



Comparison of Least Squares, Ridge Regression and Principal Component Approaches in the Presence of Multicollinearity in Regression Analysis

Soner Çankaya^{1,a,*}, Samet Eker^{2,b}, Samet Hasan Abacı^{3,c}

¹Department of Sports Management, Yaşar Doğu Faculty of Sport Sciences, Ondokuz Mayıs University 55139 Samsun, Turkey

²Ordu Provincial Directorate of Agriculture and Forestry, 52200 Ordu, Turkey

³Department of Animal Science, Faculty of Agriculture, Ondokuz Mayıs University, 55139 Samsun, Turkey

*Corresponding author

ARTICLE INFO

Research Article

Received : 04/03/2019

Accepted : 22/05/2019

Keywords:

Least Squares

Ridge Regression

Principal Component Regression

Multicollinearity

Body Measurements

ABSTRACT

The aim of this study was to compare estimation methods: least squares method (LS), ridge regression (RR), Principal component regression (PCR) to estimate the parameters of multiple regression model in situations when the underlying assumptions of least squares estimation are untenable because of multicollinearity. For this aim, the effect of some body measurements on body weights (height at withers and rumps, body length, chest width, chest girth and chest depth, front, middle and hind rump width) obtained from totally 85 Karayaka lambs at weaning period raised at Research Farm of Ondokuz Mayıs University was examined. Mean square error, R^2 value and significance of parameters were used to evaluate estimator performance. The multicollinearity, between front and middle rump width which were used to estimate live weight, was eliminated by using RR and PCR. Although research findings showed that RR method had the smallest MSE and the highest R^2 value, the estimates of PCR were determined to be more consistent when the importance tests of parameters were taken into account. The results showed that principal component regression approach should be used to estimate the live weight of Karayaka lambs at weaning period.

^a scankaya@omu.edu.tr

^{ID} <https://orcid.org/0000-0001-8056-1892>

^c shabaci37@gmail.com

^{ID} <https://orcid.org/0000-0002-1341-4056>

^b samet.eker@tarim.gov.tr

^{ID} <https://orcid.org/0000-0002-2540-0516>



This work is licensed under Creative Commons Attribution 4.0 International License

Introduction

Regression analysis is the most commonly used statistical application used to predict the quantitative correlation between a dependent variable (Y_i) and one or more independent variables (X_i). General purposes for using regression analysis can be listed as the prediction of i) future measurement values from values that can be measured early ii) the values of a feature that is hard to measure from a data set of a feature that is easy to measure iii) high-cost measurement values from low-cost measurement values (Huber and Dutter, 1974).

The most commonly used method in regression analysis is the Least Squares (LS) method with the condition of some required assumptions such as normality. This method is based on the idea of reducing the sum of the coefficients of the difference between the Y values given by the equation (theoretical) and X values given by the measurements (actual). The reliability of the resulting model depends on realization of the assumptions of LS

method. In case of significant multicollinearity between the examined independent variables, coefficients of the regression parameters predicted with LS method can cause the results to be misinterpreted.

Ridge regression is a biased regression method developed to eliminate the negative effects that will occur on parameter predictions in case of multicollinearity (occurrence of a dependent interaction between independent variables) in regression analysis. Principal Components Regression is a regression method which explains original variables which have correlation between them with the help of fewer and new variables that are linear compounds of these variables.

Prediction of results obtained in fields such as agriculture, socio-economy, medicine and biology with LS method without making the necessary assumptions can cause incorrect results. In this case, validities obtained with LS method should be suspected (Alpar, 1997). However,

in studies conducted within this context, the presence of multicollinearity, which is a significant criterion in the field of statistics, and solution techniques are not given the necessary importance. In addition, it is not clear which method is superior to other methods in case of multicollinearity in agriculture and especially in husbandry.

