The Journal of Animal & Plant Sciences, 29(6): 2019, Page: 1541-1548
ISSN: 1018-7081

PREDICTING LIVE BODY WEIGHT OF HARNAI SHEEP THROUGH PENALIZED REGRESSION MODELS

F. Iqbal[1*], M. Ali[2], Z. E. Huma[3] and A. Raziq[1]

[1]Department of Statistics, University of Balochistan, Quetta – Pakistan.
[2] Center for Advanced Studies in Vaccinology and Biotechnology, University of Balochistan, Quetta – Pakistan.
[3]Department of Zoology, Sardar Bahadur Khan Women's University, Quetta – Pakistan.
[*]Corresponding author: farhatiqb@gmail.com

ABSTRACT

For data with multicollinearity, penalized regression models i.e. Ridge regression (RR), least absolute selection and shrinkage operator (LASSO), elastic net (ENet) and adaptive LASSO (ALASSO) methods are popular alternatives to classical linear regression. The aim of this study was to comparatively examine performances of these models in order to predict the live body weight (BW) from various biometric body measurements and to select important variables in order to reduce model complexity. The data on body weight, withers height, body length, chest girth, paunch circumference, face length, length between ears, ear length, fat tail width and length were collected from 757 (247 male and 510 female) indigenous Harnai sheep of Pakistan. The present data were randomly partitioned into training and testing data sets and the hyperparameters of the penalized regression models were tuned using cross-validation. The performance of the studied models on both data sets were evaluated using the root mean squared error (RMSE), adjusted coefficient of determination (R^2_{Adj}) and Schwarz Bayesian information criterion (BIC) as goodness of fit criteria. The results revealed that all penalized regression methods provided accurate fit to the data. The ENet and ALASSO models were found to predict (on training data set) and (on testing data set) the BW with the highest accuracy for both male and female Harnai sheep.

Key words: penalized regression, body weight, sheep, ridge regression, LASSO, elastic net.

INTRODUCTION

In sheep breeding, body weight and knowledge of its association with several body measurements may provide researchers to achieve some useful clues in deciding suitable feed amounts, medication doses, slaughtering time and proper marketing. This knowledge also allows one to estimate genetic correlation between body weight and other morphological measurements which are generally considered as indirect selection criteria for improving better selection strategies relating to growth, breeding and meat production of sheep for breeding purposes (Abbasi and Ghafouri-Kesbi, 2011).

In literature, many studies have been conducted to model and predict the body weight of small ruminants i.e. sheep and goats (Eyduran *et al*., 2017 and references therein). Multiple linear regression (MLR) is a simple method widely used by researchers for revealing the relationship between body weight (dependent variable) and different body measurements (predictors) of animals. The MLR is then used to predict the dependent variable. This classical statistical method suffers from the problem of multicollinearity which occurs when some of the predictors are strongly correlated to each other. One of the consequences of multicollinearity is large standard errors of regression coefficients. Thus, any inference based on the fitted model might not be accurate (Yakubu, 2010; Dormann *et al*., 2013). To circumvent the problem of multicollinearity, use of factor and principal component scores in MLR may be adopted (Eyduran *et al*., 2013). However, various penalized regression methods have been proposed to properly handle this problem. These among others, include the Ridge regression (Hoerl and Kennard, 1970), least absolute shrinkage and selection operator (LASSO) method (Tibshirani, 1996), Elastic net (ENet) method (Cho *et al*., 2009) and adaptive LASSO (ALASSO) method (Zou, 2006).

Malau-Aduli *et al.* (2004) performed Ridge regression and principal component analyses with the objective to predict the body weight of Japanese black cattle from various body measurements. Yakubu (2009) used stepwise regression to avoid the multicollinearity problem in predicting body weight of West African dwarf goats. Tsegaye *et al*. (2013) studied the relationship between body weight and body measurements in Hararghe goats of Ethiopia using multiple linear regression. Khan *et al*. (2014) used factor scores in MLR to find the relationship between body weight and morphological traits in Harnai sheep of Pakistan. Akkol (2018) employed Ridge, LASSO and ALASSO methods for predicting the live body weight of Hair goats of Turkey to remove the multicollinearity problem. With a choice of many alternative techniques, comparative evaluation of the performance of various methods is thus indispensable to select the optimal model and identify important variables.

