DOI: 10.29050/harranziraat.504857
Harran Tarım ve Gıda Bilimleri Derg. 2019, 23(3): 277-286

Research Article/Araştırma Makalesi

Evaluating of beef cattle performance and profitability using robust regression analysis

Besi sığırı performansının ve karlılığının dayanıklı regresyon analizi ile değerlendirilmesi

Hayati KOKNAROGLU[1], Sukru ACITAS[2*], Birdal SENOGLU[3]

[1]Department of Animal Science, Isparta University of Applied Sciences, Isparta, Turkey,
[2]Department of Statistics, Eskisehir Technical University, Eskisehir, Turkey,
[3]Department of Statistics, Ankara University, Turkey,

To cite this article:
Koknaroglu, H., Acitas, S. & Senoglu, B. (2019). Evaluating of beef cattle performance and profitability using robust regression analysis. Harran Tarım ve Gıda Bilimleri Dergisi, 23(3): 277-286.
DOI: 10.29050/harranziraat.504857

Address for Correspondence:
Sukru ACITAS
e-mail:
sacitas@eskisehir.edu.tr

Received Date:
28.12.2018
Accepted Date:
21.06.2019

ABSTRACT

There are several variables affecting the beef cattle performance and profitability. In the related literature, regression analysis is performed to find the contribution of the variables on profitability in different managing systems. Least squares (LS) estimation method is mostly used in regression analysis. However, it is optimal when the distribution of the error terms is normal. In this study, we revise Koknaroglu et al. (2005) study in which regression analysis is used under normality assumption. Different from the mentioned study, we use a robust estimation method called M-estimation since the error terms do not follow a normal distribution according to Shapiro-Wilk normality test. We obtain parameter estimates and their standard errors along with the coefficient of determination. It is observed that the results obtained based on M-estimation are more reliable than their LS counterparts with respect to R^2 criterion.

Key Words: Beef cattle, Profitability, Regression, Robustness, M-estimation.

ÖZ

Besi sığırı performansını ve karlılığını etkileyen birkaç değişken vardır. İlgili literatürde, farklı yönetim sistemlerinde karlılığı etkileyen değişkenlerin katkısını bulmak için regresyon analizi kullanılmıştır. Regresyon analizinde çoğunlukla en küçük kareler (LS) tahmin yöntemi kullanılır. Ancak, bu yöntem hata terimlerinin dağılımının normal olması durumunda optimaldir. Bu çalışmada, normallik varsayımı altında regresyon analizinin kullanıldığı Koknaroğlu ve ark. (2005) çalışması revize edilmiştir. Shapiro-Wilk normallik testine göre hata terimleri için normallik varsayımını sağlanmadığından, diğer çalışmadan farklı olarak, bu çalışmada, M-tahmini adı verilen dayanıklı/robust tahmin yöntemi kullanılmıştır. Belirleme katsayısı ile birlikte parametre tahminleri ve onların standart hataları elde edilmiştir. Sonuç olarak, R^2 kriterine göre, M-tahminine dayalı olarak elde edilen sonuçların, LS tahmin edicilerinden daha güvenilir olduğu gözlenmiştir.

Anahtar Kelimeler: Besi sığırı, Karlılık, Regresyon, Dayanıklılık, M-tahmini.

Introduction

Regression analysis is widely used statistical technique in most of the applied sciences. The linear regression model is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. + \beta_p x_{ip} + \varepsilon_i,\ i = 1,2,\dots,n \quad (1)$$

where y_i is the response, $x_{i1}, x_{i2} \dots, x_{ik}$ are the predictors, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the model parameters and ε_i denote the random error term. This model can also be written by using alternative representations:

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1,2,\dots,n \qquad (2)$$

and

$$y = X\beta + \varepsilon \qquad (3)$$

where

$$y = [y_1 \ \ y_2 \ \cdots \ y_n]', \qquad x_i' = [1 \ x_{i1} \ \ x_{i2} \cdots \ x_{ip}], \qquad \beta = [\beta_0 \ \beta_1 \ \beta_2 \ \cdots \ \beta_k]',$$
$$X = [1 \ x_1 \ x_2 \cdots x_p], \qquad 1 = [1 \ 1 \ \cdots \ 1]' \ and \ \ \varepsilon = [\,\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n]'.$$

The model parameters are estimated using well-known least squares (LS) method. The idea underlying the LS method is to minimize the sum of the squares of error terms with respect to the parameters of interest. Then the well-known LS estimator is formulated by

$$\hat{\beta} = (X'X)^{-1}X'y. \qquad (4)$$

Traditionally, the error terms are assumed to be independently and identically distributed normal with mean zero and variance σ^2 in model (1). This assumption has a vital role in statistical inference, i.e. the LS estimators are the most efficient and the test statistics based on them are the most powerful under normality. However, nonnormality is more common in practical studies, see for example Geary (1947), Huber (1981), Islam and Tiku (2005), Senoglu (2005), Koknaroglu et al. (2008), Acitas et al. (2013a, 2013b), Acitas and Senoglu (2019). In the presence of nonnormality, the LS estimators may become inefficient. Furthermore, normal-theory test statistics are not reliable since they are no longer powerful.