The aim of this study was to examine and compare RR and PCR methods as an alternative to LS method, which is the most known and used method in case of multicollinearity, by considering the necessary assumptions, to find out the method which gives correct and reliable results in parameter estimation and to interpret the results.

Material and Method

Material

In this study, body measurements (height at withers (HW), height at rumps (RH), body length (BL), chest depth (CD), chest width (CW), chest girth (CG), front rump width (FR), middle rump width (MR), and hind rump width (HR)) and live weight (LW) of a total of 85 Karayaka lambs at weaning period raised at Research Farm of Ondokuz Mayıs University were used. In the multiple regression model, body measurements taken at weaning period (X variable cluster) form the independent variable group, while measurements of live weight (Y variable) form the dependent variable. SPSS and NCSS programs were used for statistical analysis.

Method

Regression analysis is one of the most commonly used methods to explain the association between one dependent and multiple independent variables. General expression of multiple regression model in Matrix form is given below (Alpar, 2011).

$$Y=X\beta+\varepsilon$$

In the equation,

- Y : n x 1 dimensional dependent variable vector
- X : n x (p+1) dimensional input matrix and the first column of this matrix consists of 1, while the other columns consist of variable values.
- β : (p+1) x 1 dimensional coefficients vector
- ε : n x 1 dimensional error vector

and equation for n observation is shown as follows.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Prediction equation of this equation is defined as follows.

$$\hat{Y} = X\hat{\beta}$$

Here, $\hat{\beta}$ is the (p+1) x 1 dimensional coefficients vector consisting of $b_0, b_1, b_2, \dots, b_p$. In the prediction of this coefficients vector, different methods are utilized based on variables' states of proving the variables.

Least Squares Method

The purpose of this method is to minimize the optimum results (Neter et al. 1990), in other words, the sum of squares of error terms, in case of error terms having a normal distribution and having homogeneous variance and thus to optimize the model.

$$Q_{(EKK)}(b) = \sum_{i=1}^n e_i^2$$

In multiple regression analysis, the following equation is utilized in the prediction of coefficients vector with LS method (Alpar, 2011).

$$\hat{\beta} = [X'X]^{-1}X'Y$$

General expression of regression model is as follows in multiplicative form;

$$Y_i = \beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2} X_{i3}^{\beta_3} \dots X_{ip}^{\beta_p} e_i \quad i=1, 2, 3, \dots, n$$

In the equation;

- Y_i : Dependent variable
- β_j : Parameters; $j= 1, 2, 3, \dots, p$
- $X_{i1}, X_{i2}, \dots, X_{ip}$: Independent variables
- e_i : Error values

When dependent variable data versus independent variable data are shown on graph, it may not always look like a linear line. That is, the association between the examined characteristics may look like a curvilinear distribution. In order to linearize this curvilinear state, observation values are exposed to logarithmical transformation in X and Y variables. This way, the regression equation, which is given in multiplicative form, is transformed into the following model (Sangun et al., 2009; Çankaya et al., 2009).

$$\ln Y_i = \ln \beta_0 + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \dots + \beta_p \ln X_{ip} + \ln e_i$$

In this equation, respectively $Y=\ln Y_i$ shows live weight, $\ln X_{i1}, \ln X_{i2}, \dots, \ln X_{ip}$ show independent parameters (height at withers, height at rumps, body length, chest width, chest girth, chest depth, front rump width, middle rump width, and hind rump width) with ($p=1,2,\dots,9$), $\beta_1, \beta_2, \dots, \beta_p$ and $a = \ln \beta_0$ show regression parameters and $\ln e_i$ shows random error (Gunst and Mason,1980; Draper and Smith, 1981; Kleinbaum et al., 1998).

However, in case of multicollinearity between independent variables, since variance increases in LS predictions, the results can remove from the actual values even if predictions are unbiased. For this reason, in case of strong linearity between independent variables in linear regression model, using methods alternative to LS method can decrease the variance and result in more stable results (Albayrak, 2006).