To the best of our knowledge, these methods and approaches have not been used in modelling the body weight of Harnai sheep. This study was undertaken to predict the body weight of both male and female Harnai sheep of Pakistan from several body measurements through some penalized regression models i.e. Ridge, LASSO, ENet and ALASSO. Multiple linear regression was performed initially and the problem of multicollinearity was detected. Then, different penalized regression models viz. Ridge, LASSO, ENet and ALASSO were employed for modelling the body weight and selecting the important predictors in the presence of multicollinearity. It is worth mentioning that the present study used a slightly different approach for modelling and predicting the body weight. The evaluated data were randomly partitioned into training and testing datasets. The training dataset was used to find the best fitted model and the testing dataset was used to measure the predictive accuracy of models. This approach not only avoids the problem of overfitting but helps to choose the best model for prediction. Models were compared and evaluated using goodness-of-fit measures such as root mean squared error, adjusted coefficient of determination and Schwarz Bayesian information criterion.

MATERIALS AND METHODS

The data consist of records of 757 Harani lambs (247 male and 510 female lambs) from 0 pair of permanent incisors (PPI) to 4PPI collected from Loralai, Ziarat, Pishin and Quetta districts of Balochistan, Pakistan. The variables taken in the study were the live body weight (BW) in Kg, and biometric body measures i.e. withers height (WH), body length (BL), chest girth (CG), paunch circumference (PC), face length (FL), length between ears (LBE), ear length (EL), fat tail width (TW) and fat tail length (TL) measured in in cm. Body weight of sheep were measured using a digital scale whereas other body measurements were measured with a measuring tape.

Consider the following multiple linear regression model

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

where $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ is a vector of observed body weights (dependent variables), $\mathbf{1}_n$ is a column vector of n ones ($i = 1,2, \dots, n$), μ is called the intercept term, $\mathbf{X}$ is an $n \times p$ matrix of body measurements (predictors), $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ is the vector of regression coefficients and $\boldsymbol{\epsilon}$ is a vector of residuals with mean $\mathbf{0}$ and variance $\mathbf{I}\sigma_\epsilon^2$.

In multiple linear regression (MLR), the parameter vector $\boldsymbol{\beta}$ is estimated by minimizing the sum of squared residuals (RSS) by the ordinary least squares (OLS) method. Mathematically,

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2, \qquad (2)$$

where $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j\right)^2$ is the quadratic loss function also called RSS.

In case of many correlated variables, the MLR suffers from the problem of multicollinearity. Ridge regression (Hoerl and Kennard, 1970) shrink the regression coefficient using l_2-norm (quadratic loss) to overcomes the problem of multicollinearity. The Ridge regression use penalized least square method to solve the regression problem.

$$\widehat{\boldsymbol{\beta}}_{Ridge} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{XB}\|^2 + \lambda\|\boldsymbol{\beta}\|^2 \qquad (3)$$

where $\lambda \geq 0$ is called the tuning (shrinkage or tuning) parameter that regulates the strength of penalty and $\|\boldsymbol{\beta}\|^2 = \sum_{j=1}^{p}\beta_j^2$ is the quadratic (l_2-norm) penalty on $\boldsymbol{\beta}$. Large value of λ means greater amount of shrinkage. The Ridge regression keeps all the predictors in the model without doing any variable selection (Zou and Hastie, 2005).

The least absolute shrinkage and selection operator (LASSO) method automatically select variables using l_1-norm (Tibshirani, 1996). In LASSO regression, the following equation is used to estimate the regression coefficients

$$\widehat{\boldsymbol{\beta}}_{LASSO} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|, \qquad (4)$$

where $\|\boldsymbol{\beta}\| = \sum_{j=1}^{p}|\beta_j|$ is the l_1-norm penalty on $\boldsymbol{\beta}$. For a suitable value of λ, the LASSO shrinks some parameters to zero.

