Prediction Model of Chick Hatching Weight with Restricted Ridge Regression

Betül Dağoğlu Hark* and Sema Alaşahan²

¹Department of Biostatistics and Medical Informatics, School of Medicine, Firat University, Elazig, Turkey.
²Department of Animal Science, Faculty of Veterinary Medicine, Hatay Mustafa Kemal University, Hatay, Turkey.

Authors’ contributions

This work was carried out in collaboration between both authors. Author BDH designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author SA wrote the first draft of the manuscript and performed data collection. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRIZ/2020/v3i230089

ABSTRACT

In this study, it is aimed to obtain prediction model for chick hatching weight by using shell ground color, shape characteristics, shell characteristics, white + yellow weight and weight loss property values during hatching of quail hatching eggs. Multicollinearity is one of the most common problems in prediction models. It is defined as linear dependence between explanatory variables in linear regression analysis. In this case, the ordinary least squares (OLS) estimator is not a predictor with minimum variance between unbiased estimators. Accordingly, the parameter estimates are away from the real value. One proposed approximation approach to eliminate this problem is restricted ridge regression.

According to the findings obtained from 523 chick hatching eggs, elongation, egg shell weight, shell weight + embryo waste (g), absolute weight loss after transfer (g), relative weight loss after transfer (%) and absolute weight loss in 0-14 days (g) were determined as factors affecting chick hatching weight. In the linear modeling approach, which is used to determine these factors, multicollinearity problem is revealed. Therefore, in order to obtain a stable set of parameters,
restricted ridge regression analysis was applied and estimations with lower standard errors than standard errors of the parameters obtained from the (least squares) LSQ approach were obtained.

Keywords: Chick hatching weight; linear regression; multicollinearity; restricted ridge regression.

1. INTRODUCTION

In general, the purpose of the regression model;

a. To determine the extent to which the dependent variable can be estimated by independent variables,

b. To reveal the power of a theoretical relationship between dependent and independent variables [1].

The term “multicollinearity” was first used in the literature by Frisch (1934) [2]. The existence of a strong linear dependence between two or more explanatory variables is defined as multicollinearity [3].

It is important to understand which explanatory variables are causing the problem in the presence of multi-collinearity, and the easiest solution would be to remove them from the analysis. However, doing this is not always a good solution because it changes depending on whether the purpose of the analysis is explanation or prediction. If the goal is to prediction and both variables have the same “meaning”, removing one of the explanatory variables from the analysis will not cause any problems. However, for explanation purposes, explanatory variables are selected according to a theoretical rule, so it is not possible to exclude any variable from the analysis: the model, including every selected variable, must be fully explained. In the literature, method recommendations for dealing with multicollinearity without removing the variable from the model are presented. Restricted ridge regression analysis is one of the proposed methods. Restricted ridge regression is an advanced approach based on the combination of restricted least squares and ridge estimators [1].

It has been reported in many studies that the hatching weight of poultry eggs is affected by egg characteristics and egg weight loss during hatching [4-11]. The aim of this study was to determine the factors affecting the chick hatching weight by restricted ridge regression approach. For this purpose, firstly, the multi-collinearity problem in the data set was determined and the factors affecting the chick hatching weight were determined by OLS and restricted ridge estimators. A secondary objective is to show that the parameter estimates obtained by the restricted ridge approach are more stable than the OLS approach in case of multi-collinearity.

2. MATERIALS AND METHODS

2.1 Egg Material

The study material was obtained from hatching eggs of 3 months old Japanese quail (Coturnix coturnix japonica) breeding flocks in 1 day. Hatching and data collection process were performed in Mustafa Kemal University, Faculty of Veterinary Medicine Alternative Poultry Application and Research Unit. The hatching eggs were examined visually for shell defects and cracked, broken-shelled and abnormally shaped eggs were separated. Then a total of 752 eggs were loaded on the hatching machine. Furthermore, the hatching eggs were individually numbered before being loaded into the machine. The individual measurement values determined from the hatching eggs were noted in the prepared chart.