Ridge Regression

In Ridge regression (RR) model, Ridge regression is obtained by adding small k values ($k \geq 0$) to the diagonal factors of $X'X$ matrix in the form of $X'X$ matrix correlation (Hoerl and Kennard 1970).

$$\hat{\beta}(k) = (X'X + kI)^{-1} X'y, \dots k \geq 0$$

Ridge regression method is used in i) showing the instability that occurs in coefficients on chart in the presence of strong multicollinearity, ii) obtaining smaller variance predictions than LS predictions when independent variables are correlated in multi linear regression model, iii) eliminating multicollinearity in independent variables, iv) in decreasing mean squared error (MSE) by changing variance in regression with coefficient bias.

Finding the k parameter value of Ridge regression model depends on eigenvalue. Ridge trace plot is examined or the value of k parameter is found out at which point Ridge regression process becomes stable or is closer to eigenvalue 1.

A great number of researchers have suggested various formulas to find out k value. Among these formulas, Kurtuluş (2001) utilized condition index in finding out k constant based on eigenvalue and obtained the following equation:

$$k \leq \frac{\lambda_{\max} - 100\lambda_{\min}}{99}, k \neq 0$$

By using this equation, the point where k parameter makes VIF value closest to 1 is found (Anderson, 1998; Üçkardeş et al., 2012). As cited by Albayrak (2005) from Anderson (1998), other criteria used in the selection of optimum k value can be listed as k constant approaches which provide the coefficients' suitability to hypothetic expectations, their stability, reasonable size, acceptable error sum of squares and minimum VIFs (VIF values close to 1 together for independent variables).

Principal Component Regression (PCR)

Principal Component Regression (PCR) method is a technique which predicts the coefficients of variables in multiple regression analysis without the need to delete independent variables in case of multicollinearity between independent variables. PCR standard errors are decreased by adding a bias degree to regression predictions (Albayrak, 2006, Hintze, 2007).

For PCR analysis, first of all, after all the variables (dependent and independent) are subtracted from their averages, they are divided by their own standard deviations and standardized. Later, independent variables are transformed to principal components and they are mathematically expressed with the following equation.

$$X'X = PDP' = Z'Z$$

In the equation X'X; D describing PCR model shows correlation matrix for independent variables, P shows the diagonal matrix of X'X eigenvalue and X'X shows eigenvector matrix and Z shows data matrix (Albayrak, 2006; Hintze, 2007; Topal et al., 2010).

As a result of these operations, Z (Z₁, ..., Z_n) variables which express the weighted average of X (X₁, ..., X_n) original independent variables are derived. Since these new variables are principal components, the correlation between principal components is zero. It will be possible to detect multicollinearity for very small eigenvalue. In order to eliminate multicollinearity data, generally compounds with small eigenvalue (Z) that can consist of one or two are removed. When the compound with small eigenvalue is removed from the analysis, multicollinearity problem won't exist when regression is performed to independent variables on dependent variables. Later, the results are turned into X scale to obtain B predictions. It is thought that these predictions will be biased; however, it is expected that the extent of this bias will be higher than compensated with the decrease in variance (Albayrak, 2006; Hintze, 2007; Topal et al., 2010).

Mathematical prediction equation is given in the following equation.

$$\hat{A} = (Z'Z)^{-1}Z'Y = D^{-1}Z'Y$$

This prediction equation is similar to ordinary least squares regression applied on a different independent variables cluster. Two sets of regression coefficients such as A and B are associated with the following formulas.

$$\begin{aligned} \underline{A} &= P' \underline{B} \\ \underline{B} &= P \underline{A} \end{aligned}$$

Removal of a principal component can occur by equalizing the corresponding A element to zero (Albayrak, 2006; Hintze, 2007; Topal et al., 2010).

Results and Discussion

In this study, the descriptive statistics of some body measurements and live weights obtained from Karayaka lambs at weaning period raised at Research Farm of Ondokuz Mayıs University are given in Table 1.