The Elastic net (ENet) is another extension of LASSO suitable for a situation when predictors are strongly correlated (Cho et al., 2009). The ENet estimator can be obtained by

$$\widehat{\boldsymbol{\beta}}_{ENet} = \left(1 + \frac{\lambda_2}{n}\right)\left\{\arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda_1\|\boldsymbol{\beta}\| + \lambda_2\|\boldsymbol{\beta}\|^2\right\} \qquad (5)$$

where λ_1 and λ_2 are LASSO and Ridge regression penalties, respectively.

Both LASSO and ENet are popular variable selection methods; however, they lack the oracle property. As defined by Fan and Li (2001), the subset of true parameters with zero coefficients are estimated as exactly zero with almost sure probability in an oracle procedure. This efficient variable selection method produces asymptotically unbiased and normally distributed estimates of the nonzero coefficients (Zou, 2006).

The adaptive LASSO (ALASSO) can be used to avoid the problem of lack of oracle property of the LASSO method. The ALASSO is a modification of the original LASSO penalty. This method adds weights for each parameter to the penalty term (Zou, 2006). The ALASSO estimator is given by

$$\hat{\boldsymbol{\beta}}_{ALASSO} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda, \sum_{j=1}^{p} \hat{\omega}_j |\beta_j|, \quad (6)$$

where $\hat{\omega}_j = 1/|\hat{\beta}_j^{ini}|^{\gamma}$ is a known weight, $\gamma > 0$ is a constant and $\hat{\beta}_j^{ini}$ is the initial consistent estimator obtained from OLS or ridge regression when multicollinearity problem is present (Ogutu *et al.*, 2012). These data-defined weights, $\hat{\omega}_j$, control the shrinkage of the zero coefficients more than the non-zero coefficients.

Different regression methods mentioned earlier were used to reveal the relationship between BW and other morphological traits. The data were randomly partitioned into two parts; the training dataset (80 %) and testing dataset (20 %). Fivefold cross validation was used to obtain optimal values of tuning parameters. The Ridge, ENet, LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) methods were fitted using a training dataset. Then, the predictive performances of these methods are evaluated on a testing dataset. Commonly-used goodness-of-fit measures were estimated in this study to compare the results in the model selection. These include the root mean squared error (RMSE), the adjusted coefficient of determination (R^2_{Adj}) and Schwarz Bayesian information criterion (BIC). The predictive performance of these models was then evaluated using the testing dataset.

The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{SSE}{n}},$$

where $SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ is the sum of square errors, n is the total number of observations and $\hat{Y}_i$ is the estimated/predicted body weight of i-th animal. The adjusted coefficient of determination is defined as

$$R^2_{\text{Adj}} = 1 - (1 - R^2)\frac{n-1}{n-p-1},$$

where R^2 is the coefficient of determination and p is the number of predictors in the model.

The BIC Schwarz is

$$\text{BIC} = n\ln\left(\frac{SSE}{n}\right) + p\ln(n).$$

The model with the highest R^2_{Adj} and lowest RMSE and BIC values is considered to be the best among competing models. Besides, standard deviation ratio (SD ratio), which is defined as the ratio of standard deviation of body weight to standard deviation of error term is also calculated. The SD ratio of below 0.40 indicates a good fit. The R program (R Development Core Team 2014) was used for conducting all Statistical analysis.

RESULTS AND DISCUSSION

In this study, there were 247 male (32.63 %) and 510 female (67.37 %) sheep. Descriptive statistics of body weight (BW) and other body measurements such as WH, BL, CG, PC, FL, LBE, EL, FW, and FL and independent sample t-test for all variables in both genders are provided in Table 1. Except for BW, BL and CG, significant differences ($P < 0.05$) were observed between the genders for all biometric measurements.

The results of the study found a significant difference between genders in terms of body measurements. All measurement for male sheep were found significantly larger than the female apart from withers height and paunch circumference. Khan et al. (2014) reported similar findings for most of the variables though FL and LBE were not found significantly different in genders. WH and CG were also found significantly different by Akbas and Saatci (2016).