Robust statistical methods are frequently used when existing of nonnormality. The aim of the robust statistical methods is to reduce the large effects of the outliers which cause nonnormality. This is done by giving small weights to the outlying observations. There are several robust estimation methods such as M (Huber; 1964, 1981), MM (Yohai, 1987), least median squares (LMS) (Rousseeuw, 1984), modified maximum likelihood (MML) (Tiku; 1967, 1968) and so on. M-estimation is the most widely used and popular method among them. See also Karadavut and Taşkın (2017) in which M estimation method is used for determination of outliers in Japanese quail body weight data.

In this study, we use model (1) to identify variables that affect beef cattle profitability. Main purpose of animal production is to make living by earning money thus, profitability is the driving force behind animal production. There are several documented variables affecting beef cattle performance and profitability (Koknaroglu et al., 2005). Thus finding contribution of variables on profitability in different managing systems becomes important. Then, the model is formulated by

$$profit_i = \beta_0 + \beta_1 fedprice + \beta_2 feedgain + \beta_3 fe + \beta_4 corn + \beta_5 adg + \varepsilon_i \qquad (5)$$

where i refers to the pen of cattle. See Koknaroglu et al. (2005) for detailed information on descriptions of the variables used in Model (5).

We revise Koknaroglu et al. (2005) study in which same model is used. Different than the mentioned study, we here use M-estimation method. The reason using M-estimation method is that the error terms do not follow a normal distribution according to Shapiro-Wilk normality test. It should be noted that we do not consider other assumptions regarding to regression model (5) since our aim is to provide a different approach to explore the contribution of several variables on beef cattle profitability in terms of robust regression aspect by revising Koknaroglu et al. (2005).

The rest of the paper is organized as follows. Material and methods are given in section 2. Section 3 is reserved to the results and the discussion. The paper ends with a conclusion section.

Material and Methods

In this section, data structure, the results of Shapiro Wilk normality test and brief information about the M-estimation method are provided.

Data structure

Close-out information, consisting of data from cattle that were placed on feed between January 1988 and December 1997, which had been submitted by Iowa cattle producers using the Iowa State University Feedlot Performance and Cost Monitoring program, was examined to determine factors affecting beef cattle performance and profitability.

The model given in equation (5) is employed to identify variables that affect cattle feeding profitability. All of the variables used in model (5) are continuous. Detailed information on materials and methods on how data are obtained and categorized is provided in (Koknaroglu et al., 2005).

The effects of season, housing, sex, body weight (BW) and concentrate level on the profit are also considered. For this purpose, model (5) is applied for different levels of the following variables: season, housing, sex, BW and concentrate level. A brief information about these variables are given as follows. Season has four levels as known well: Winter, spring, summer and fall. Housing includes three levels named confinement, overhead shelter and open lot. Steers and Heifers are two levels of Sex. Initial BW has three levels: cattle weighing <273 kg, between 273 and 364 kg, and >364 kg and finally concentrate level consists of three levels: low (<75%), intermediate (between 75 and 85%), and high (>85%). Therefore, in total 15 regression models should be taken into account during the analysis.

M-estimation method

M-estimators of the model parameters are solutions of the following minimization problem:

$$\hat{\beta}_M = \underset{\beta}{\operatorname{argmin}} \sum \rho\left(\frac{y_i - x_i'\beta}{\sigma}\right) \quad (6)$$

where $\rho(\cdot)$ is the objective function. After taking derivative in equation (6) with respect to β and setting equal to zero, $\hat{\beta}_M$ can also be obtained as solution of the following equation:

$$\sum \psi\left(\frac{y_i - x_i'\beta}{\sigma}\right) x_i = 0 \quad (7)$$

where $\psi(\cdot) = \rho'(\cdot)$. It should be noted that LS estimators are obtained when $\rho(x) = x^2$. Different choices of $\rho(\cdot)$ function are considered to capture the robustness in the related literature. Indeed, $\rho(\cdot)$ function should satisfy some properties for the sake of robustness. These properties are given in Maronna et al. (2006) on page 31. Jureckova and Picek (2010) can also be seen for different $\rho(\cdot)$ functions.