Table 1 Descriptive statistics of the traits examined in Karayaka lambs

Traits	n	Ln (Mean)	Mean of original data	Std. deviation	Variation Coefficient (VC) (%)
Live weight (LW)	85	3.27	27.37	6.659	24.3
Height at withers (WH)	85	3.97	53.29	5.115	9.6
Height at rumps (RH)	85	3.98	53.56	4.809	9.0
Body length (BL)	85	3.92	50.71	5.496	10.8
Chest depth (CD)	85	3.29	26.87	2.853	10.6
Chest width (CW)	85	2.57	13.53	3.365	24.9
Chest girth (CG)	85	4.46	86.99	10.722	12.3
Front rump width (FR)	85	2.55	13.13	2.923	22.3
Middle rump width (MR)	85	2.60	13.85	3.022	21.8
Hind rump width (HR)	85	2.50	12.44	2.238	18.0

Table 2 Correlation coefficients between examined traits and significance test results

Traits	LW	WH	RH	BL	CD	CW	CG	FR	MR
WH	0.63**								
RH	0.60**	0.79**							
BL	0.49**	0.51**	0.51**						
CD	0.65**	0.55**	0.57**	0.43**					
CW	0.26*	0.11-	0.18-	0.37**	0.37**				
CG	0.73*	0.62**	0.62**	0.62**	0.73**	0.43**			
FR	0.30**	0.28*	0.30**	0.43**	0.42**	0.51**	0.58**		
MR	0.33**	0.31**	0.34**	0.42**	0.43**	0.57**	0.56**	0.96**	
HR	0.35**	0.33**	0.60**	0.46**	0.38**	0.38**	0.56**	0.86**	0.86**

LW: Live weight; WH: Height at withers; RH: Height at rumps; BL: Body length; CD: Chest depth; CW: Chest width; CG: Chest girth; FR: Front rump width; MR: Middle rump width; HR: Hind rump width; *: P<0.05; **: P<0.01

Table 3 Regression analysis results according to Least Squares Method

Traits	Coefficients	Std Deviation	Standardized regression coefficient	t-value	P	Tolerance	VIF
Constant (b ₀)	-6.672	1.010	0.159	-6.604	**		
WH	0.517	0.385	0.071	1.342	-	0.314	3.180
RH	0.234	0.391	-0.006	0.598	-	0.341	2.930
BL	-0.034	0.264	0.191	-0.127	-	0.532	1.881
CD	0.523	0.295	-0.018	1.770	-	0.423	2.366
CW	-0.018	0.115	0.545	-0.159	-	0.533	1.875
CG	1.296	0.317	-0.485	4.089	**	0.275	3.642
FR	-0.643	0.343	0.286	-1.873	-	0.071	14.069
MR	0.380	0.370	0.071	1.026	-	0.061	16.476
ASHR	0.119	0.247	0.159	0.480	-	0.216	4.629

** : P<0.01 * : P<0.05 ∴ P>0.05

Table 4 Correlation eigenvalue and the number of conditions

Number	Eigenvalue	Condition index
1	4.99	1.00
2	1.64	3.04
3	0.75	6.61
4	0.58	8.53
5	0.42	11.83
6	0.21	23.28
7	0.20	24.00
8	0.14	34.36
9	0.04	142.2

With Kolmogorow – Smirnov normality assumption test applied on the data of the characteristics examined in the study, it was found that error terms of the data were normally distributed (P>0.05). Pearson correlation coefficients between live weights and some body measurements taken from Karayaka lambs and significance test results are shown in Table 2.

There is positive correlation between the live weights and examined body measurements of Karayaka lambs at weaning period. While the highest correlation was found between front rump width and middle rump width (r=0.96, P<0.01), the lowest correlation was found between height at withers and chest width (r=0.11, P>0.05). In case of correlation coefficients between the variables examined being around 90%, multicollinearity problem should be considered. For this purpose, multiple regression analysis results of the variables examined are given below respectively according to LS, RR and PCR methods used.