Figure 1 shows the Pearson's coefficient of correlation between all variables for both genders. Significant positive correlations (correlation coefficient greater than 0.5) were found between most of the body measurements for both male and female sheep (P<0.05). The BW was found strongly correlated with all predictors for both genders (P<0.01). The correlation coefficients ranged between 0.52 and 0.94. Similar findings were reported by other studies. For example, Eyduran et al. (2013) reported that CG and WH are strongly associated with BW for commercial goats of Pakistan; Tsegaye et al. (2013) found BW to be positively correlated with BL, HG, and EL for Hararghe goats of Ethopia; Das et al. (2015) reported strong correlation of BW with BL and HG for Jamunapari goats of India; Sam et al. (2016) also found BW to be strongly correlated with WH and BL for Western African Dwarf goats of Nigeria and high correlation between BW with WH, CG and BL were observed by Akkol (2018) for Hair goats of Turkey. Besides, strong positive correlations (correlation coefficient greater than 0.8) were found between biometric measurements as predictors handled in the present study.

Table 1. Descriptive statistics of live body weight and biometric body measurements for male and female Harnai sheep.

Variable	Male ($n = 247$)		Female ($n = 510$)	
	Mean $\pm$ SE	CV(%)	Mean $\pm$ SE	CV(%)
BW	33.07 $\pm$ 0.94	44.88	32.68 $\pm$ 0.65	44.94
WH	47.65 $\pm$ 0.59*	19.46	45.12 $\pm$ 0.33	16.41
BL	42.25 $\pm$ 0.48	17.72	42.99 $\pm$ 0.31	16.36
CG	46.54 $\pm$ 0.50	16.98	46.90 $\pm$ 0.39	18.69
PC	49.52 $\pm$ 0.57*	18.14	44.26 $\pm$ 0.35	17.64
FL	16.30 $\pm$ 0.32*	30.89	18.50 $\pm$ 0.28	33.77
LBE	12.90 $\pm$ 0.15*	17.84	14.33 $\pm$ 0.13	19.88
EL	4.96 $\pm$ 0.12*	39.43	6.82 $\pm$ 0.13	42.69
FW	16.70 $\pm$ 0.19*	18.20	19.40 $\pm$ 0.20	23.45
FL	13.53 $\pm$ 0.20*	22.73	15.63 $\pm$ 0.20	29.55

* Difference from female is statistically significant at 1%. Body weigh in kg, all other body measurements in cm.

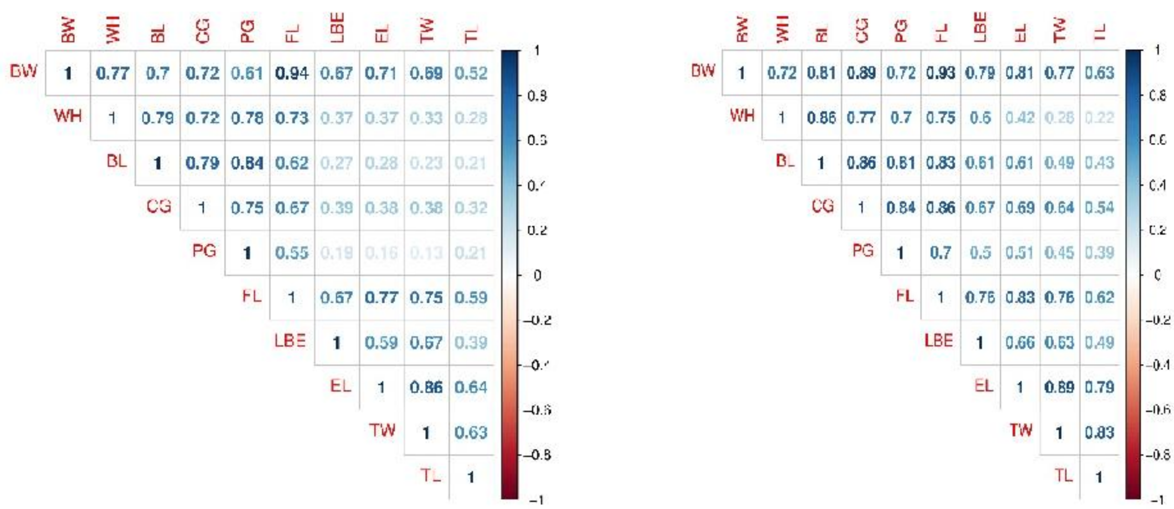

Figure 1. Pearson correlation coefficient between body weight and biometric body measurements of male (left) and female (right) Harani sheep.

