



The prediction of live weight of hair goats through penalized regression methods: LASSO and adaptive LASSO

Suna Akkol

Biometry and Genetic Unit, Department of Animal Science,
Faculty of Agriculture, Van Yuzuncu Yil University, Van, Turkey

Correspondence: Suna Akkol (sgakkol@yyu.edu.tr)

Received: 7 August 2018 – Revised: 3 November 2018 – Accepted: 8 November 2018 – Published: 19 November 2018

Abstract. The least absolute selection and shrinkage operator (LASSO) and adaptive LASSO methods have become a popular model in the last decade, especially for data with a multicollinearity problem. This study was conducted to estimate the live weight (LW) of Hair goats from biometric measurements and to select variables in order to reduce the model complexity by using penalized regression methods: LASSO and adaptive LASSO for $\gamma = 0.5$ and $\gamma = 1$. The data were obtained from 132 adult goats in Honaz district of Denizli province. Age, gender, forehead width, ear length, head length, chest width, rump height, withers height, back height, chest depth, chest girth, and body length were used as explanatory variables. The adjusted coefficient of determination (R_{adj}^2), root mean square error (RMSE), Akaike's information criterion (AIC), Schwarz Bayesian criterion (SBC), and average square error (ASE) were used in order to compare the effectiveness of the methods. It was concluded that adaptive LASSO ($\gamma = 1$) estimated the LW with the highest accuracy for both male ($R_{adj}^2 = 0.9048$; RMSE = 3.6250; AIC = 79.2974; SBC = 65.2633; ASE = 7.8843) and female ($R_{adj}^2 = 0.7668$; RMSE = 4.4069; AIC = 392.5405; SBC = 308.9888; ASE = 18.2193) Hair goats when all the criteria were considered.

1 Introduction

Native goat breeds play important socio-economic roles in the livelihood strategies of poorer farmers, especially those in rural and hard-to-reach areas of the world. Turkey has one of the largest goat populations in the world and has one of the highest breeding rates. The total number of goats in the country is about 10.3 million and the dominant goat breed is the “Common”, or “Hair”, goat, which constitutes approximately 92 % of the total goat population in the country (TUIK, 2017). Goats have been kept for milk, meat, skin, and hair for several centuries in Anatolia (Gokdal, 2013).

Studies to define adult live weights and body measurements are of great importance for the characterization of farm animal breeds. The prediction of body weight (BW) and the determination of its relationships with other biometric measurements generates considerable knowledge for breeding research relating to meat production per animal

(Iqbal et al., 2013; Yılmaz et al., 2013; Khan et al., 2014). Multiple linear regression (MLR), based on ordinary least squares (OLS), is a traditional, simple method that has been used by researchers in order to predict the complex relationship between live weight and some body measurements in goat, sheep, cattle, fish, etc. (Francis et al., 2002; Pesmen and Yardimci, 2008; Yılmaz et al., 2013). However, when a multicollinearity problem exists among explanatory variables, the OLS method produces poor predictions (Montgomery et al., 2001; Yakubu, 2010; Dormann et al., 2013; Khan et al., 2014). The multicollinearity problem implies that the standard errors of regression coefficients are higher than expected, and thus it is difficult to find out the accuracy and robustness of the prediction models (Weisberg, 2005; Yakubu, 2009, 2010; Sangun et al., 2009).

Penalized methods based on minimizing the residual sum of squares are an alternative to OLS method for data with multicollinearity problems. Ridge regression is one of them;

it overcomes the multicollinearity problem by using l_2 -norm in order to shrink the regression coefficients (Hoerl and Kennard, 1970; Marquardt and Snee, 1975; Dormann et al., 2013). Ridge regression has been previously used by some researchers working on the prediction of live weight (Malau-Aduli et al., 2004; Yakubu, 2009; Topal et al., 2010). It works by keeping all the explanatory variables in the model; however, it cannot perform variable selection (Zou and Hastie, 2005). However, variable selection is as important as prediction in a model with a large number of explanatory variables. The other penalized method used in the current study is the least absolute shrinkage and selection operator (LASSO) proposed by Tibshirani (1996). LASSO uses l_1 -norm and executes both automatic variable selection and continuous shrinkage simultaneously (Zou and Hastie, 2005; Wang et al., 2011). These properties make LASSO a popular variable selection method (Wang et al., 2011; Ogutu et al., 2012; Akkol et al., 2018). However, LASSO has some important limitations in practice (Zou and Hastie, 2005). One of them is that LASSO selects only one or a few variables and shrinks the rest to 0 if the model includes a number of correlated explanatory variables (Zou and Hastie, 2005; Wang et al., 2011). This might be an undesirable feature in many studies. Fan and Li (2001) showed that LASSO does not produce unbiased estimates for large coefficients and that LASSO does not possess oracle properties. Zou (2006) introduced the adaptive LASSO (ALASSO) estimators to remedy the problem, by adding data-defined weights to the original LASSO version. He showed that ALASSO can have oracle properties if the weights are dependent on the data and are wisely chosen. In his study, Zou used LASSO and ALASSO for $\gamma = 0.5$, $\gamma = 1$, and $\gamma = 2$ and revealed that ALASSO is closer to the true model than LASSO and also that ALASSO for $\gamma = 1$ is closer to the true model than the one for $\gamma = 0.5$.

