Kafkas Univ Vet Fak Derg
17 (5): 687-692, 2011
DOI:10.9775/kvfd.2010.3974

RESEARCH ARTICLE

Comparision of Principal Component Regression with the Least Square Method in Prediction of Internal Egg Quality Characteristics in Japanese Quails

Tülin ÇİÇEK RATHERT * Fatih ÜÇKARDEŞ ** Doğan NARİNÇ *** Tülin AKSOY ***

* Kahramanmaraş Sütçü İmam Üniversitesi Ziraat Fakültesi Zootekni Bölümü, TR-46100 Kahramanmaraş - TÜRKİYE
** Ordu Üniversitesi Ziraat Fakültesi Zootekni Bölümü, TR-52200 Ordu - TÜRKİYE
*** Ankdeniz Üniversitesi Ziraat Fakültesi Zootekni Bölümü, TR-07100 Antalya - TÜRKİYE

Makale Kodu (Article Code): KVFD-2010-3974

Summary

The purpose of this study is to determine the inner egg quality characteristics albumen height, albumen width, albumen length, yolk diameter and yolk height using principal component regression. For this reason 104 eggs of Japanese quails between the 20th and 24th week of age were analysed. The birds had not been object of selection and they were paired randomly. The problem of multicollinearity occurs in regression models between independent variables in case of high correlation. In case of multicollinearity, parameter estimations and hypothesis results contradict each other through large standard deviation caused by Least Square method (LS). One of the methods that are applied in such a case is principal component regression (PCR). PCR leads to small standard deviation and more accurate and more reliable regression equations. For this research the external egg quality traits egg weight (X_1), egg width (X_2), egg length (X_3) and shape index (X_4) were used as variables. By processing these variables using LS and PCR the inner egg quality traits albumen height, albumen width, albumen length, yolk diameter and yolk height were estimated. In either method the regression estimating equations of the inner egg quality were significant (P<0.01). The goodness of fit of the regression estimating equations was 29.24%-68.25% when LS and 29.12%-66.72% when PCR was applied.

Keywords: *Principal component regression, Multicollinearity, Least square method, Japanese quail*

Japon Bıldırcınlarında Yumurta İç Kalite Özelliklerinin Tahmin Edilmesinde Temel Bileşenler Regresyon Yöntemi İle En Küçük Kareler Yönteminin Karşılaştırılması

Özet

Bu çalışmanın amacı Japon bıldırcınlarında iç kalite özelliklerinden olan Ak yüksekliği, Ak genişliği, Ak uzunluk, Sarı uzunluk ve Sarı Yüksekliğinin Temel Bileşenler Regresyon yöntemiyle belirlenmesidir. Bu amaçla 20-24 haftalık yaşlar arasında bulunan, seleksiyon uygulanmamış, rastgele çiftleşmiş Japon bıldırcınlarından toplanan yumurtalar kullanılmıştır. Regresyon modelinde bağımsız değişkenler arasında yüksek bir korelasyon durumunda çoklu bağlantı adı verilen bir problem meydana gelir. Çoklu bağlantı durumunda parametre tahminleri en küçük kareler yöntemi ile standart hataları büyük ve hipotez sonuçları çelişki içindedir. Çoklu bağlantı (multicollinearity) problemi ile uğraşan yöntemlerden biri temel bileşenler regresyon yöntemidir. Temel Bileşenler Regresyon yöntemi kullanarak küçük standart hata, daha doğru ve güvenilir regresyon denklemleri elde edilir. Bu çalışmada Japon bıldırcınlarında yumurta dış kalite özelliklerinden; yumurtanın ağırlığı (X), yumurtanın genişliği (X_2), yumurtanın uzunluğu (X_3) ve şekil indeksi (X_4) değişkenleri kullanılmıştır. Bu değişkenlerle yumurta iç kalite özelliklerinden; ak yüksekliği, ak genişliği, ak uzunluğu, sarı uzunluğu ve sarı yüksekliği hem en küçük kareler yöntemi hem de temel bileşenler regresyon yöntemi ile tahmin edilmiştir. Her iki yöntemde yumurta iç kalite özelliklerinin regresyon tahmin denklemleri önemli bulunmuştur (P<0.01). Regresyon tahmin denklemlerinin uyum iyiliği LS yönteminde %29.24-%68.25, PCR yönteminde ise %29.12-%66.72 aralığında bulunmuştur.