The results of multiple linear regression are given in Table 2. The regression coefficients along with their standard errors and variance inflation factor (VIF) are shown for both genders. All predictors included in the model, explained 92.3 % of the variation in BW for males and 92.0 % for females. The VIF values of more than 5 observed for most of the variables may be considered a cause of concern (Sheater, 2009) whereas a value of more than 10 is an indication of severe multicollinearity (Kutner et al., 2004). These findings along with high correlation coefficient values found between biometric measurements as predictors may indicate the presence of multicollinearity. Multicollinearity results in higher standard errors of regression coefficients and hence the accuracy and robustness of fitted models are difficult to be found (Sangun et al., 2009; Yakubu, 2010; Eyduran et al., 2013), hence the data need further examination. We adopted alternative approaches to circumvent this problem of multiple linear regression.

Table 2. OLS estimates of regression coefficients in MLR and VIF.

Variable	Male		Female	
	Coefficients	VIF	Coefficients	VIF
Intercept	-35.12 ± 3.18*	0	-39.31 ± 2.29*	0
Withers height	0.20 ± 0.06*	4.31	0.23 ± 0.07*	6.75
Body length	0.33 ± 0.08	5.08	-0.12 ± 0.07	7.13
Chest girth	0.13 ± 0.06	3.41	0.54 ± 0.06*	7.83
Paunch circumference	-0.06 ± 0.06	4.78	0.01 ± 0.05	3.90
Face length	1.70 ± 0.15*	8.46	0.63 ± 0.10*	11.21
Length between ears	0.79 ± 0.17*	2.16	0.74 ± 0.11*	2.59
Ear length	0.53 ± 0.13	5.07	0.22 ± 0.17	7.23
Fat tail width	0.12 ± 0.20	5.26	0.75 ± 0.12*	8.74
Fat tail length	-0.04 ± 0.12	1.91	-0.94 ± 0.08	3.60
RMSE	4.045		4.147	
R^2	0.925		0.921	
R^2_{Adj}	0.923		0.920	

* P<0.05; VIF: variance inflation factor; RMSE: root mean squared error;
R^2: coefficient of determination; R^2_{Adj}: adjusted coefficient of determination

This study utilized Ridge, ENet, LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) regressions as alternative methods to MLR in the presence of multicollinearity. The results of applying Ridge, LASSO, ENet and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) regressions are reported in Table 3 for male sheep and in Table 4 for female sheep. For both genders, Ridge regression included all predictors, whereas ENet, LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) regression models consisted of reduced number of predictors. For male sheep, two predictors, i.e. PC and FL were not included by LASSO, ENet and ALASSO ($\gamma = 1$) whereas these variables besides FW were not included by ALASSO ($\gamma = 0.5$). For female sheep, three predictors BL, PC and FL by LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) were not included in the model whereas ENet excluded two predictors, BL and FL.

Next, we compare the performance of Ridge, LASSO, ENet and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) models by goodness-of-fit measures i.e. RMSE, R^2_{Adj} and BIC. Table 5 displays the results of fitting these models to the training dataset. All models handled in the present study were found to fit the training data well with very high adjusted R^2 values (90.74 % - 91.51 %). The adjusted R^2 values of Ridge regression (which included all predictors in the model) for males and females were 90.74 % and 91.25 %, respectively were found the lowest. The ENet, LASSO and ALASSO methods with $\gamma = 0.5$ and $\gamma = 1$ were used not only to remove the problem of multicollinearity but to reduce the model complexity by selecting important predictors. For male sheep data, the values of adjusted coefficient of determination were found 91.49 %, 91.51 % and 90.93 % for LASSO, ENet and ALASSO ($\gamma = 1$) models, respectively. Each of these models consisted of 7 variables whereas the ALASSO ($\gamma = 0.5$) model which selected 6 variables had the adjusted R^2 value of 90.98 %. For female data, ENet selected 7 variables with highest adjusted coefficient of determination (91.58 %). LASSO and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) models which consisted of 6 predictors also indicated similar performances in adjusted R^2 values. The SD ratios of all models were found below 0.40 for both training and testing datasets which further confirm the better fit of these models.