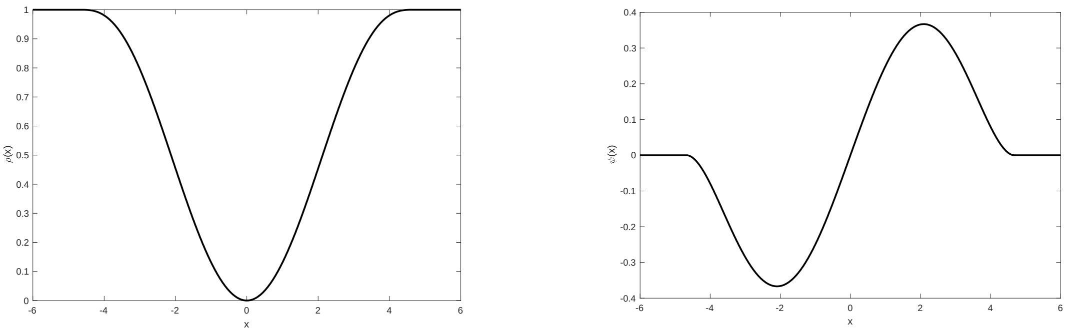
In this study, we use Tukey's bisquare (biweight) function:

$$\rho(x) = \begin{cases} 1 - \left[1 - \left(\frac{x}{k}\right)^2\right]^3 & if \quad |x| \le k \\ 1 & if \quad |x| > k. \end{cases} \quad (8)$$

Therefore, $\psi(\cdot)$ function is obtained by

$$\psi(x) = \begin{cases} \frac{6}{k^2} x \left[1 - \left(\frac{x}{k}\right)^2\right]^2 & if \quad |x| \le k \\ 0 & if \quad |x| > k \end{cases}. \quad (9)$$

Here, k is the robustness tuning constant which is used for adjusting the trade-off between the robustness and the efficiency. Therefore, it is taken as 4.68. Tukey's bisquare function is mostly used in robust statistical analyzes since ψ function is redescending, i.e. ψ tends to 0 for ∞. The advantage of using a redescending ψ function is that it provides more robustness (Maronna et al., 2006). The plots of $\rho(x)$ and $\psi(x)$ functions are given in Figure 1.



(a) $\boldsymbol{\rho(x)}$ (b) $\psi(x)$

Figure 1. Plots of $\rho(x)$ and $\psi(x)$ functions
Şekil 1. $\rho(x)$ ve $\psi(x)$ fonksiyonlarının grafikleri

It is clear that solutions of equation (7) cannot be obtained explicitly. Therefore, numerical methods should be performed. In the related literature, iteratively reweighted least squares (IRLS) method is mostly utilized to compute the M-estimates, see for example Montgomery (2012). The computations in this study are done using "robustfit" function of MATLAB software. Robustfit function gives M-estimates of the model parameters in addition to robust estimate of the scale. We also compute the robust coefficient of determination (R^2) given by

$$R_w^2 = 1 - \frac{\sum w_i (y_i - \hat{y}_i)^2}{\sum w_i (y_i - \bar{y}_w)^2} \quad (10)$$

where

$$\bar{y}_w = \frac{1}{\sum w_i} \sum w_i y_i, \quad \hat{y}_i = y_i - x_i' \hat{\beta}_M$$

and w_i $(i = 1,2,\dots,n)$ are the weights, see Renaud and Victoria-Fesser (2010). Obviously, the higher values of R^2 implies better fitting.

Results and Discussion

In this part of the study, we first explore the normality of the data set. Then we analyze the data set using the M-estimation method and interpret the results.

Normality test results

In this study, we use the Shapiro-Wilk (1965), one of the widely used and the most powerful goodness of fit test, to explore the normality of the data set. The appropriate hypotheses are stated as follows:

Ho: The error terms have a normal distribution

H1: The error terms do not have a normal distribution

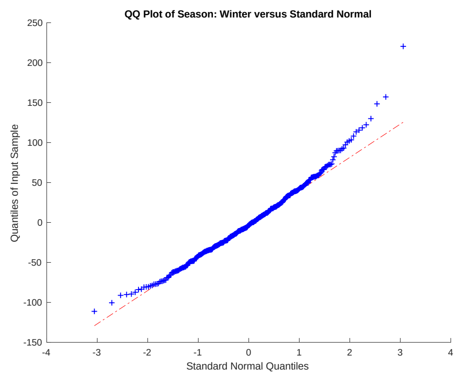
The test is conducted using LS residuals. Indeed, we first compute the LS estimates of the model parameters using equation (4) and then obtain the residuals based on them. The test is conducted in MATLAB using Oner and Deveci's (2017) code file which is available at the website provided in references.