The aim of this study was to estimate the LW of Hair goats from biometric measurements for the purpose of selection for genetic improvement and breeding program in the field to select variables in order to reduce the model complexity and to determine the best model to explain the change in LW by performing ALASSO. Therefore, multiple linear regression was performed to determine a potential multicollinearity problem; then the Ridge, LASSO, and ALASSO methods for $\gamma = 0.5$ and for $\gamma = 1$ were compared to each other in order to obtain the best fit model.

2 Material and methods

2.1 Material

The data of the study comprised measurements from a total of 132 Hair goats from the Honaz district of Denizli province in Turkey. The data included age, gender, live weight, and 10 biometric measures of goats: forehead width (FW), ear length (EL), head length (HL), chest width (CW), rump height (RH), withers height (WH), back height (BH), chest

depth (CD), chest girth (CG), and body length (BL) were recorded in the breeding season. Live weights of the goats were determined with a digital scale. CW, RH, WH, BH, CD, and BL were measured with a measuring stick, and FW, EL, HL, and CG were measured with a measuring tape.

2.2 Methods

The basic multiple linear regression model used to predict the live weight with the LASSO and ALASSO model:

$$Y = \mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + e, \quad (1)$$

where $Y = (y_1, y_2, \dots, y_n)^T$ is a vector of observed dependent variables, $\mathbf{1}_n$ is a column vector of n variables ($i = 1, 2, 3, \dots, n$), μ is the intercept, \mathbf{X} is an $n \times p$ matrix of explanatory variables, $\boldsymbol{\beta}$ is the vector of regression coefficients, and e is the vector of the residuals with a mean of zero and a variance $I\sigma_e^2$. It was assumed that observed independent variables have been mean-centered in regularized linear regression.

2.2.1 LASSO regression

In the OLS method, $\boldsymbol{\beta}$ coefficients are estimated by minimizing the sum of residuals squares (RSSs). This is expressed as an optimization problem by the following equation

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} |Y - X\boldsymbol{\beta}|^2. \quad (2)$$

The following equation in the Lagrangian form is used to calculate the regression coefficients with LASSO.

$$\hat{\boldsymbol{\beta}}(\text{lasso}) = \arg \min_{\boldsymbol{\beta}} |Y - X\boldsymbol{\beta}|^2 + \lambda |\boldsymbol{\beta}|_1, \quad (3)$$

where $|Y - X\boldsymbol{\beta}|^2 = \sum_{i=1}^n (y_i - x_i^T \boldsymbol{\beta})^2$ is the loss function, $|\boldsymbol{\beta}|_1 = \sum_{j=1}^p |\beta_j|$ is the l_1 -norm penalty on $\boldsymbol{\beta}$, and $\lambda \geq 0$ is a tuning (penalty or shrinkage) parameter which regulates strength of penalty and is important for the success of LASSO. For the LASSO estimate Eq. (3) is rewritten without an intercept (Hastie et al., 2009):

$$\hat{\boldsymbol{\beta}}(\text{lasso}) = \arg \min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (4)$$

The penalty function called ℓ_1 is important for the success of LASSO.

2.2.2 Adaptive LASSO regression

ALASSO modifies the original LASSO penalty by adding weights for each parameter to the penalty term. These weights are data-defined weights, $\hat{\omega}_j$, and they control the

shrinking of the zero coefficients more than the non-zero coefficients. The ALASSO estimates $\hat{\beta}$ (alasso) are given by

$$\hat{\beta}(\text{alasso}) = \arg \min_{\beta} \left\| y_i - \sum_{j=1}^p x_{ij} \beta_j \right\|^2 + \lambda \sum_{j=1}^p \hat{\omega}_j |\beta_j| \quad (5)$$

where $\hat{\omega}_j = 1/|\hat{\beta}_j^{\text{ini}}|^\gamma$ is a known weights vector, γ is a positive constant > 0 , and $\hat{\beta}_j^{\text{ini}}$ is the initial consistent estimator of β obtained from ordinary least square or ridge regression if there is a multicollinearity problem (Zou, 2006; Ogutu et al., 2012). When the parameter estimates produced by ALASSO are defined by $\hat{\beta}(\lambda_n)$, then

$$\hat{\beta}(\lambda_n) = \arg \min \left\{ \left\| Y - \sum_{j=1}^p X_j \beta_j \right\|^2 + \sum_{j=1}^p \lambda_n \hat{\omega}_j |\beta_j| \right\}. \quad (6)$$

It was proved that ALASSO has the oracle property when $\lambda_n \rightarrow \infty$ and $\lambda_n/\sqrt{n} \rightarrow 0$ (Fan and Li, 2001; Zou, 2006).

2.2.3 Model selection

The adjusted coefficient of determination (R_{adj}^2), the Akaike information criterion (AIC), the Schwarz Bayesian information criterion (SBC), and the average square error (ASE) are cohesion criteria used to compare LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) results in the model selection. They are called goodness-of-fit measurements, and for a statistical model this shows inconsistency between the observed and expected values (Maydeu-Olivares and García-Forero, 2010).

$$R_{\text{adj}}^2 = 1 - \left(1 - R^2\right) \frac{n-1}{n-p-1} \quad (7)$$

In Eq. (7), R^2 shows the coefficient of determination, p is the total number of explanatory variables in the model not including the constant, and n shows the sample size. AIC (Akaike, 1974) and SBC (Schwarz, 1978) are

$$\begin{aligned} \text{AIC} &= -2\text{ll} + 2p, \\ \text{SBC} &= n \ln(\text{SSE}/n) + p \ln(n). \end{aligned} \quad (8)$$

“ll” shows the log likelihood, and SSE is the sum of square error. The ASE is another cohesion criterion.

$$\text{ASE} = \frac{\sum_{i=1}^n \left(Y_{\text{new}} - \left(\hat{\beta}_0 + \sum_{i=1}^{p-1} \hat{\beta}_j X_{\text{new},j} \right) \right)^2}{n}, \quad (9)$$

where Y_{new} and X_{new} express new data that are unusable to estimate the coefficients of β . The model having minimum AIC, SBC, and ASE values is determined to be the best when selecting the model.