Results of Least Square Method

Prediction coefficient, standard error, test statistics and VIF values of each parameter in LS method applied to form the equation of the association between live weights of Karayaka lambs at the period of weaning and their morphological traits (WH, RH, etc) used as independent variable are given in Table 3.

According to multiple regression analysis results conducted by using LS method, regression coefficients of height at withers, height at rumps, body length, chest depth, chest width and front, middle and hind rump width, which are body measurements used in the prediction of live weight, were found to be statistically insignificant. In addition, multicollinearity was found between independent variables front rump and middle rump width (VIF>10) (Table 3). This result shows that inconsistent parameter predictions were made since standard error increased (for example, while the regression coefficient of RH is 0.234, the standard error of this coefficient is 0.319)

Results of Ridge Regression Method

By using Ridge regression technique, it was found that 62.7% of the linear association between live weights and morphological characteristics of Karayaka lambs at weaning period was explained and that the regression equation to be used in the prediction of this association was significant ($P < 0.001$).

When the condition index (CI) coefficients calculated for Equation X were examined, multicollinearity problem was considered since $CI > 10$ (Table 4).

Table 5 shows Ridge regression analysis results determining k bias coefficient which eliminates multicollinearity problem and gives the highest R^2 value, while Table 6 shows variance inflation factors.

As can be seen in Tables 5 and 6, it can be seen that VIF values are lower than 10, in other words, multicollinearity is eliminated and k value that gives the highest R^2 value is 0.011 (1.1%). At the same time, as reported by Anderson (1998), it can be seen that the 0.011

(1.1%) is the k value that gives the sum of acceptable least error squares ($0.199^2 = 0.0396$) and minimum VIF (VIF value < 10) and maximum R^2 (62.7%) for all variables. Table 7 shows regression coefficients predicted according to $k = 1.1\%$ bias constant and Ridge regression method and significance tests of these.

When VIF values are examined, it can be seen that multicollinearity problem between front rump and middle rump measurements, which are body measurements used for live weight prediction, was eliminated with RR method (Table 7). In addition, a significant difference was found between the predicted Ridge regression parameters and LS parameters given in Table 4 in terms of both coefficients and the standard deviations of these coefficients. While this coefficient predicted with LS method was statistically insignificant especially due to the decrease in standard deviation value of regression coefficient predicted for chest width, it was found to be statistically significant when predicted with RR technique.

Table 5 k parameter selection

k	R^2	Sigma	B'B	Average. VIF	Max VIF
0.00	0.634	0.198	0.687	5.671	16.490
0.01	0.628	0.199	0.548	4.212	10.278
0.011	0.627	0.199	0.539	4.111	9.867
0.02	0.623	0.201	0.472	3.398	7.097
0.03	0.618	0.202	0.423	2.877	5.243
⋮	⋮	⋮	⋮	⋮	⋮
0.08	0.600	0.206	0.311	1.702	2.018
0.09	0.597	0.207	0.299	1.579	1.866

Table 6 Variance Inflation Factor (VIF) values

k	WH	RH	BL	CD	CW	CG	FR	MR	HR
0.00	3.14	2.92	1.89	2.38	1.88	3.70	13.98	16.49	4.66
0.01	2.84	2.69	1.79	2.23	1.70	3.33	8.99	10.28	4.07
0.011	2.81	2.67	1.78	2.21	1.69	3.30	8.65	9.87	4.02
0.02	2.59	2.49	1.70	2.09	1.59	3.04	6.39	7.10	3.60
0.03	2.39	2.32	1.63	1.97	1.49	2.79	4.85	5.24	3.21
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.08	1.70	1.68	1.33	1.51	1.20	1.95	1.99	1.95	2.02
0.09	1.60	1.59	1.28	1.44	1.16	1.83	1.75	1.70	1.87