As it is indicated previously, there are five variables. Model (5) is used for each level of these variables. Therefore, we have 15 regression models belonging to the different levels of season, housing, sex, body weight and concentrate level. Shapiro-Wilk test is carried out to LS residuals obtained from these 15 regression models. In other words, first LS residuals are obtained for each 15 regression models. Then, they are used to check the normality assumption via Shapiro-Wilk test. The results are given in Table 1. It reports value of the test statistics and the corresponding p −values.

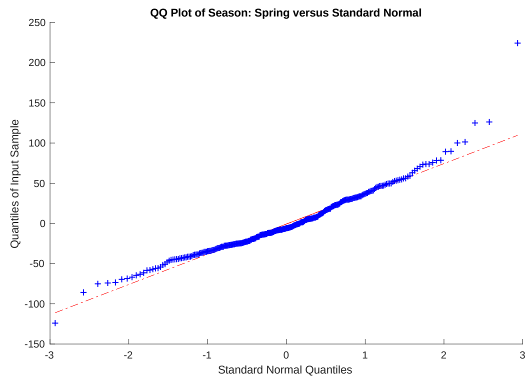
The results are interpreted at $\alpha = 0.05$ significance level as follows. It is clear that the normality assumption is not satisfied for Winter, Spring and Summer while it is satisfied for Fall. For three levels of housing variable, the normality of the error terms is rejected. While the error terms are distributed normally for Heifers, the distribution is not normal for Steers. First two levels of BW, the normality assumption is satisfied. However, third level (>364) does not follow a normal distribution. For the concentrate levels (75 to 85 and >85), the normality assumption is met. On the other hand, for the level <75, the null hypothesis is rejected.

Table 1. Shapiro-Wilk normality test results
Çizelge 1. Shapiro-Wilk normallik testi sonuçları

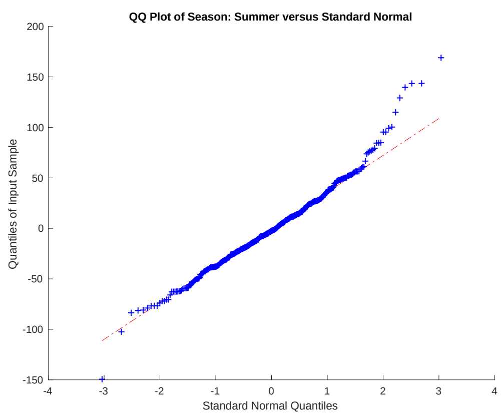
		Sample Size (Örneklem hacmi)	Test Statistic (Test İstatistiği)	p −value (p −değeri)	Not reject/Reject (Red değil/Red) (0/1)
Season (Mevsim)	Winter (Kış)	448	0.9798	0.0001	1
	Spring (İlkbahar)	300	0.9559	0.0001	1
	Summer (Yaz)	420	0.9769	0.0001	1
	Fall (Sonbahar)	685	0.9964	0.1171	0
Housing (Barınma)	Confinement (Kapalı ahır)	456	0.9089	0.0001	1
	Overhead shelter (Yarı açık ahır)	470	0.9930	0.0273	1
	Open lot (Açık ahır)	927	0.9929	0.0002	1
Sex (Cinsiyet)	Steers (Erkek Dana)	1429	0.9834	0.0001	1
	Heifers (Dişi Dana)	424	0.9942	0.1046	0
BW (Vücut Ağırlığı)	<273	371	0.9832	0.0003	1
	273 to 364	964	0.9699	0.0001	1
	>364	518	0.9947	0.0713	0
Concentrate Level (Yoğunluk Düzeyi)	<75	180	0.9895	0.2062	0
	75 to 85	1020	0.9928	0.0001	1
	>85	653	0.9505	0.0001	1