Table 1. Descriptive statistics regarding live weight and some body measurements for male and female goats.

Variable	Male (n = 35)		Female (n = 97)	
	Mean \pm SE	CV	Mean \pm SE	CV
LW	79.298 \pm 1.986*	14.82	58.405 \pm 0.927	15.62
FW	15.650 \pm 0.144*	5.44	13.165 \pm 0.108	8.02
EL	19.980 \pm 0.440	13.04	20.036 \pm 0.351	17.21
HL	18.430 \pm 0.276	8.850	17.915 \pm 0.122	6.70
CW	22.150 \pm 0.253*	6.76	20.567 \pm 0.175	8.39
RH	84.900 \pm 0.886*	6.17	77.217 \pm 0.389	4.96
WH	89.150 \pm 1.010*	6.70	79.619 \pm 0.390	4.82
BH	82.800 \pm 0.949*	6.78	75.258 \pm 0.406	5.35
CD	39.950 \pm 0.489*	7.24	34.041 \pm 0.186	5.40
CG	101.000 \pm 1.115*	6.53	89.211 \pm 0.492	5.43
BL	82.150 \pm 0.740*	5.33	74.923 \pm 0.401	5.27
Age	3.25 \pm 0.741*	28.01	4.16 \pm 1.129	30.06

* Differences from the females are statistically significant ($P < 0.05$). SE: standard error; CV: coefficient of variation; LW: live weight; FW: forehead width; EL: ear length; HL: head length; CW: chest width; RH: rump height; WH: withers height; BH: back height; CD: chest depth; CG: chest girth; BL: body length.

The statistical evaluations were performed by using MEANS, CORR, GLM, and GLMSELECT procedures in SAS (2014). The R program was used to create a figure showing the correlations. The GLM procedure was used to eliminate age effect before performing OLS, and then the Ridge, LASSO, and ALASSO methods were applied.

3 Results

There were 35 male (26.52 %) and 97 female (73.48 %) goats in the study. Descriptive statistics regarding LW and biometric measurements (CW, RH, WH, BH, CD, BL, FW, EL, HL, CG, and age) and the results of univariate analysis of variance for all of variables in both genders are given in Table 1. It was observed that there were significant differences ($P < 0.05$) between the genders for all the biometric measurement of Hair goats, except for EL and HL.

The analyses were made after the data were corrected according to age. Pearson correlation coefficients displaying relationships between live weight and body measurements of Hair goats are presented by gender in Fig. 1. The values for males are shown in Fig. 1a, and those for females are shown in Fig. 1b. In Fig. 1, correlation coefficients greater than 0.5 were found to be statistically significant for males ($P < 0.01$); whereas for females, coefficients greater than 0.26 were significant ($P < 0.01$). There were correlation coefficients of over 0.8 between the explanatory variables in both genders, which made these data suitable for examination.

Regression coefficients, standard errors, tolerance values (TVs), and variance inflation factor (VIF) values are shown in Table 2 for both genders. The results revealed that all explanatory variables in the model explained 88.62 % of the variation in BL for males and 76.45 % for females. As

Table 2. OLS coefficients in multiple linear regression, tolerance, and VIF values for male and female goats.

Variable	Male			Female		
	Coefficients	TVs	VIF	Coefficients	TV	VIF
Intercept	$(-145.47 \pm 31.923)^b$		0	$(-119.73 \pm 11.65)^b$		0
FW	(2.972 ± 2.634)	0.164	6.08	(0.437 ± 0.565)	0.574	1.74
EL	(-0.546 ± 0.482)	0.502	1.92	$(-0.311 \pm 0.144)^a$	0.830	1.21
HL	(0.809 ± 0.771)	0.523	1.91	(0.312 ± 0.459)	0.673	1.49
CW	(0.648 ± 0.962)	0.320	2.51	(0.637 ± 0.341)	0.591	1.70
RH	(0.731 ± 1.490)	0.013	79.67	(0.121 ± 0.500)	0.056	17.90
WH	$(2.060 \pm 0.696)^a$	0.050	20.93	(0.523 ± 0.345)	0.094	10.69
BH	(-1.467 ± 1.160)	0.020	51.36	(-0.383 ± 0.370)	0.079	12.67
CD	(-1.880 ± 1.054)	0.090	11.24	(0.104 ± 0.514)	0.229	4.36
CG	$(0.875 \pm 0.377)^a$	0.134	7.47	$(0.992 \pm 0.198)^a$	0.221	4.52
BL	(0.272 ± 0.549)	0.143	7.00	$(0.611 \pm 0.162)^a$	0.500	1.20
RMSE	3.963			4.429		
R^2	0.946			0.789		
R^2_{adj}	0.886			0.765		

^a $p < 0.05$. ^b $p < 0.01$. TVs: tolerance values; VIF: variance inflation factor values; LW: live weight; FW: forehead width; EL: ear length; HL: head length; CW: chest width; RH: rump height; WH: withers height; BH: back height; CD: chest depth; CG: chest girth; BL: body length; RMSE: root mean square error; R^2 : the coefficient of determination; R^2_{adj} : the adjusted coefficient of determination.