The adjusted coefficient of determination value of 91.51 % (for male Harnai sheep) for ENet found in the present study was higher than that reported by Khan et al. (2014) who estimated 87.60 % for same males using multiple linear regression with scores of explanatory factor analysis. However, the present R^2_{Adj} value (91.58 %) for female Harnai sheep was slightly lower than their reported value (91.90 %). The highest R^2 value of 80.30 % reported by Sam et al. (2016) for African Dwarf goats, and the adjusted R^2 values of 90.48 % and 76.68 % reported by Akkol (2018) for male and female Hair goats, respectively, were also found lower than adjusted R^2 values given in the present study.

Table 3. Estimates of Ridge, LASSO, Elastic Net and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) in MLR for male sheep.

	Ridge	LASSO	Elastic Net	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)
Withers height	0.259	0.170	0.194	0.146	0.151
Body length	0.290	0.278	0.285	0.316	0.316
Chest girth	0.153	0.108	0.117	0.039	0.052
Paunch circumference	0.045	–	–	–	–
Face length	1.099	1.729	1.606	1.825	1.810
Length between ears	0.921	0.756	0.789	0.814	0.816
Ear length	0.894	0.486	0.576	0.539	0.549
Fat tail width	0.384	0.083	0.142	–	0.001
Fat tail length	0.065	–	–	–	–

Table 4. Estimates of Ridge, LASSO, Elastic Net and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) in MLR for female sheep.

	Ridge	LASSO	Elastic Net	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)
Withers height	0.232	0.192	0.204	0.210	0.205
Body length	0.102	–	–	–	–
Chest girth	0.392	0.543	0.533	0.514	0.518
Paunch circumference	0.083	–	0.005	–	–
Face length	0.498	0.679	0.653	0.692	0.695
Length between ears	0.789	0.744	0.747	0.757	0.758
Ear length	0.472	0.146	0.200	0.180	0.176
Fat tail width	0.599	0.644	0.638	0.650	0.646
Fat tail length	0.023	–	–	–	–

Table 5. Evaluating Ridge, LASSO, Elastic Net and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) models on training dataset for male and female sheep.

	Male					Female				
	Ridge	LASSO	Elastic Net	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)	Ridge	LASSO	Elastic Net	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)
p	9	7	7	6	7	9	6	7	6	6
RMSE	4.3180	4.1420	4.1540	4.2900	4.2900	4.2760	4.2050	4.2020	4.2060	4.2050
R^2	0.9116	0.9175	0.9177	0.9121	0.9116	0.9146	0.9170	0.9171	0.9169	0.9169
R^2_{Adj}	0.9074	0.9149	0.9151	0.9098	0.9093	0.9125	0.9157	0.9158	0.9156	0.9156
BIC	629.80	597.36	598.52	606.09	606.05	1169.4	1138.8	1138.2	1138.9	1138.8
SDratio	0.2757	0.2708	0.2710	0.2770	0.2768	0.2798	0.2746	0.2750	0.2748	0.2750

p: number of variables; RMSE: root mean squared error; R^2: coefficient of determination; R^2_{Adj}: adjusted coefficient of determination; BIC: Schwarz Bayesian information criterion

Table 6. Evaluating Ridge, LASSO, Elastic Net and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) models on testing data set for male and female sheep.