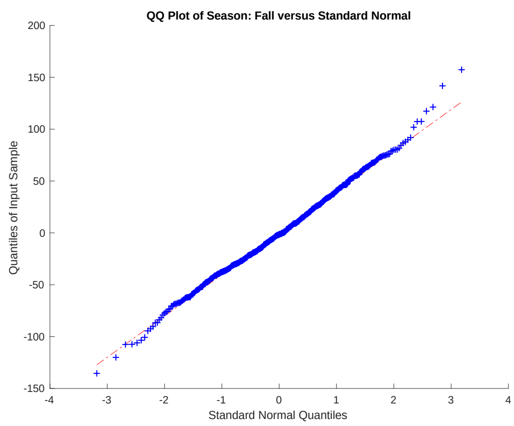
(a) Winter



(b) Spring



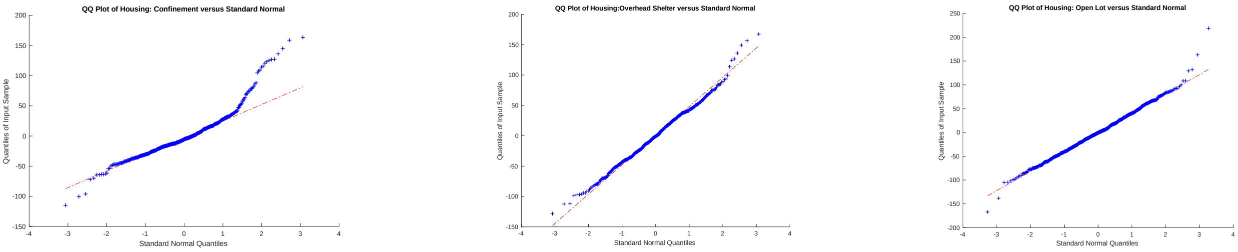
(c) Summer



(d) Fall

Figure 2. Q-Q plots for season
Şekil 2. Mevsim için Q-Q grafikleri

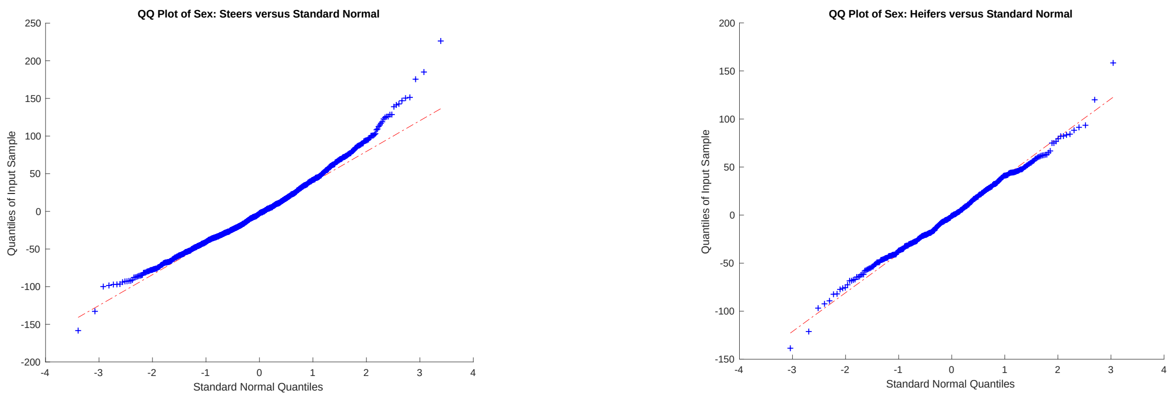
These results are also supported by Figures 2 - 6 in which Q-Q plots are given for all categories. It should be noted that the normality assumption is satisfied if the quantile pairs do not deviate too much from the straight line.