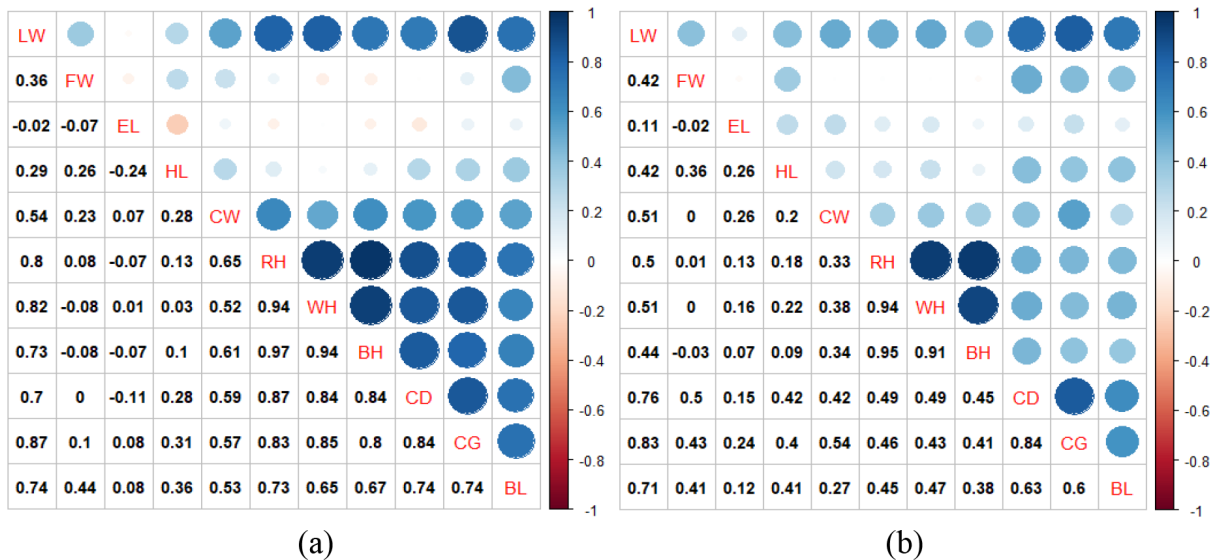


Figure 1. Pearson correlation coefficients between live weight and biometric body measurements for male (a) and female (b) goats.

shown in Table 2, there were VIF values of more than 10. VIF values for RH, WH, BH, and CD were found to be 77, 21, 51, and 11, respectively, in males. VIF values of RH, WH, and BH for females were 18, 11, and 13, respectively.

The coefficients and the standardized coefficients of Ridge, LASSO, and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) in multiple linear regression are given in Table 3 for males and in Table 4 for females. The estimation equation for Ridge included all explanatory variables for both males and females, whereas LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) reduced the number of explanatory variables. In order to compare the

methods some goodness-of-fit measurements such as R^2_{adj} , AIC, SBC, and ASE are presented in Table 5, which shows that R^2_{adj} varied between 79.62 % and 90.48 % for males and between 74.95 % and 76.68 % for females.

In the current study we present the coefficient progression with AIC in Fig. 2a and b because we use AIC as a selection criterion. The selection process was done solely as visualized in Fig. 2. When the lowest AIC value was provided, the variable selection process was completed. As seen in Fig. 2, seven explanatory variables were selected for males: FW, EL,

Table 3. Coefficients and standardized coefficients of Ridge, LASSO, and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) in multiple linear regression for male goats.

VN	Ridge		LASSO		ALASSO ($\gamma = 0.5$)		ALASSO ($\gamma = 1$)	
	Coefficients	Standardized coefficients	Coefficients	Standardized coefficients	Coefficients	Standardized coefficients	Coefficients	Standardized coefficients
FW	3.767 ± 1.373	0.273	4.570 ± 2.285	0.330	4.556 ± 0.88	0.330	4.564 ± 0.984	0.331
EL	-0.108 ± 0.438	-0.024	-0.204 ± 0.158	-0.047	-2.135 ± 0.106	-0.047	-0.241 ± 0.131	-0.053
HL	0.462 ± 0.739	0.0644	0.703 ± 0.578	0.100	0.727 ± 0.606	0.100	0.742 ± 0.551	0.103
CW	-0.024 ± 0.876	-0.003	-	-	-	-	-	-
RH	0.265 ± 0.184	0.118	-	-	-	-	-	-
WH	0.796 ± 0.209	0.796	1.623 ± 0.557	0.826	1.802 ± 0.353	0.917	1.928 ± 0.331	0.981
BH	-0.007 ± 0.203	-0.003	-0.362 ± 0.192	-0.173	-0.527 ± 0.369	-0.249	-0.637 ± 2.408	-0.305
CD	-0.371 ± 0.508	-0.091	-0.951 ± 0.831	-0.234	-0.1.022 ± 1.753	-0.252	-1.125 ± 0.571	-0.277
CG	0.674 ± 0.223	0.378	0.769 ± 0.204	0.432	0.771 ± 0.292	0.433	0.797 ± 0.345	0.448
BL	0.096 ± 0.325	0.036	-	-	-	-	-	-

VN: variable name; LW: live weight; FW: forehead width; EL: ear length; HL: head length; CW: chest width; RH: rump height; WH: withers height; BH: back height; CD: chest depth; CG: chest girth; BL: body length.