	Male					Female				
	Ridge	LASSO	Elastic Net	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)	Ridge	LASSO	Elastic Net	ALASSO ($\gamma = 0.5$)	ALASSO ($\gamma = 1$)
RMSE	4.7110	4.1680	4.2920	4.2690	4.1270	4.0380	4.0210	4.0200	4.0270	4.0290
R^2	0.9178	0.9355	0.9320	0.9342	0.9379	0.9300	0.9308	0.9309	0.9307	0.9307
R^2_{Adj}	0.9139	0.9335	0.9299	0.9325	0.9363	0.9283	0.9297	0.9298	0.9296	0.9296
BIC	664.50	599.85	611.53	604.16	590.65	1125.4	1104.4	1104.2	1105.6	1105.9
SDratio	0.3340	0.3126	0.3169	0.3213	0.3216	0.3097	0.3133	0.3125	0.3146	0.3144

RMSE: root mean squared error; R^2: coefficient of determination; R^2_{Adj}: adjusted coefficient of determination; BIC: Schwarz Bayesian information criterion

When considering other goodness-of-fit measures (RMSE and BIC), the LASSO method had the smallest values for male Harnai sheep. ENet was found to provide the smallest values of RMSE and BIC for female sheep whereas Ridge regression had the highest values of these measure for both genders. In the present study, the LASSO model selected FL, LBE, EL, BL, WH, CG and TW, in order of the highest significant effect, as important predictors for BW of male Harnai sheep. ENet found LBE, FL, FW, CG, and WH, in order of highest significant effect, as important predictors of BW of female Harnai sheep. Previous studies also identified few of these variables as significant predictors for BW of sheep and goats (Eyduran et al., 2013; Khan et al., 2014; Das and Yadav, 2015; Sam et al., 2016; Akkol, 2018).
Generally, researchers are more interested in the predictive performance of the models used in their study. A model that fits the data well is not necessary to provide the best prediction. Hence, we evaluated the predictive performance of all the models considered in the present study on a separate testing dataset. The results of different evaluation measures for both male and female data were reported in Table 6. For males, the RMSE value of 4.1270 was found the least for ALASSO ($\gamma = 1$) followed by the LASSO regression. The BIC values for these models were also found smaller than other models used in the present study. The R^2_{Adj} values varied between 91.39 % (Ridge) and 93.63 % (ALASSO with $\gamma = 1$). For females, ENet was found to provide the best results with minimum RMSE (4.0200), BIC (1004.2) and maximum R^2_{Adj} (92.98 %). Hence, we concluded that the ALASSO ($\gamma = 1$) and ENet are the best predictive models for predicting the BW of male and female Harnai sheep, respectively. The ALASSO model was also found to perform better than Ridge and ALASSO models by previous researchers (see for example, Ogutu *et al.*, 2012 and Akkol, 2018).

The present study utilized penalized regression methods such as Ridge, LASSO, ENet and ALASSO ($\gamma = 0.5$ and $\gamma = 1$) in modelling the body weight based on morphological traits of indigenous Harnai sheep breed of Pakistan in the presence of multicollinearity. The results of evaluation measures, when models were fitted using training dataset, revealed that all models fit the data well with ENet and ALASSO providing better fits. The predictive performance of models, evaluated on testing data, also confirmed the superior predictive ability of these models. In conclusion, we found that in the presence of multicollinearity, alternative methods may be employed to obtain better fit and accuracy. However, care must be taken when using these methods as sometimes they ignore nonsignificant variables that may, nevertheless, be interesting or important. The results of the present study may help practitioners and researchers to choose a suitable approach for better modelling and accurate prediction of live weights of animals when data have problem of multicollinearity.

REFERENCES

Abbasi, M.A. and F. Ghafouri-Kesbi (2011). Genetic (Co)variance Components for Body Weight and Body Measurements in Makooei Sheep. Asian-Aust. J. Anim. Sci. 24: 739 – 743.

Akbas, A.A. and M. Saatci (2016). Growth, slaughter, and carcass characteristics of Honamlı, Hair, and Honamlı×Hair (F1) male goat kids bred under extensive conditions. Turk. J. Vet. Anim. Sci. 40: 459–467. DOI: https://doi.org/10.3906/vet-1511-5.