Table 7 Regression analysis results according to $k = 1.1\%$ bias constant and Ridge regression method

Independent Variable	Coefficients	Std. Deviation	Standardized regression coefficients	t-value	Significance level	VIF
Constant	-6.640					-
WH	0.517	0.365	0.167	1.417	-	2.81
RH	0.237	0.374	0.073	0.635	-	2.67
BL	-0.011	0.260	-0.004	-0.041	-	1.78
CD	0.544	0.290	0.196	2.221	*	2.21
CW	-0.010	0.110	-0.008	-0.088	-	1.69
CG	1.243	0.306	0.520	4.063	**	3.30
FR	-0.516	0.273	-0.392	-1.890	-	8.65
MR	0.257	0.290	0.196	0.885	-	9.87
HR	0.119	0.234	0.072	0.510	-	4.02

** : $P < 0.01$, * : $P < 0.05$, - : $P > 0.05$

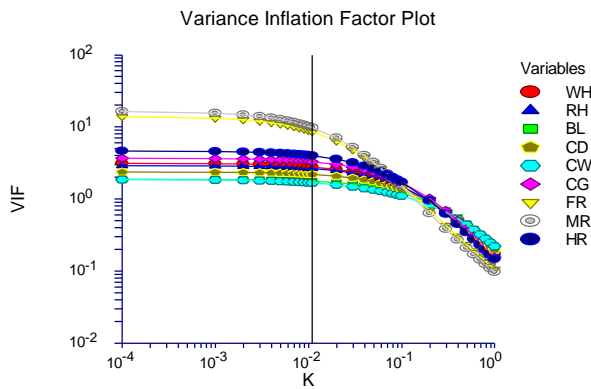


Figure 1 Variance Inflation Factor Plot for Ridge Regression with k = 1.1% bias constant

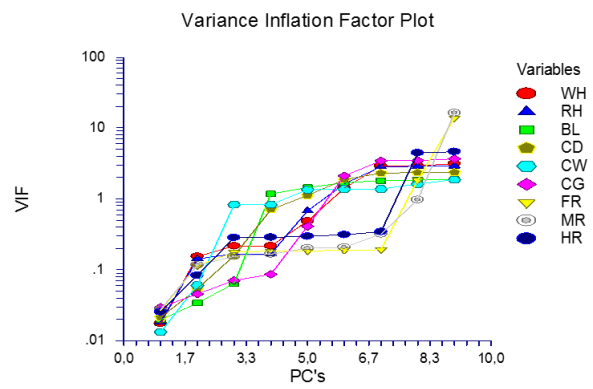


Figure 2 Variance Inflation Factor Plot for Principal Component Regression Analysis

Table 8 Descriptive statistics of principal component regression analysis

Principal Component	PC Coefficient	Individual R-Squared	Eigenvalue
PC1	-0.089	0.414	4.993
PC2	-0.092	0.146	1.641
PC3	-0.03	0.007	0.755
PC4	0.049	0.015	0.585
PC5	0.054	0.013	0.422
PC6	-0.091	0.019	0.215
PC7	0.047	0.005	0.208
PC8	0.056	0.005	0.145
PC9	-0.171	0.011	0.035

Table 9 Principal component regression analysis results

Independent Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient	t	P	VIF
Intercept	-6.670					
WH	0.629	0.377	0.204	1.669	*	2.958
RH	0.191	0.393	0.059	0.486	-	2.906
BL	-0.060	0.268	-0.021	-0.222	-	1.862
CD	0.564	0.302	0.204	1.866	*	2.369
CW	0.039	0.109	0.033	0.359	-	1.637
CG	1.195	0.318	0.499	3.762	**	3.506
FR	-0.164	0.127	-0.124	-1.289	-	1.851
MR	-0.160	0.092	-0.122	-1.745	*	0.976
HR	0.187	0.249	0.113	0.753	-	4.495

Table 10 Comparison of LS, RR and PCR analysis results

Methods	MSE	R ²	R ² _{adj}	CV (%)	Significance level of the method
LS	0.0390	0.634	0.590	0.1910	<0.001
RR	0.0393	0.627	0.587	0.0609	<0.001
PCR	0.0429	0.623	0.549	0.0613	<0.001

Table 4 shows multicollinearity problem in VIF values found in front and middle rump widths. Figure 1 shows the elimination and stabilization of this problem with k = 1.1% bias constant.