Table 4. Coefficients and standardized coefficients of Ridge, LASSO, and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) in multiple linear regression for female goats.

VN	Ridge		LASSO		ALASSO ($\gamma = 0.5$)		ALASSO ($\gamma = 1$)	
	Coefficients	Standardized coefficients	Coefficients	Standardized coefficients	Coefficients	Standardized coefficients	Coefficients	Standardized coefficients
FW	0.608 ± 0.405	0.070	-	-	0.608 ± 0.405	-0.056	-	-0.0720
EL	-0.184 ± 0.116	-0.070	-	-	-0.184 ± 0.116	-	-	-
HL	0.455 ± 0.350	0.060	-	-	0.455 ± 0.350	-	-	-
CW	0.728 ± 0.245	0.138	0.276 ± 0.077	0.052	0.728 ± 0.245	0.064	0.276 ± 0.077	0.074
RH	0.119 ± 0.092	0.049	-	-	0.119 ± 0.092	-	-	-
WH	0.197 ± 0.106	0.083	0.127 ± 0.056	0.054	0.197 ± 0.106	0.085	0.127 ± 0.056	0.092
BH	-0.034 ± 0.096	-0.015	-	-	-0.034 ± 0.096	-	-	-
CD	0.663 ± 0.251	0.134	0.299 ± 0.59	0.060	0.663 ± 0.251	-	0.299 ± 0.59	-
CG	0.641 ± 0.094	0.340	0.936 ± 0.119	0.497	0.641 ± 0.094	0.579	0.936 ± 0.119	0.580
BL	0.548 ± 0.114	0.237	0.621 ± 0.079	0.269	0.548 ± 0.114	0.292	0.621 ± 0.079	0.299

VN: variable name; LW: live weight; FW: forehead width; EL: ear length; HL: head length; CW: chest width; RH: rump height; WH: withers height; BH: back height; CD: chest depth; CG: chest girth; BL: body length.

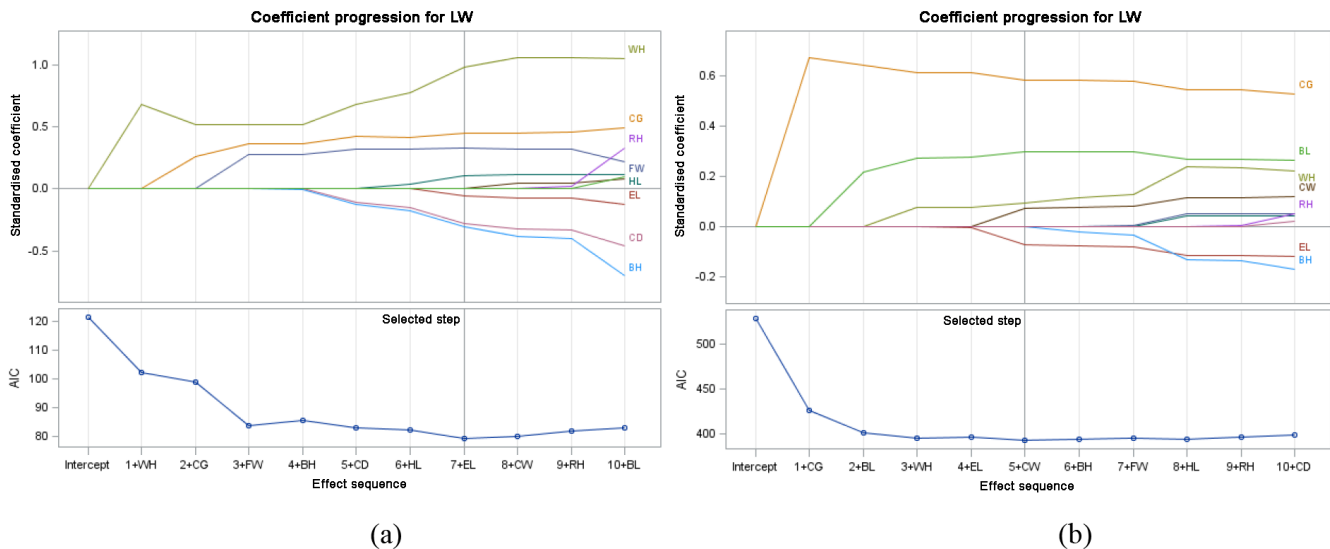


Figure 2. Coefficient progression with ALASSO ($\gamma = 1$) for male (a) and female (b) goats.

Table 5. Goodness-of-fit criteria for estimation equations of Ridge, LASSO, and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) for male and female goats.