2.2 Chick Material

After the initial egg weights of the hatching eggs were determined, the eggs were placed in the hatching machine trays. Egg plates were randomly placed in the hatching machine (growth + hatch) with a temperature of 37.5°C degrees and a relative humidity of 65%. The hatching machine room temperature was set to 33-35°C for the measurement during transfer of the hatching eggs from the developmental part to the hatching part. The hatching eggs were placed in outlet boxes with individual egg chambers and transferred back to the hatching machine for the last 3 days of embryo development. At the end of the incubation period, 523 hatching eggs were used.

2.3 Features

The variables examined before, during and after the hatching of 523 hatching chick hatcher are presented below to determine the factors affecting the hatching weight;
2.3.1 Egg background color
The quail eggs were examined visually for shell background color and were divided into two classes as grayish white and grayish brown background.

2.3.2 Egg weight (g)
The weight of the eggs was determined by weighing the electronic scales as the initial egg weight and the transfer (14th day of embryo development) egg weight. An electronic balance of 0.01 g was used to weigh the hatching egg weight values.

2.3.3 Egg shape features
Egg length, egg width and shape index values were determined in hatching eggs. Determination of values [12];

\[
\text{Egg length (mm)} = 13.04938 \times (\text{Initial egg weight})^{0.373272} \\
\text{Egg width (mm)} = 8.01571 \times (\text{Initial egg weight})^{0.448338} \\
\text{Egg Shape Index} (\%) = (\text{Egg width} / \text{Egg length}) \times 100 \\
\text{Elongation value} = \text{Egg length} / \text{Egg width}
\]

2.3.4 Egg shell properties
Shell thickness, shell weight, shell pore number and shell pore density were calculated as hatching egg shell characteristics. Equations used in the calculation [13,14,15];

\[
\text{Egg shell Weight (g)} = 0.0524 \times (\text{Initial egg weight})^{1.113} \\
\text{Egg shell thickness (mm)} = 0.0546 \times (\text{Egg length})^{0.441} \\
\text{Number of shell pores (pcs)} = 304 \times (\text{Starting egg weight})^{0.767} \\
\text{Shell pore density (pcs / cm}^2) = \text{Number of shell pores} / \text{Egg surface area} \\
\text{Surface area (cm}^2) = 4,835 \times (\text{Initial egg weight})^{0.662}
\]

2.3.5 Egg internal structure feature
Together weights of hatching egg white and yellow were calculated;

\[
\text{White + yellow weight (g)} = (\text{Initial egg weight} - \text{Shell weight})
\]

2.3.6 Egg weight loss
Hatching Eggs were transferred to hatching machine on the 14th day of embryo development. Absolute and relative weight loss values were calculated before transfer using the egg weight on the 14th day (transfer).

\[
\text{Absolute weight loss between days 0 and 14 days (g)} = \text{Initial egg weight} - \text{Transfer egg weight} \\
\text{Relative weight loss between days 0 and 14 days (\%)} = [(\text{Initial egg weight} - \text{Transfer egg weight}) / \text{Initial egg weight}] \times 100
\]

2.3.7 Chick hatching weight
The chick hatching weights after the 17th day of the hatching period were determined by recording the individual egg numbers of the live chicks after drying their feathers. At the same time, embryo waste shell weight was recorded after hatching.

Absolute and relative weight loss values after transfer were calculated using chick hatching weight and embryo waste shell weight.

\[
\text{Absolute weight loss after transfer (g)} = \text{Transfer egg weight} - (\text{Chick hatch weight + Embryo waste shell weight}) \\
\text{Relative weight loss after transfer (\%)} = [(\text{Transfer egg weight} - (\text{Chick hatch weight + Embryo waste shell weight}) / \text{Transfer egg weight}] \times 100
\]

2.4 Methods
Linear model; The nx1 dimensional y response vector, n, the number of observations and p, the number of the regressor variables, is modeled by the nxp dimensional X explanatory variable matrix. So the regression model,

\[
y = X\beta + \varepsilon
\]