Principal Components Regression Method

Principal components regression analysis results of morphological characteristics measured at weaning period of Karayaka lambs are shown in Tables 8 and 9, respectively.

According to PCR analysis, principal component 9 was found to have a low eigenvalue. The study was repeated by deleting two and three eigenvalue and a decrease was

found in the explanatory power. For this reason, it was decided to delete only one component with very low eigenvalue, that is Z value, and to interpret the results.

Figure 2 shows the elimination and stabilization of this problem with one low eigenvalue disposal with PCR analysis.

It was found that the multicollinearity problem found for front rump and middle rump width, which are among independent variables used for live weight prediction of Karayaka lambs at weaning period was eliminated with principal components regression analysis (Table 9). According to the results obtained, it was found that WH, CD, CG and FR were statistically significant while predicting regression equation.

Comparison Results of the Method

Table 10 shows mean squared error (MSE), coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}) and Coefficient of Variation (CV%) values of LS, RR and PCR methods used in the live weight prediction of Karayaka lambs at weaning period.

When Table 10 is examined, it was found that all of the models obtained with three different methods were statistically significant ($P < 0.001$). It was found that following LS method, RR method had the lowest MSE and the highest R^2 value. However, when Table 7 is taken into consideration, it was found that regression coefficients of all variables except for chest girth and chest depth were statistically insignificant, while principal components regression analysis equation showed that coefficients of height at withers and height at rumps were statistically significant besides chest girth and chest depth (Table 10). For this reason, it was found that PCR predictions were more consistent.

Conclusion

In agricultural researches based on cause and effect relation, multiple regression analysis is used in the assessment of data, while least squares method (LS) is preferred in the prediction of coefficients of regression equation since they are easily calculated and the results are easy to understand. However, validity of this method is decreasing since the data of the characteristics which are the subject of the research (for example, using body measurements to predict live weight in husbandry) most of the time cannot meet the necessary assumptions such as no significant association between independent variables. The factor that causes the breakdown of the related assumption and the LS method to become indefensible is the multicollinearity problem between independent variables. Due to these problems, multicollinearity problem should be eliminated so that the data obtained from agricultural researches can be interpreted healthily.

In our study, first of all parameter estimations obtained with least squares (LS) method were found. Since a multicollinearity problem was found for parameters obtained with LS method, the assumption of not having an internal association between independent variables was not met. As a result of this problem, analysis results obtained by using LS method are erroneous and they can cause model prediction to be wrong (Ergüneş, 2004). Thus, for front and middle rump widths, VIF values, which are one of the indicators of multicollinearity problem, were found to be higher than the values obtained from Ridge regression and principal components regression.

On the other hand, MSE value obtained as a result of analysis with LS method was found to be lower than that obtained with RR and PCR, while R^2 value was found to be higher and these results were similar to the results of Ergüneş (2004), Çamdeviren et al. (2005), Topal et al. (2010) and Üçkardeş et al. (2012).

As a conclusion, multi regression equation predictions were obtained with the help of different statistics programs (SPSS, NCSS) and with the help of the problems used, the effects of multicollinearity problem of independent variables were eliminated directly or indirectly. This way, a regression equation with lower errors, more consistency and thus with stronger prediction power was found. Thus,

using biased predictors Ridge regression (RR) and Principal Components Regression (PCR) instead of LS predictor to eliminate the effects of multicollinearity between independent variables will contribute to healthier interpretation of the results. In addition, it is thought that studies in which the sensitivity and validity of these methods are tested with different sample sizes will contribute to researchers working in this field.