GFC	Male				Female			
	Ridge	LASSO	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)	Ridge	LASSO	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)
NV	10	7	7	7	10	5	5	5
R^2	0.893	0.935	0.938	0.940	0.7756	0.764	0.777	0.779
R^2_{adj}	0.797	0.896	0.9028	0.905	0.7495	0.752	0.765	0.767
RMSE	5.590	3.783	3.682	3.625	4.5669	4.549	4.426	4.407
AIC	–	81.008	79.918	79.297	–	398.701	393.396	392.541
SBC	–	66.974	65.884	65.263	–	315.150	309.844	308.989
ASE	–	8.588	8.133	7.884	–	19.414	18.381	18.219

GFC: goodness-of-fit criteria; NV: number of variables; R^2 : coefficient of determination; R^2_{adj} : adjusted coefficient of determination; RMSE: root mean square error; AIC: Akaike information criterion; SBC: Schwarz Bayesian criterion; ASE: average square error.

HL, WH, BH, CD, and CG. Five variables (FW, CW, WH, CG, and BL) were selected for females.

4 Discussion

The present results show that there was a significant difference between the genders in terms of body measurements in this study ($P < 0.05$), with all measurements larger in males than females apart from ear length, despite females being on average older than the males. Similar results were reported by other researchers (Khan et al., 2014; Akbaş and Saatci, 2016). EL and HL were not measured in the study of Akbaş and Saatci (2016).

The correlation between LW and CG was found to be 0.87 for males and 0.83 for females (Fig. 1). The highest correlation coefficient with LW was revealed by CG for both genders. This was in agreement with the finding of previous studies (Pesmen and Yardimci, 2008; Cam et al., 2010; Tsegaye et al., 2013; Das and Yadav, 2015; Sam et al., 2016). The present study was focused the correlations between explanatory variables. Because there were high and significant correlations between explanatory variables, this study examined whether there was a multicollinearity problem. Previous studies have reported that when the tolerance values were less than 0.1 and VIF values were more than 10, the data had a multicollinearity problem (Montgomery et al., 2001; Yakubu, 2010; Dormann et al., 2013). According the results of OLS methods in MLR, the tolerance values found for RH, WH, BH, and CD in males were 0.01255, 0.04779, 0.01947, and 0.08894, respectively, and corresponding VIF values were 77, 21, 51, and 11 (Table 2). Tolerance and VIF values for RH, WH, and BH in females were 0.05589, 0.09356, and 0.07891 and 18, 11, and 13 (Table 2). This result revealed that the current data set had a multicollinearity problem for both genders. It was emphasized by researchers that the multicollinearity implies that standard errors of regression coefficients are higher than expected, and, thus, it is difficult to find out the accuracy and robustness of the predic-

tion models (Weisberg, 2005; Yakubu, 2009, 2010; Sangun et al., 2009).

In this study, where the variable selection for the data with multicollinearity is important, stepwise regression was not discussed because a previous study proposed that stepwise regression had some limitations and problems (Fan and Li, 2001; Shen and Ye, 2002; Whittingham et al., 2006). The body weight has been predicted from body structural and udder morphological traits in Frizarta dairy sheep, and it has been claimed that stepwise and LASSO regression selected the same variables with equal goodness-of-fit measurements (Kominakis et al., 2009). However, Kominakis et al. (2009) did not mention the multicollinearity problem.

In Ridge regression (in which coefficients of all explanatory variables are estimated), the adjusted R^2 values were 78.62 % for males and 74.94 % for females. Also, variable selection could not accomplished as reported in previous research (Pimentel et al., 2007; Topal et al., 2010; Ogutu et al., 2012; Orhan et al., 2016). Subsequently, LASSO and ALASSO for both $\gamma = 0.5$ and $\gamma = 1$ were performed to overcome the multicollinearity problem and also to select explanatory variables for the purpose of reducing model complexity. In all three methods, models consisted of seven variables for males and five variables for females. The adjusted coefficient of determination was 89.63 % for LASSO and 90.18 % and 90.48 % for ALASSO (for $\gamma = 0.5$ and $\gamma = 1$ methods, respectively) for male Hair goats (Table 3). ALASSO ($\gamma = 1$) had the highest adjusted coefficient of determination. According to the model, FW, EL, HL, WH, BH, CD, and CG were selected as significant explanatory variables. The adjusted coefficient of determination of female Hair goats for the three methods was found to be 75.15 % (LASSO), 76.47 % (ALASSO, $\gamma = 0.5$), and 76.66 % (ALASSO, $\gamma = 1$) (Table 4). The method giving the highest adjusted R^2 was again ALASSO ($\gamma = 1$), which selected the variables FW, CW, WH, CG, and BL. When all methods were evaluated in terms of an adjusted coefficient

of determination, it was found that Ridge regression gave the lowest coefficient in both genders of Hair goats.

When considering goodness-of-fit measurements for all methods (RMSE, AIC, SBC, and ASE), except for Ridge regression, ALASSO ($\gamma = 1$) had the smallest value in both male and female goats. From this finding it was concluded that the best model explaining the change in LW was ALASSO ($\gamma = 1$) in both genders of Hair goats. This is the first study to examine the ALASSO method with multilevel linear regression method to predict live weight from some biometric measurements and to select variables. Consequently, this study revealed that the best method explaining the variation in LW of male and female Hair goats is ALASSO ($\gamma = 1$). The fact that ALASSO was a better method than LASSO was consistent with the findings of previous researchers (Fan and Li, 2001; Zou, 2006; Huang et al., 2008; Ogutu et al., 2012). They proposed that the ALASSO method was more advantageous compare to LASSO method due to its oracle property.