Here; \(\beta\), px1 dimensional unknown constant vector and \(\varepsilon\), nx1 dimensional random error vector (\(\varepsilon \sim N(0, \sigma^2)\)). Parameter estimations under OLS estimator,
\[ \hat{\beta} = (X'X)^{-1}X'y \]
\[ \text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \]

\[ X_j, \text{ is the j-th column vector of the matrix X;} \]
\[ \sum_{j=1}^{p} t_j X_j = 0 \]

If there is a set of constants \( t_1, t_2, ..., t_p \), at least one of which is zero, the vectors \( X_1, X_2, ..., X_p \) are linearly dependent. If the equation (3) is approximately true for a subset of some of the columns of \( X \), there will be a close linear dependence and a multicollinearity problem will arise [16].

Methods used to determine multicollinearity:

1. Investigation of the correlation matrix: The non-diagonal elements (\( r_{ij} \)) of the \( X'X \) matrix take values close to one, indicating linear dependence.
2. Analysis of eigenvalues of the \( X'X \) matrix: If there is one or more linear dependencies in the data, one or more eigenvalues will be small. In addition, the number of conditions is calculated based on eigenvalues, and greater than 100 indicates the presence of a strong multicollinearity.
3. Determination of variance inflation factor (VIF): Any VIF value greater than 10 indicates the presence of multicollinearity.
4. Evaluating the determinant of the matrix \( X'X : |X'X| = 0 \) represents a full multicollinearity.

Assumptions for linear model: linearity, normality, constant variance and independence. In the presence of multicollinearity, the first three assumptions can be achieved, but the assumption of independence is violated. In this case, the parameter estimates obtained are unbiased. However, it is not a predictor with minimum variance between unbiased estimators. Therefore, parameter estimates are no longer true. The proposed estimation approaches to eliminate this problem have led to the emergence and examination of biased estimators.

First, ridge regression was proposed by Hoerl and Kennard (1970) [17]. This approach is based on the addition of the \( k \) bias parameter to the \( \beta \) estimates.

\[ \hat{\beta}(k) = (X'X + kI)^{-1}X'y \]

In this approach, the choice of \( k \) affects the performance of the estimator. Hoerl and Kennard (1970) suggested using the ridge trace to find the appropriate \( k \)-value at which regression coefficients are stabilized [17]. Ridge trace is a graph plotted according to parameter estimates obtained with ridge estimator versus \( k \) value defined in the range 0-1. As \( k \) increases, the predictions become smaller and after a certain \( k \), the estimates become stable. The smallest \( k \) value at which the stable point is provided is determined as the most appropriate \( k \) value. However, many approaches have been proposed to determine the appropriate \( k \)-value [17,18,19].

As an alternative to ridge regression, the restricted ridge regression proposed by Sarkar (1992) [20] and developed by Groß (2003) [21] is a more advanced approach based on the combination of restricted least squares and ridge estimators. In this approach, it is assumed that the matrix \( X \) is not a full rank, that is, \( \beta \) provides the RDR = \( r \) linear constraint for the case of multicollinearity. Here, the R matrix with \( m <p \) rank is \( m \times p \) and \( r \) vector is \( m \times 1 \). Groß (2003) recommended restricted ridge estimator [21]:

\[ \hat{\beta}(k, \beta_0) = \hat{\beta}(k, \beta_0) - S_k^{-1}R'[RS_k^{-1}R']^{-1}(R\hat{\beta}(k, \beta_0) - r), \]
\[ k \geq 0 \]
\[ \hat{\beta}(k, \beta_0) = (X'X + kI)^{-1}(X'y + k R(R'R)^{-1}R) \]
\[ S_k = X'X + kI \]

2.5 Statistical Analysis

IBM SPSS Statistics Version 22.0 [22] package program and R version 3.5.0 program [23] were used for statistical analysis of the data. Continuous measurements were summarized as mean and standard deviation. Pearson Correlation Coefficient and related \( p \) value were obtained to examine the relationship between continuous measurements. Backward linear regression was used to determine the factors affecting the hatching weight. Multi-collinearity problem was presented by correlation matrix, eigenvalue analysis and VIF values. In the case of multi-collinearity, parameter estimations were performed with linear and restricted ridge regression analyzes.