(a) Confinement (b) Overhead shelter (c) Open lot

Figure 3. Q-Q plots for housing

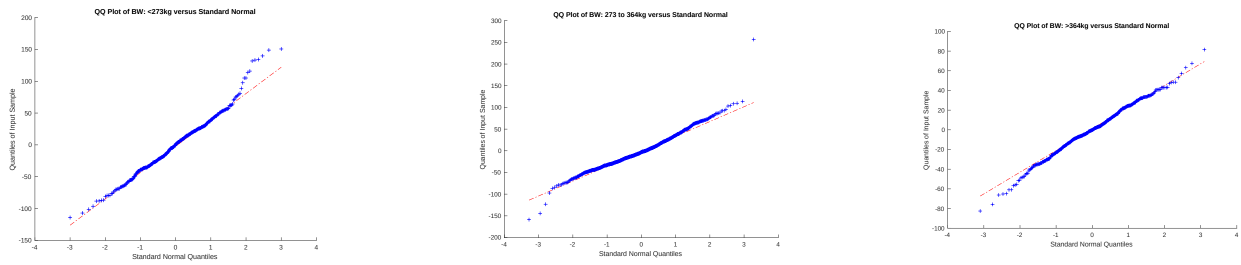
Şekil 3. Barınak için Q-Q grafikleri



(a) Steers (b) Heifels

Figure 4. Q-Q plots for sex

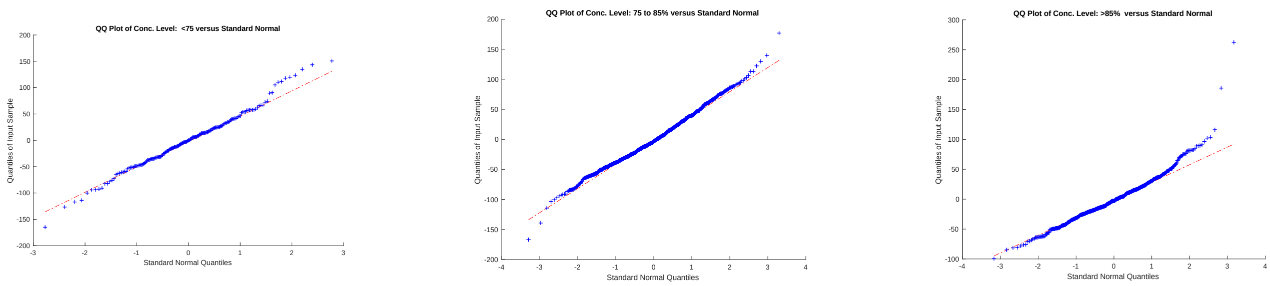
Şekil 4. Cinsiyet için Q-Q grafikleri



(a) <273 (b) 273 to 364 (c) >364

Figure 5. Q-Q plots for BW

Şekil 5. BW için Q-Q grafikleri



(a) <75 (b) 75 to 85 (c) >85

Figure 6. Q-Q plots for concentrate level

Şekil 6. Yoğunluk düzeyi için Q-Q grafikleri

Table 2. Estimated regression coefficients of the factors explaining profitability, $\hat{\sigma}$ and R^2. The values given in parenthesis are the standard error of the regression estimates

Çizelge 2. Karlılığı açıklayan faktörler için tahmin edilmiş regresyon katsayıları, $\hat{\sigma}$ ve R^2. Parantez içinde verilen değerler regresyon tahminlerinin standart hatalarıdır

				$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\sigma}$	R^2(%)
Season (Mevsim)	Winter (Kış)	LS	Coeff.	90.28	9.76	-4.90	-27.98	-29.65	-64.57	45.49	61
			SE	(63.46)	(0.43)	(0.29)	(3.36)	(5.55)	(17.02)		
		M	Coeff.	54.51	9.98	-5.12	-25.96	-27.52	-52.49	44.43	68
			SE	(62.10)	(0.42)	(0.28)	(3.28)	(5.43)	(16.65)		
	Spring (İlkbahar)	LS	Coeff.	86.54	10.46	-6.01	-27.90	-34.63	-25.22	39.63	67
			SE	(66.36)	(0.54)	(0.32)	(3.17)	(6.66)	(17.03)		
		M	Coeff.	60.65	11.11	-6.76	-24.74	-45.38	4.12	36.31	77
			SE	(61.41)	(0.50)	(0.30)	(2.93)	(6.17)	(15.76)		
	Summer (Yaz)	LS	Coeff.	-99.89	11.50	-5.40	-24.73	-21.85	-26.55	40.36	71
			SE	(56.74)	(0.43)	(0.23)	(2.45)	(6.66)	(11.91)		
		M	Coeff.	-93.99	12.37	-6.76	-20.67	-34.25	0.52	34.32	84
			SE	(48.98)	(0.37)	(0.20)	(2.12)	(5.75)	(10.28)		
	Fall (Sonbahar)	LS	Coeff.	230.99	8.60	-4.68	-34.61	-41.47	-58.06	40.52	67
			SE	(40.24)	(0.30)	(0.19)	(1.85)	(5.19)	(10.08)		
		M	Coeff.	235.02	8.64	-4.75	-35.19	-40.94	-57.56	40.72	72
			SE	(40.44)	(0.30)	(0.19)	(1.86)	(5.21)	(10.13)		
Housing (Barınma)	Confinement (Kapalı ahır)	LS	Coeff.	207.38	9.60	-6.15	-31.94	-41.75	-34.17	37.02	74
			SE	(49.84)	(0.34)	(0.21)	(2.84)	(6.59)	(12.51)		
		M	Coeff.	136.07	10.29	-6.95	-27.72	-45.67	9.30	26.24	89
			SE	(36.59)	(0.25)	(0.15)	(2.09)	(4.84)	(9.18)		
	Overhead shelter (Yarı açık ahır)	LS	Coeff.	74.29	9.99	-4.45	-34.69	-20.65	-68.17	45.63	61
			SE	(62.99)	(0.42)	(0.28)	(2.88)	(6.52)	(15.30)		
		M	Coeff.	47.67	10.19	-4.53	-33.96	-20.29	-58.43	45.68	66
			SE	(63.06)	(0.42)	(0.28)	(2.89)	(6.53)	(15.32)		
	Open lot (Açık ahır)	LS	Coeff.	117.66	9.44	-4.98	-30.09	-30.55	-46.86	41.44	66
			SE	(37.15)	(0.27)	(0.18)	(1.63)	(3.71)	(8.72)		
		M	Coeff.	122.75	9.54	-5.19	-29.88	-31.21	-43.74	41.41	71
			SE	(37.12)	(0.27)	(0.18)	(1.63)	(3.71)	(8.72)		
Sex (Cinsiyet)	Steers (Erkek Dana)	LS	Coeff.	165.34	9.76	-5.22	-34.21	-33.06	-60.76	43.34	65
			SE	(33.23)	(0.23)	(0.15)	(1.60)	(3.48)	(8.17)		
		M	Coeff.	155.72	10.02	-5.52	-34.00	-33.22	-51.59	42.23	72
			SE	(32.39)	(0.22)	(0.15)	(1.56)	(3.39)	(7.97)		
	Heifers (Dişi Dana)	LS	Coeff.	3.38	9.33	-4.41	-25.26	-20.04	-36.36	38.77	66
			SE	(49.28)	(0.39)	(0.24)	(2.03)	(5.33)	(14.53)		
		M	Coeff.	-20.61	9.42	-4.41	-24.23	-19.32	-30.24	38.96	70
			SE	(49.53)	(0.39)	(0.24)	(2.04)	(5.36)	(14.60)		