Anahtar sözcükler: *Temel bileşenler regresyon, Çoklu bağlantı, En küçük kareler metodu, Japon bıldırcını*

İletişim (Correspondence)
+90 242 3102480
zootekni@akdeniz.edu.tr

INTRODUCTION

Japanese quails are used as a model in poultry breeding because of their various characteristics. They are also used commercially in meat and egg production in many countries [1-6]. Inner egg quality characteristics such as albumen and yolk indices are of high importance in determining egg quality in the context of commercial egg production. These characteristics both indicate the commercial value of the products and help estimate the quality of the chick and breed stock, respectively [1,7]. For these reasons receiving information on inner egg quality without breaking the shell is crucial. Without breaking the egg, regression prediction equations are used to determine internal egg quality characteristics. However, some problems can be faced in the case that estimated independent variables highly dependent each other. Principal ones of these problems are multicollinearity and non-significant regression parameters that are close to zero. The methods that were developed to cope with these problems are called as biased regression predictors. The most commonly used one is Principal Component Regression (PCR). In this study this method is also considered to be used. One of the aims of this study is to determine multicolinearity and to present the comparison PCR method with Least Squares Method. The purpose of this study is to determine the inner egg quality characteristics albumen height, albumen width, albumen length, yolk diameter and yolk height using principal component regression.

MATERIAL and METHODS

In this study, 104 eggs of Japanese quails between the 20th and 24th week of age were analyzed. The birds had not been object of selection and they were paired randomly. The birds were fed with starter fodder containing 24% HP, 2.900 kcal/kg ME for the first 3 weeks, with mixed fodder containing 20% HP, 2.800 kcal/kg ME between the 4th and the 6th week and with laying fodder containing 17% HP, 2.700 kcal/kg ME after the 6th week. During the first 3 weeks 23 h/day lighting were applied, in the following periods 16 h lighting and 8 h darkness. The eggs were examined for egg weight (g), egg length (mm), egg width (mm), shell thickness (mm) and shell weight (g). Egg weights were measured by digital balance to the nearest 0.1 g. Egg width, egg length, yolk width, albumen length and -width were measured by digital compass to the nearest 0.01 mm. Yolk and -albumen heights were measured by micrometer to the nearest 0.01 mm. Egg quality characteristics were calculated by using the Haugh unit formula [1,7,8].

One method for estimating parameters in multi-regression is Least Square method (LS). LS is not only advantageous for estimating parameters but it is also an unbiased method that reduces the deviance between observed value and estimated value in a model to a minimum. To ensure the validity of this model, assumptions such as independence of errors, normal distribution and non-correspondence of independent variables must be valid. If any of these assumptions is not given, the reliability of the model is reduced and interpretations can become wrong. To estimate the inner egg qualities albumen height, -width and - length, yolk length and -height, it was made use of the outer egg quality characteristics egg weight (X_1), egg width (X_2), egg length (X_3) and shape index (X_4). To estimate the inner egg qualities LS based multi-regression was applied. In case that the assumptions mentioned above do not apply, the regression parameters calculated by LS method drift apart from the real values. Different measurements of the same experimental object (egg) can lead to strong correlations among independent valuables. This is called multicollinearity in regression models [9,11,13]. In such a case parameters variance increases significantly and appear non-significant according to t-test. A lot of researchers do not attend to the levels of signification when using regression equations. To avoid this problem the use of Principal Component Regression (PCR), a biased estimation method, instead of LS is preferable [10,13]. In this study, both LS based multi regression and PCR were used because significant relations had to be expected among the different measurements due to the fact that more than one measurement was taken to estimate inner egg quality traits. Moreover, detailed data was given on multicollinearity.

Multicollinearity as a problem in regression equations can be categorized through some criteria [11]. It is possible to range the most frequent methods for detecting multi-collinearity as follows:

1. Simple correlation coefficient: High correlation between independent variables ($r \geq 0.75$) suggests multi-collinearity.

2. Variance Inflaction Factor (VIF): Variance inflaction factor is a method to detect multicollinearity. Diagonal elements of the matrix including $(X'X)^{-1} = C$ are likely to create multicollinearity [12]. In calculating VIF values, partial correlation coefficients are utilized. The VIF factor is calculated with the following formula:

$$C_{ij} = \frac{1}{1 - R_{ij}} \qquad \text{(i)}$$

Here R_{ij} is the partial correlation coefficient. If the VIF factor $C_{ij} \geq 10$, multicollinearity is assumed [10].