Akkol, S. (2018). The prediction of live weight of hair goats through penalized regression methods: LASSO and adaptive LASSO. Arch. Anim. Breed. 61: 451-458.

Cho S., H. Kim, S. Oh, K. Kim, and T. Park (2009). Elastic-net regularization for genomewide association studies of rheumatoid arthritis. BMC proceedings. 3(suppl 7): s25.

Das, S.K., and B.P.S. Yadav (2015). Productive performance of Jamunapary goat under semi intensive system of management. Indian J. Anim. Res. 49: 254–257. DOI: https://doi.org/10.5958/0976-555.2015.00114.4.

Dormann, C.F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J.R.G. Marquéz, B. Gruber, B. Lafourcade, P.J. Leitão, T. Münkemüller, C. McClean, P.E. Osborne, B. Reineking, B. Schröder, A.K. Skidmore, D. Zurell, and S. Lautenbach (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, Ecography. 36: 27–46. DOI: https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Eyduran, E., A. Waheed, M.M. Tariq, F. Iqbal, and S. Ahmed (2013). Prediction of live body weight from morphological characteristics of commercial goat in Pakistan using factor and principal component scores in multiple linear regression. The J. Anim. Plan. Sci. 23(6): 1532-1540.

Eyduran, E., D. Zaborski, A. Waheed, S. Celik, K. Karadas, and W. Grzesiak (2017). Comparison of the Predictive Capabilities of Several Data Mining Algorithms and Multiple Linear Regression in the Prediction of Body Weight by Means of Body Measurements in the Indigenous Beetal Goat of Pakistan. Pakistan J. Zool. 49(1): 257-265.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. As. 96: 1348-1360.

Hoerl, A.E. and R.W. Kennard (1970). Ridge regression: biased estimation for nonorthogonal problems. Technomterics. 12: 55-67.

Khan, M.A., M.M. Tariq, E. Eyduran, A. Tatlıyer, M. Rafeeq, F. Abbas, N. Rashid, M.A. Awan, and K. Javed (2014). Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. The J. Anim. Plant Sci. 24: 120– 126.

Kutner, M. H., C.J. Nachtsheim, and J. Neter (2004). Applied Linear Regression Models (4th ed.). McGraw-Hill Irwin.

Malau-Aduli, A.E.O., M.A. Aziz, T. Kojina, T. Niibayashi, K. Oshima, and M. Komatsu (2004). 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.

R Core Team R. 2014. A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Ogutu, J.O., T. Schulz-Streeck, and H. Piepho (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions, BMC Proc. 6: 10. DOI: https://doi.org/10.1186/1753-6561-6-S2-S10.

Sam, I., J. Ekpo, U. Ukpanah, G. Eyoh, and M. Warrie (2016). Relationship between linear body measurement and live body weight in West African Dwarf Goats in Obio Akpa. J. Biol. Agricult. Healthcare. 6: 118–124.

Sangun, L., S. Cankaya, G.T. Kayaalp, and M. Akar (2009). 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.

Sheater, S. (2009). A Modern Approach to Regression with R. New York, NY: Springer.

Tsegaye, D., B. Belay, and A. Haile (2013). Linear body measurements as predictor of body weight in Hararghe Highland goats under farmers environment Ethiopia. Global Veterin. 11: 649–656. DOI: https://doi.org/10.5829/idosi.gv.2013.11.5.7615.

Tibshirani R. (1996). Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B. 58: 267-288.

Yakubu, A. (2009). Fixing collinearity instability in the estimation of body weight from morpho-biometrical traits of West African dwarf goats. Trakia J. Sci. 7: 61–66.

Yakubu, A. (2010). Fixing multicollinearity instability in the prediction of body weight from morphometric traits of White Fulani cows. J. Centr. Eur. Agricult. 11: 487–492.

Zou, H., and T. Hastie (2005). Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B. 67: 301–320. DOI: https://doi.org/10.1111/j.1467-68.2005.00503.x.

Zou, H. (2006). The adaptive lasso and its oracle properties. J. Am. Stat. As. 101: 1418-1429.