Table 2. (Cont.)
Çizelge 2. (Devamı)

				$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\sigma}$	R^2(%)
BW (Vücut Ağırlığı)	<273	LS	Coeff.	-132.46	9.94	-3.84	-21.84	-37.18	35.73	44.08	64
			SE	(64.76)	(0.46)	(0.22)	(3.20)	(6.79)	(21.66)		
		M	Coeff.	-124.52	10.10	-3.85	-23.05	-42.95	38.39	42.83	70
			SE	(63.09)	(0.45)	(0.22)	(3.12)	(6.61)	(21.10)		
	273 to 364	LS	Coeff.	42.37	10.30	-5.86	-25.79	-39.63	5.93	36.48	74
			SE	(33.54)	(0.23)	(0.16)	(1.68)	(3.40)	(9.43)		
		M	Coeff.	22.60	10.52	-6.04	-24.20	-40.59	11.51	35.10	80
			SE	(32.32)	(0.22)	(0.16)	(1.62)	(3.28)	(9.09)		
	>364	LS	Coeff.	-16.96	11.42	-8.03	-16.97	-35.58	39.97	23.72	89
			SE	(31.54)	(0.22)	(0.16)	(1.42)	(3.43)	(7.77)		
		M	Coeff.	-0.99	11.57	-8.32	-16.55	-40.57	43.56	23.62	91
			SE	(31.41)	(0.22)	(0.16)	(1.42)	(3.42)	(7.74)		
Concentrate Level (Yoğunluk Düzeyi)	<75	LS	Coeff.	-19.63	10.86	-4.57	-29.20	-29.41	-33.59	53.36	60
			SE	(94.89)	(0.81)	(0.40)	(4.05)	(10.64)	(23.47)		
		M	Coeff.	-100.44	11.40	-4.52	-27.25	-27.50	-19.91	51.86	67
			SE	(92.68)	(0.79)	(0.39)	(3.95)	(10.39)	(22.92)		
	75 to 85	LS	Coeff.	104.35	9.70	-4.96	-31.08	-29.23	-48.68	40.59	67
			SE	(35.56)	(0.25)	(0.16)	(1.69)	(3.74)	(9.03)		
		M	Coeff.	83.55	9.69	-4.93	-30.18	-28.47	-41.02	40.73	71
			SE	(35.68)	(0.25)	(0.16)	(1.70)	(3.75)	(9.07)		
	>85	LS	Coeff.	159.23	9.88	-5.94	-32.44	-39.70	-26.98	35.62	74
			SE	(38.59)	(0.27)	(0.19)	(1.93)	(4.35)	(9.55)		
		M	Coeff.	172.36	10.43	-7.18	-28.60	-48.36	-1.34	29.18	87
			SE	(32.02)	(0.23)	(0.16)	(1.60)	(3.61)	(7.92)		

Regression analysis results

This section includes LS and M-estimates of the model parameters. Estimated regression coefficients of the variables explaining profitability are given in Table 2. Furthermore, LS and robust estimates of scale parameter ($\hat{\sigma}$) and the coefficient of determination (R^2) are provided in Table 2. The standard errors (SEs) are also given under the corresponding regression estimate in parenthesis. It should also be noted that robust estimate of scale parameter and standard errors for M-estimates of regression parameters are obtained from robustfit function. Following results can be deduced from Table 2.

Season

The normality assumption is not satisfied for Winter, Spring and Summer. Therefore, standard errors of the M-estimates are less than their LS counterparts. Furthermore, $\hat{\sigma}_M < \hat{\sigma}_{LS}$ implies that robust linear regression model is more reliable. This conclusion is also supported by R^2 values, i.e. $R_M^2 > R_{LS}^2$. For the Fall, normality assumption is valid. Therefore, LS results are preferable here.

Housing

The normality assumption violated for three levels of housing variable. Therefore, M-estimates are more reliable for this case. This conclusion is obtained from the standard errors of the estimated regression coefficients and the standard deviation of the error terms. $R_M^2 > R_{LS}^2$ also implies to use robust statistical methods.