References

- Albayrak AS. 2005. Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifleri yanlı tahmin teknikleri ve bir uygulama. ZKÜ sosyal bilimler dergisi. 1(1): 105-126.
- Albayrak AS. 2006. Uygulamalı çok değişkenli istatistik teknikleri. 1 Baskı. Mamak/Ankara. Asil yayını. ISBN: 975-9091-98-4.
- Alpar R. 1997. Uygulamalı çok değişkenli istatistiksel yöntemler. 1. Baskı. Kızılay/Ankara. Kültür ofset.
- Alpar R. 2011. Uygulamalı çok değişkenli istatistiksel yöntemler. 3. Baskı. Kızılay/Ankara. Detay Yayıncılık. ISBN:978-605-5437-42-8
- Anderson B. 1998. Scandinavian evidence on growth and age structure, ESPE 1997 Conference at Uppsala University.
- Çamdeviren H, Demir N, Kanık A, Keskin S. 2005. Use of Principal component scores in multiple linear regression models for prediction of *Chlorophyll-a* in reservoirs, Ecological Modelling 181: 581-589. DOI: 10.1016/j.ecolmodel.2004.06.043
- Çankaya S, Altop A, Kul E, Erener G. 2009. Faktör analiz skorları kullanılarak karayaka kuzularında canlı ağırlık tahmini. Anadolu tarım bilim. derg, 24(2): 98-102.
- Draper NR, Smith H. 1981. Applied regression analysis. 2nd Edition. New York: John Wiley & Sons, Inc.
- Ergüneş E. 2004. En küçük kareler yöntemi ile ridge regresyon yönteminin karşılaştırılması olarak incelenmesi, Yüksek Lisans Tezi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü.
- Gunst RF, Mason RL. 1980. Regression analysis and its application: A data-oriented approach. New York: Marcel Dekker.
- Hintze JL. 2007. NCSS User's Guide III - Regression and Curve Fitting, Chapter 340- Principal Components Regression, Kaysville/Utah. NCSS Statistical System.
- Hoerl AE, Kennard RW. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1): 55-67.
- Huber PJ, Dutter R. 1974. Numerical solution of robust regression problems. In: Compstat 1974 (Ed: G. Brickmann) 165 – 172. Phsika Verlag, Wein.
- Huber PJ. 1974. Numerical solution of robust regression problems. In COMPSTAT 1974, Proc. Symposium on Computational Statistics. Physike Verlag.
- Kleinbaum DG, Kupper LL, Muller KE, Nizam A. 1998. Applied regression analysis and other multivariable methods, R.R. Donnelly and Sons Duxbury Press, USA.
- Kurtuluş M. 2001. Ridge regresyon üzerine bir çalışma. Yüksek Lisans Tezi. Gazi Üniversitesi, Fen Bilimleri Enstitüsü. Ankara.
- Sangun L, Çankaya S, Kayaalp GT, Akar M. 2009. Use of factor analysis scores in multiple regression model for estimation of body weight from some body measurements in lizardfish. J. Anim. Vet. Adv. 8: 47-50.
- Topal M, Eyduran E, Yağanoğlu AM, Sönmez A, Keskin S. 2010. Çoklu doğrusal bağlantı durumunda ridge ve temel bileşenler regresyon analiz yöntemlerinin kullanımı. Atatürk Üniversitesi Ziraat Fakültesi Dergisi, 41(1): 53-57.
- Üçkardeş F, Ercan E, Nariç D, AKSOY T. 2012. Japon bildircinlarında yumurta ak indeksinin ridge regresyon yöntemiyle tahmin edilmesi. Akademik Ziraat Dergisi, 1(1): 11-20