3. (X'X) Eigenvalues of the matrix: To detect the severity of multicollinearity (X'X) it is benefited from the eigenvalues of the matrix. In case that there is no multicollinearity, the value of the eigenvalues equals 1. When at least 1 eigenvalue is different from 1 or at least 1 eigenvalue is near 0, multicollinearity is proved. Examining eigenvalues separately, however, is not very meaningful. Therefore, Akdeniz and Çabuk [9] suggested a condition index based on the biggest and smallest eigenvalue. In calculating the smallest square estimators for the condition index, the eigenvalues of the used correlation matrix X'X are shown as;

$$(\lambda_{max} = \lambda_1 > \lambda_2 > ... > \lambda_p = \lambda_{min})$$

The condition index is calculated with the following formula:

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad \text{(ii)}$$

If CI < 10, there is little multicollinearity and a serious problem cannot be observed. Multicollinearity is medium-leveled in 10 ≤ CI ≤ 30, while 30 < CI indicates a severe multicollinearity and more than one multicollinearity must be assumed [14,15].

Principal Component Regression

LS estimator and multilinear regression model have the following form in matrix rotation:

$$Y = \beta X + e \quad \text{(iii)}$$

Y represents the dependent variable, X the independent variable, β the estimated coefficient and e the error in the model.

The estimator variables in the regression model have the following form after the necessary transformations in equation iv have been carried out:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad \text{(iv)}$$

In PCR, mean subtraction from both dependent and independent variables is performed and the subtraction result is divided by standard deviation. Thus dependent variables are transformed to main components for PCR analysis. This transformation is expressed mathematically as follows:

$$(X'X) = PDP' = Z'Z$$

Here X'X is the diagonal matrix of the eigenvalues, P: X'X the eigenvector matrix, and Z the data matrix. P'P = I.

After this transformation the correlation between the components is removed. The variable X is replaced with variable Z.

When PCR is applied, the smallest component that causes multicollinearity is removed. For this purpose eigenvalues are used. The model which has the smallest component near to zero among the components is removed from the system. As a result of removing this component the problem of multicollinearity is solved with the utmost probability [16].

To estimate inner egg quality variables by evaluating outer egg quality traits LS, multiregression and PCR were used; necessary calculations were performed with NCSS software [16].

RESULTS

The correlation matrix between inner and outer egg quality variables are presented in *Table 1*. The results given in *Table 1* indicated that correlation between shape index and egg length, between albumen length and yolk width were non-significant (P>0.05), apart from them they were significant (P<0.05). The Least Square Method and Principal Component Regression values are given in *Table 2*. Multicollinearity was found in the parameters $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ through LS method. However, multicollinearity was not found in PCR method.

Table 1. *Correlation matrix of dependent and independent variables*

Tablo 1. *Bağımlı ve bağımsız değişkenlerin korelasyon matrisi*

Variables	Egg Weight (X_1)	Egg Width (X_2)	Egg Length (X_3)	Shape Index (X_4)
Egg weight (X_1)	1.000			
Egg width (X_2)	0.760**	1.000		
Egg length (X_3)	0.668**	0.768**	1.000	
Shape index (X_4)	0.252**	0.494**	-0.174	1.000
Albumen height	0.600**	0.578**	0.492**	0.217*
Albumen width	0.603**	0.782**	0.749**	0.198*
Albumen length	0.625**	0.753**	0.736**	0.160
Yolk width	0.492**	0.480**	0.488**	0.075
Yolk height	0.496**	0.590**	0.442**	0.304**

*** P<0.01, * P<0.05*

***Table 2.** Least squares method and PCR multicollinearity*

***Tablo 2.** En küçük kareler tekniği ve temel bileşenler analizinde çoklu bağlantı*

Parameters	**LS**	**PCR**
	VIF	**VIF**
$\hat{\beta}_1$	2.569	2.452
$\hat{\beta}_2$	319.609 $^{\psi}$	0.682
$\hat{\beta}_3$	241.423 $^{\psi}$	0.810
$\hat{\beta}_4$	131.850 $^{\psi}$	0.708