Sex

The distribution of the error terms is not normal for Steers. Therefore, robust statistical methods should be used. The results show that standard errors of the M-estimates are less than those of LS estimates. However, LS estimates are more preferable for Heifers since normality assumption is satisfied for this case.

BW

First two levels of BW do not satisfy the normality assumption. Therefore, M-estimation method should be used to estimate the model parameters for these levels. Indeed, standard errors for M-estimates are smaller than their LS counterparts. This is also true for estimate of the scale parameter. The coefficient of determination for M-estimation is also higher. However, for the third level, LS estimates are more preferable since normality is satisfied.

Concentrate level

The normality assumption is satisfied for the first level (<75). Therefore, LS estimates should be preferred for this case. On the other hand, for the remaining levels (75 to 85% and >85%), the normality is not satisfied. Indeed, the M-estimates are more reliable for these cases.

Conclusion

Regression analysis is a widely used method in animal sciences. For example, the beef cattle performance and profitability can be determined using the regression analysis. The parameters of the linear regression model are frequently estimated using LS method. It is a well-known fact that LS method is optimal when the distribution of the error terms is normal. However, nonnormality of the error terms is more common in practice. The motivation for this study comes from this fact. Koknaroglu et al. (2005) evaluate the beef cattle performance using linear regression model under normality. However, in our analyzes, we find that normality is not satisfied for most of the cases. We therefore use robust regression methods, i.e. M-estimation. The advantage of using a robust method is that it is not sensitive to the outliers and also to the nonnormality. Therefore, it gives more reliable results in presence of outliers and nonnormality.

References

Acitas, S., Kasap, P., Senoglu, B. & Arslan, O. (2013a). Robust estimation with the skew t2 distribution. *Pakistan Journal of Statistics*, *29*(4), 409-430.

Acitas, S., Kasap, P., Senoglu, B. & Arslan, O. (2013b). One-step M-estimators: Jones and Faddy's skewed t-distribution. *Journal of Applied Statistics*, *40*(7), 1545-1560.

Acitas, S., & Senoglu, B. (2019). Ridge-Type MML Estimator in the Linear Regression Model. *Iranian Journal of Science and Technology, Transactions A: Science*, *43*(2), 589-599.

Geary, R. C. (1947). Testing for normality. *Biometrika*, *34*(3/4), 209-242.

Huber, P. J. (1964). Robust estimation of a location parameter. *The annals of mathematical statistics*, *35*(1), 73-101.

Huber, P. J. (1981). *Robust statistics*. New York: John Wiley & Sons, Inc.

Islam, M. Q. & Tiku, M. L. (2005). Multiple linear regression model under nonnormality. *Communications in Statistics-Theory and Methods*, *33*(10), 2443-2467.

Jureckova, J. & Picek, J. (2005). *Robust statistical methods with R*. Chapman and Hall/CRC.

Karadavut, U. & Taşkın, A. (2017). Determination of Outliers in Growing Quail's Data with Different Sample Size, *Harran Tarım ve Gıda Bilimleri Dergisi*, 21(1), 99-111.

Koknaroglu, H., Loy, D. D., Wilson, D. E., Hoffman, M. P. & Lawrence, J. D (2005). Factors affecting beef cattle performance and profitability. *The Professional Animal Scientist*, *21*(4), 286-296.

Koknaroglu, H., Turan C. & Toker, M. T. (2008). Animal science application of robust tests: effect of zeolite and initial weight on fattening performance of cattle. *Animal Science Papers and Reports*, *26*(2), 107-115.

Maronna, R. A., Martin, R. D. & Yohai, V. (2006). *Robust statistics: Theory and Methods*. Chichester: John Wiley & Sons.

Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012). *Introduction to linear regression analysis*. New Jersey: John Wiley & Sons.

Oner, M. & Deveci Kocakoc, I. (2017). JMASM 49: A Compilation of Some Popular Goodness of Fit Tests for Normal Distribution: Their Algorithms and MATLAB Codes (MATLAB). *Journal of Modern Applied Statistical Methods*, *16*(2), 547-575.

https://www.mathworks.com/matlabcentral/fileexchange/60147-normality-test-package.

Renaud, O. & Victoria-Feser, M. P. (2010). A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference*, *140*(7), 1852-1862.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*, 871–880.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611.

Senoglu, B. (2005). Robust 2^k factorial design with Weibull error distributions. *Journal of Applied Statistics*, *32*(10), 1051-1066.

Tiku, M. L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, *54*, 155-165.

Tiku, M. L. (1968). Estimating the parameters of normal and logistic distributions from censored samples. *Australian & New Zealand Journal of Statistics*, *10*(2), 64-74.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, *15*, 642-656.