In this study, the results from ALASSO ($\gamma = 1$) revealed that WH had the highest significant effect on LW in male goats, and the second main significant effect was CG. These were in agreement with the findings of the previous study (Yakubu, 2009), whereas many studies propose CG as the most important predictor (Cam et al., 2010; Tsegaye et al., 2013; Sam et al., 2016; Das and Yadav, 2015). The analysis of data having a multicollinearity problem should be treated with caution since the problem has been shown to be associated with unstable estimates of regression coefficients (Montgomery et al., 2001; Yakubu, 2010; Dormann et al., 2013; Khan et al., 2014). This justifies the use of ALASSO methods for prediction. However, the results of female Hair goats showed that CG was the main significant effect in LW. The same result was supported by Kominakis et al. (2009), Cam et al. (2010), Tsegaye et al. (2013), and Das and Yadav (2015).

5 Conclusions

In this study, LW was predicted from biometric measurement with high accuracy for both male and female Hair goats by using ALASSO ($\gamma = 1$). However, the variable selection was performed by ALASSO ($\gamma = 1$), unlike in Ridge. New statistical techniques like penalized regression methods can be successfully implemented in the investigation of relationships between LW and biometric measurements in goat, sheep, cattle, fish, etc.

Data availability. Data sets are available upon request by contacting the correspondence author.

Competing interests. The author declares that there is no conflict of interest.

Acknowledgements. I would like to thank Ibrahim Cemal and agricultural engineer Mustafa Varol for their permission to use the project data.

Edited by: Manfred Mielenz

Reviewed by: Ghobad Asgari Jafarabadi and one anonymous referee

References

- Akaike, H.: A New Look at the Statistical Model Identification, *IEEE T. Automat. Contr.*, 19, 716–724, <https://doi.org/10.1109/TAC.1974.1100705>, 1974.
- Akbaş, A. A. and Saatci, M.: Growth, slaughter, and carcass characteristics of Honamlı, Hair, and Honamlı × Hair (F1) male goat kids bred under extensive conditions, *Turk. J. Veterin. Anim. Sci.*, 40, 459–467, <https://doi.org/10.3906/vet-1511-5>, 2016.
- Akkol, S., Arpalı, D., and Yagmur, M.: Adaptive Lasso analysis for grain yield and yield components in two-rowed barley under rainfed conditions, *Comptes rendus de l'Académie bulgare des Sciences*, 71, 1279–1287, <https://doi.org/10.7546/CRABS.2018.09.17>, 2018.
- Cam, M. A., Olfaz, M., and Soydan, E.: Possibilities of using morphometrics characteristics as a tool for body weight prediction in Turkish hair goats (Kilkeci), *Asian J. Anim. Vet. Adv.*, 5, 52–59, <https://doi.org/10.3923/ajava.2010.52.59>, 2010.
- Das, S. K. and Yadav, B. P. S.: Productive performance of Jamunapary goat under semi intensive system of management, *Indian J. Anim. Res.*, 49, 254–257, <https://doi.org/10.5958/0976-0555.2015.00114.4>, 2015.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, *Ecography*, 36, 27–46, <https://doi.org/10.1111/j.1600-0587.2012.07348.x>, 2013.
- Fan, J. and Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Stat. Assoc.*, 96, 1348–1360, <https://doi.org/10.1198/016214501753382273>, 2001.
- Francis, J., Sibanda, S., and Kristensen, T.: Estimating body weight of cattle using linear body measurements, *Zimbabwe Veterin. J.*, 33, 15–21, <https://doi.org/10.4314/zvj.v33i1.5297>, 2002.
- Gokdal, O.: Growth, slaughter and carcass characteristics of Alpine × Hair goat, Saanen × Hair goat and Hair goat male kids fed with concentrate in addition to grazing on rangeland, *Small Rumin. Res.*, 109, 69–75, <https://doi.org/10.1016/j.smallrumres.2012.07.013>, 2013.
- Hastie, T. J., Tibshirani, R., and Friedman, J.: *The Elements of Statistical Learning: Prediction, Inference and Data Mining*, 2nd Edn., Springer Verlag, California, 2009.
- Hoerl, A. E. and Kennard, R. W.: Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, 12, 55–67, <https://doi.org/10.1080/00401706.1970.10488634>, 1970.