3. RESULTS AND DISCUSSION

In Table 1, descriptive statistics for all variables in the regression model are given to determine the factors affecting the chick hatch weight.
The model containing significant explanatory variables was determined by the backward linear regression approach. According to this approach, the obtained explanatory variables; elongation, egg shell weight, shell weight + embryo waste (g), absolute weight loss after transfer (g), relative weight loss after transfer (%) and absolute weight loss between 0-14 days (g).

In order to determine the multicollinearity problem, the correlation matrix of the response variable and meaningful variables is given in Table 2. There is a high correlation between the explanatory variables.

Secondly, the eigenvalues of the $X'X$ matrix were analyzed in order to determine the multicollinearity problem. The obtained eigenvalues; $\lambda_1 = 2.632$, $\lambda_2 = 2.027$, $\lambda_3 = 0.943$, $\lambda_4 = 0.392$, $\lambda_5 = 0.004$ and $\lambda_6 = 0.002$. One or more small eigenvalues imply that there are near-linear dependences among the explanatory variables. Since two eigenvalue values are close to zero, it can be mentioned that multiple connections exist in the dataset. The number of conditions obtained was found to be $\kappa = 121.5$ and the presence of a strong multicollinearity between the explanatory variables was determined because the number of conditions was greater than 100. Finally, the obtained VIF values are 207.392, 233.249, 1.744, 160.399 and 130.772. All VIF values except egg shell weight + embryo waste value is higher than 10, indicating the presence of multicollinearity.

Fig. 1 shows the ridge trace plotted according to the standard parameter estimates obtained with the ridge estimator versus the k value defined in the range 0 to 0.010. The smallest k value at which the stable point was obtained was determined to be 0.008.

Table 1. Descriptive statistics for variables

| Variables                              | mean±sd    |
|----------------------------------------|------------|
| Chick Hatching Weight (g)              | 8.881±0.910|
| Egg Weight (g)                         | 12.219±1.094|
| Egg Length (mm)                        | 33.184±1.123|
| Egg Width (mm)                         | 24.595±0.998|
| Egg Shape Index (%)                    | 74.101±0.508|
| Elongation Value                       | 1.350±0.009|
| Egg Shell Weight (g)                   | 0.850±0.085|
| Egg Shell Thickness (mm)               | 0.256±0.004|
| Number of Shell Pores (pcs)            | 2071.578±142.946|
| Shell Pore Density (pcs / cm$^2$)      | 81.742±0.783|
| Surface Area (cm$^2$)                  | 25.328±1.512|
| White + Yellow Weight (g)              | 11.369±1.010|
| Transfer Egg Weight (g)                | 11.403±1.086|
| Shell Weight + Embryo Waste (g)        | 0.989±0.110|
| 0-14. Absolute weight loss between days (g) | 0.816±0.285|
| 0-14. Relative weight loss between days (%) | 6.697±2.303|
| Absolute weight loss after transfer (g) | 1.532±0.419|
| Relative weight loss after transfer (%) | 6.697±2.303|

Table 2. Correlation matrix

|                           | 1     | 2      | 3      | 4      | 5      | 6      | 7      |
|---------------------------|-------|--------|--------|--------|--------|--------|--------|
| Chick Hatch Weight (1)    | 1.00  | -0.91  | 0.91   | 0.58   | -0.01  | 0.07   | -0.28  |
| Elongation Value (2)      | 1.00  | -0.99  | -0.58  | -0.16  | -0.37  | -0.05  |
| Egg Shell Weight (3)      | 1.00  | 0.58   | 0.16   | 0.37   | 0.04   |
| Shell Weight + Embryo Waste (4) | 1.00  | 0.00   | -0.01  | -0.23  |
| Absolute weight loss between 0-14 days (5) | 1.00  | 1.00   | 0.94   |
| Absolute weight loss after transfer (6) | 1.00  | 1.00   |
| Relative weight loss after transfer (7) | 1.00  |       |

