.

Miller SJ. 2006. The method of least squares. Mathematics Department Brown University. 1-7.

Montgomery DC, Peck EA, Vining GG. 2012. Introduction to linear regression analysis (Vol. 821). Wiley.

Nsoso SJ, Aganga AA, Moganetsi BP, Tshwenyane SO. 2003. Body weight, body condition score and heart girth in indigenous Tswana goats during the dry and wet seasons in southeast Botswana. Livest Res Rural Dev, 15 (4).

Olive DJ, Hawkins DM. 2003. Robust regression with high coverage. Stat Probabil Lett, 63 (3), 259-266.

Palmer Pol A, Pascual MB, Vázquez PC. 2006. Robust estimators and bootstrap confidence intervals applied to tourism spending. Tourism Manage, 27 (1), 42-50.

Pérez B, Molina I, Peña D. 2013. Outlier detection and robust estimation in linear regression models with fixed group effects. J Stat Comput Sim, (ahead-of-print), 1-18.

Rousseeuw P, Yohai V. 1984. Robust regression by means of S-estimators. In Robust and nonlinear time series analysis, Springer US, 256-272.

Rousseeuw PJ, Leroy A. 1987. Robust regression and outlier detection, John Wiley & Sons, New York.

Rousseeuw PJ. 1984. Least median of squares regression. J Am Stat Assoc, 79 (388), 871-880.

Sarti FM, Castelli L, Bogani D, Panella F. 2009. The measurement of chest girth as an alternative to weight determination in the performance recording of meat sheep. Ital J Anim Sci, 2 (2): 123-129.

SAS Institute Inc. 2009. SAS/STAT ® 9.2, User's Guide, Chapter 74: The ROBUSTREG Procedure, Second Edition, Cary, NC: USA.

SAS Intstitute Inc. 2002. SAS/STAT Software, Version 9. (Cary, NC, USA).

Schumacker RE, Monahan MP, Mount RE. 2002. A comparison of OLS and robust regression using S-PLUS. Multiple Linear Regression Viewpoints, 28, 10-13.

SPSS. 1999. SPSS for Windows (Release 10.0). Standard Version. SPSS, Chicago, IL.

Stromberg AJ. 1993. Computation of high breakdown nonlinear regression parameters. J Am Stat Assoc, 88 (421), 237-244.

Topal M, Macit M. 2004. Prediction of body weight from body measurements in Morkaraman sheep. J Appl Anim Res, 25 (2): 97-100.

Yadav DK, Arora R, Bhatia S, Singh G, Bhar LM. 2011. Robust regression for determining selection criterion of breeding rams. Indian J Small Rumin, 17 (2), 166-169.

Yohai VJ. 1987. High breakdown-point and high efficiency robust estimates for regression. Ann. Stat, 642-656.

Zaman A, Rousseeuw PJ, Orhan M. 2001. Econometric applications of high-breakdown robust regression techniques. Economics Letters, 71 (1), 1-8.