- Huang, J., Ma, S., and Zhang, C. H.: Adaptive Lasso for sparse high-dimensional regression models, *Statist. Sin.*, 18, 1603–1618, 2008.
- Iqbal, M., Javed, K., and Ahmad, N.: Prediction of body weight through body measurements in Beetal goats, *Pakistan J. Sci.*, 65, 458–461, 2013.
- Khan, M. A., Tariq, M. M., Eydurán, E., Tatliyer, A., Rafeeq, M., Abbas, F., Rashid, N., Awan, M. A., and Javed, K.: Estimating body weight from several body measurements in harnai sheep without multicollinearity problem, *J. Anim. Plant Sci.*, 24, 120–126, 2014.
- Kominakis, A. P., Papavasiliou, D., and Rogdakis, E.: Relationships among udder characteristics, milk yield and, non-yield traits in Frizarta dairy sheep, *Small Rumin. Res.*, 84, 82–88, <https://doi.org/10.1016/j.smallrumres.2009.06.010>, 2009.
- Malau-Aduli, A. E. O., Aziz, M. A., Kojina, T., Niibayashi, T., Oshima, K., and Komatsu, M.: Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle, *J. Anim. Veterin. Adv.*, 3, 856–863, 2004.
- Marquardt, D. W. and Snee, R. D.: Ridge regression in practice, *Am. Statist.*, 29, 3–19, <https://doi.org/10.1080/00031305.1975.10479105>, 1975.
- Maydeu-Olivares, A. and García-Forero, C.: Goodness-of-fit testing, *Int. Encyclop. Educ.*, 7, 190–196, 2010.
- Montgomery, D. C., Peck, E. A., and Vining, G. G.: Introduction to Linear Regression Analysis, 3rd Edn., Jon Wiley and Sons, New York, USA, 2001.
- Ogutú, J. O., Schulz-Streeck, T., and Piepho, H.: Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions, *BMC Proc.*, 6, 10, <https://doi.org/10.1186/1753-6561-6-S2-S10>, 2012.
- Orhan, H., Eydurán, E., Tatliyer, A., and Saygici, H.: Prediction of egg weight from egg quality characteristics via ridge regression and regression tree methods, *Revista Brasileira de Zootecnia*, 45, 380–385, <https://doi.org/10.1590/S1806-92902016000700004>, 2016.
- Pesmen, G. and Yardimci, M.: Estimating the live weight using some body measurements in Saanen goats, *Archiva Zootechnica*, 11, 30–40, 2008.
- Pimentel, E. D. C. G., Queiroz, S. A. D., Carvalheiro, R., and Fries, L. A.: Use of ridge regression for the prediction of early growth performance in crossbred calves, *Genet. Molec. Biol.*, 30, 536–544, <https://doi.org/10.1590/S1415-47572007000400006>, 2007.
- Sam, I., Ekpo, J., Ukpanah, U., Eyoh, G., and Warri, M.: Relationship between linear body measurement and live body weight in West African Dwarf Goats in Obio Akpa, *J. Biol. Agricult. Healthcare*, 6, 118–124, 2016.
- Sangun, L., Cankaya, S., Kayaalp, G. T., and Akar, M.: Use of factor analysis scores in multiple regression models for estimation of body weight from some body measurements in Lizardfish, *J. Anim. Veterin. Adv.*, 8, 47–50, 2009.
- SAS: SAS/STAT, SAS Institute Incorporation, Cary, NC, USA, 2014.
- Schwarz, G.: Estimating the dimensions of a model, *Ann. Statist.*, 6, 461–464, <https://doi.org/10.1214/aos/1176344136>, 1978.
- Shen, X. and Ye, J.: Adaptive Model Selection, *J. Am. Stat. Assoc.*, 97, 210–221, <https://doi.org/10.1198/016214502753479356>, 2002.
- Tibshirani, R.: Regression shrinkage and selection via the Lasso, *J. Roy. Stat. Soc. Ser. B*, 58, 267–288, 1996.
- Topal, M., Eydurán, E., Yağanoğlu, A. M., Sönmez, A. Y., and Keskin, S.: Çoklu doğrusal bağlantı durumunda ridge ve temel bileşenler regresyon analiz yöntemlerinin kullanımı Atatürk Üniversitesi Ziraat Fakültesi Dergisi, 41, 53–57, 2010.
- Tsegaye, D., Belay, B., and Aynalem Haile, A.: Linear body measurements as predictor of body weight in Hararghe Highland goats under farmers environment Ethiopia, *Global Veterin.*, 11, 649–656, <https://doi.org/10.5829/idosi.gv.2013.11.5.76135>, 2013.
- TUIK: http://www.tuik.gov.tr/VeriBilgi.do?tb_id=46andust_id=13, last access: January 2017.
- Wang, S., Nam, B., Rosset, S., and Zhu, J.: Random Lasso, *Ann. Appl. Stat.*, 5, 468–485, <https://doi.org/10.1214/10-AOAS377>, 2011.
- Weisberg, S.: Applied Linear Regression, 3rd Edn., John Wiley and Sons Publishers, New York, 2005.
- Whittingham, M. J., Stephens, P. A., Bradbury, R. B., and Freckleton, R. P.: Why do we still use stepwise modelling in ecology and behaviour?, *J. Anim. Ecol.*, 75, 1182–1189, <https://doi.org/10.1111/j.1365-2656.2006.01141.x>, 2006.
- Yakubu, A.: Fixing collinearity instability in the estimation of body weight from morpho-biometrical traits of West African dwarf goats, *Trakia J. Sci.*, 7, 61–66, 2009.
- Yakubu, A.: Fixing multicollinearity instability in the prediction of body weight from morphometric traits of White Fulani cows, *J. Centr. Eur. Agricult.*, 11, 487–492, 2010.
- Yılmaz, O., Cemal, İ., and Karaca, O.: Estimation of mature live weight using some body measurements in Karya sheep, *Trop. Anim. Health Product.*, 45, 397–403, <https://doi.org/10.1007/s11250-012-0229-7>, 2013.
- Zou, H.: The adaptive lasso and its oracle properties, *J. Am. Stat. Assoc.*, 101, 1418–1429, <https://doi.org/10.1198/016214506000000735>, 2006.
- Zou, H. and Hastie, T.: Regularization and variable selection via the elastic net, *J. Roy. Stat. Soc. B*, 67, 301–320, <https://doi.org/10.1111/j.1467-9868.2005.00503.x>, 2005.