According to the centralized data, parameter estimations obtained by linear and restricted ridge regression approaches and standard errors of parameter estimations are given in Table 3. According to Table 3, the regression model obtained with the LSQ estimator;
\[ \hat{y}_{\text{chick hatching weight}} = -11.1566x_1 + 11.7205x_2 - 0.9999x_3 - 0.9998x_4 - 0.9992x_5 - 0.0001x_6 \]

Regression model obtained with restricted ridge estimator:

\[ \hat{y}_{\text{chick hatching weight}} = -8.1447x_1 + 8.7091x_2 - 2.5357x_3 - 1.3944x_4 - 1.8671x_5 - 0.6787x_6 \]

When the two approaches are compared, it is seen that the standard error of the parameter estimates obtained by the restricted ridge regression approach is lower than that obtained by the LSQ.

**Fig. 1.** The ridge trace according to the parameter estimates obtained with the ridge estimator against the k value defined in the range of 0 to 0.10

**Table 3.** Parameter estimates obtained by linear and restricted ridge regression approaches

| Variables                              | OLS       | Restricted Ridge Estimator |
|----------------------------------------|-----------|-----------------------------|
|                                        | \( \hat{\beta} \) | se(\( \hat{\beta} \)) | p value | \( \hat{\beta} \) | se(\( \hat{\beta} \)) | p value |
| Elongation Value (1)                   | -11.1566  | 0.0232                      | <0.001  | -8.1447            | 0.0001               | <0.001  |
| Egg Shell Weight (2)                   | 11.7205   | 0.0027                      | <0.001  | 8.7091             | 0.0001               | <0.001  |
| Shell Weight + Embryo Waste(3)         | -0.9999   | 0.0002                      | <0.001  | -2.5357            | 0.0000               | <0.001  |
| 0-14. Absolute weight loss between days (4) | -0.9998  | 0.0001                      | <0.001  | -1.3944            | 0.0000               | <0.001  |
| Absolute weight loss after transfer (5) | -0.9992  | 0.0005                      | <0.001  | -1.8671            | 0.0002               | <0.001  |
| Relative weight loss after transfer (6) | -0.0001  | 0.0001                      | 0.075   | -0.6787            | 0.0002               | -        |

**4. CONCLUSION**

Multicollinearity leads to high sample variance of regression parameters estimated in OLS estimation approach. In order to overcome this problem, biased but low variance estimators have been developed [16]. Restricted ridge regression is one of these predictors. This regression approach is based on the combination of restricted LSQ and ridge regression approach.
Studies have found that in cases where the constraints are defined correctly, the restricted ridge estimator gives estimates with lower standard errors than both the restricted LSQ approach and the ridge regression approach [16,21]. Therefore, the factors affecting chick hatching weight were obtained by restricted ridge regression analysis. The standard error of the regression parameters obtained was found to be significantly lower than the OLS approach due to multicollinearity.

Factors affecting chick hatching weight; elongation, egg shell weight, shell weight + embryo waste (g), absolute weight loss after transfer (g), relative weight loss after transfer (%) and 0-14. absolute weight loss between days (g). However, the multicollinearity problem that arose due to the strong correlation between the explanatory variables was solved by restricted ridge regression analysis.

The study has many limitations. One of them is that there are many different approaches for choosing the ridge parameter k, but these approaches are not included in this study. Another limitation is not taken into account in model fit assessments.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Giacalone M, Panarello D, Mattera R. Multicollinearity in regression: An efficiency comparison between I p-norm and least squares estimators. Quality & Quantity. 2018;52(4):1831-1859.

2. Frisch R. Statistical confluence by means of complete regression systems, vol. 5. Universitets økonomiske Institut, Copenhagen; 1934.

3. Leahy K. Multicollinearity: When the solution is the problem. Data Mining Cookbook. 2000;106-108.

4. Saatci M, Kirmizibayrak T, Aksoy AR, Tilki M. Egg weight, shape index and hatching weight and interrelationships among these traits in native turkish geese with different coloured feathers. Turk J Vet Anim Sci. 2005;29:353-357.

5. Oral Toplu HD, Dereli Fidan E, Nazlıgül A. Japon bildircinlarda kuluçkalık yumurta ağırlığı ve depolama süresinin kuluçka özelliklileri ve civciv çıkış ağırlığı üzerine etkileri. Erçiyes Univ Vet Fak Derg. 2007;4(1):11-16.

6. Daikwo SI, Dim NI, Momoh MO. Hatching characteristics of japanese quail eggs in a tropical environment. International Journal of Poultry Science. 2011;10(11):876-878.

7. Nwachukwu EN, Ogbu CCE, Ifeanacho NG. Egg quality characteristics and hatchability of two colour variants of japanese quails (Coturnix coturnix japonica). International Journal of Livestock Research. 2015;5(11):46-54.

8. Yamak US, Sarica M, Boz MA, Ucar A. The effect of eggshell thickness on hatching traits of partidges. Brazilian Journal of Poultry Science. 2016;18(Special Issue):13-18.

9. Iqbal J, Mukhtar N, Rehman ZU, Khan SH, Ahmad T, Anjum MS, Pasha RH, Umar S. Effects of egg weight on the egg quality, chick quality and broiler performance at the later stages of production (week 60) in broiler breeders. J Appl Poult Res. 2017;26:183-191.

10. Mirahmetoglu C, Celen MF. Japon Bildircinlarda (Coturnix coturnix japonica) Kuluçkalık Yumurta Ağırlığı ve Kuluçkanın Son Döneminde Oksijen Eklentimesinin Kuluçka Sonuçları ve Çıkış Sonrası Performansa Etkileri. Batman University Journal of Life Sciences. 2017;7(2/2):137-149.

11. Ewonetu KS and Kasaye A. Effect of egg weight on hatch performance of white leghorn chicken breed from day old to laying age. Journal of Poultry Research. 2018;15(2):16-22.

12. Alasahan S, Dagoglu Hark B, Eltas O. Prediction of the length and width of quail egg using linear regression analysis. Eurasian Journal of Veterinary Sciences.2019;35(3):152-157.

13. Paganeli CV, Olszowka A, Ar A. The avian egg: Surface area, volume and density. The Condor. 1974;76:319-325.

14. Rahn H, Paganeli CV. Shell mass, thickness and density of avian eggs derived from the tables of Schönwetter. Journal fur Ornithologie. 1989;130:59-68.

15. Rahn H, Paganeli CV. Gas fluxes in avian eggs: Driving forces and the pathway for exchange. Comparative Biochemistry and Physiology Part A. 1990;95(1):1-15.
16. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. John Wiley & Sons. 2012:821.
17. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55-67.
18. Hoerl AE, Kennard RW. Ridge regression iterative estimation of thebiasing parameter. Communications in Statistics-Theory and Methods. 1976;5(1):77-88.
19. Ozkale MR. Comment on ridge estimation to the restricted linear model. Communications in Statistics - Theory and Methods. 2009;38(7):1094-1097.
20. Sarkar N. A new estimator combining the ridge regression and the restricted least squares methods of estimation. Communications in Statistics-Theory and Methods. 1992;21(7):1987-2000.
21. Groß, J. Restricted ridge estimation. Statistics & Probability Letters. 2003;65(1):57-64.
22. SPSS. IBM SPSS Statistics 22.0 for Windows. Armonk,NY; 2013.
23. R foundation for statistical computing. R: A Language and Environment for Statistical Computing, Vienna, Austria: R Core Team; 2013.