$^{\psi}$ Since some VIF's are higher than 10, multicollinearity is a problem

***Table 3.** Eigenvalues of correlations and condition index*

***Tablo 3.** Korelasyonların özdeğerleri ve koşul indeksi*

No	**Eigenvalue**	**Condition Index**
PC1	2.544	1.00
PC2	1.161	1.48
PC3	0.294	2.94
PC4 $^{\psi}$	0.001	41.92

$^{\psi}$ Multicollinearity

Eigenvalues of correlations and condition index are given in *Table 3*. Particularly, eigenvalue and condition index values were higher in LS method than values. In PCR method, results were below the expected critical results.

Estimation equations of inner egg quality characteristics received from LS and PCR analyses and goodness of fit are presented in *Table 4*.

DISCUSSION

Table 1 shows that except for egg height-shape index all correlations among outer egg quality traits were found significantly positive (P<0.01). Particularly the correlations between egg weight and egg width (76%) and egg width and egg length (76.8%) respectively were high. The results of this study were in accordance to the results of Poyraz [17], Akbaş et al.[18] and Kul and Şeker [19]. Examining the relation between inner and outer egg quality, it is asserted that the correlations between shape index and yolk width as well as between shape index and albumen length were non-significant (P>0.05). These results were similar to the results of Alkan et al.[1]. Although the correlations between shape index, albumen height, albumen width and yolk height

***Table 4.** Estimation equations of inner egg quality characteristics received from LS and PCR analyses (standart errors in parantheses) and goodness of fit*

***Tablo 4.** En küçük kareler yöntemi ve temel bileşenler regresyon analizinden elde edilen yumurta iç kalitesinin parametre tahminleri (standart hata değerleri parantez içinde) ve uyum iyiliği*

Inner Egg Quality Variables	**Methods**	**Parameters**						
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	**R^2**	**Sig**
Albumen height	LS	-0.326	0.196 ^{NS}	0.030	0.012	0.003	0.3958	**
		(5.105)	(0.065)	(0.206)	(0.157)	(0.064)		
	PCR	-0.484	0.196	0.024	0.017	0.005	0.3958	**
		~	(0.064)	(0.010)	(0.009)	(0.005)		
Albumen width	LS	-101.895 ^{NS}	-0.086	-3.462	4.092 ^{NS}	1.422 ^{NS}	0.6825	**
		(48.997)	(0.626)	(1.974)	(1.509)	(0.619)		
	PCR	5.035	-0.379	0.864	0.786	0.068	0.6672	**
		~	(0.625)	(0.093)	(0.090)	(0.046)		
Albumen length	LS	-2.427	0.420	0.362	1.126	0.174	0.6281	**
		(57.979)	(0.740)	(2.335)	(1.785)	(0.723)		
	PCR	8.770	0.389	0.815	0.780	0.032	0.6280	**
		~	(0.723)	(0.108)	(0.103)	(0.054)		
Yolk height	LS	-2.018	0.372	-0.193	0.262	0.073	0.2924	**
		(14.824)	(0.189)	(0.597)	(0.456)	(0.187)		
	PCR	4.262	0.355	0.061	0.068	-0.609	0.2912	**
		~	(0.185)	(0.028)	(0.026)	(0.014)		
Yolk width	LS	11.706	0.432	0.656	-0.136	-0.037	0.3541	**
		(35.832)	(0.458)	(1.440)	(1.103)	(0.452)		
	PCR	2.152	0.458	0.270	0.159	0.084	0.3537	**
		~	(0.447)	(0.067)	(0.064)	(0.033)		

***Sig:** Significance level, ** P<0.01, **NS:** Non-significant , ~ Not revealed* were significant, they were found low (P<0.05). Apart from these, the correlations between inner and outer egg quality variables were found to be highly significant and at a high grade positive (P<0.01). These results were in accordance to the results of Alkan et al.[1] Narinç et al.[5] Kul and Şeker [19], Üçkardeş et al.[20].

LS method was applied in several studies, such as Alkan et al.[1] Akbaş et al.[18] and Kul and Şeker [19]. These studies report a high correlation among inner egg quality traits. However, they did not use regression estimating equations. It is likely that in these studies multicollinearity occurred. The fact that especially correlations between different outer egg quality variables appeared significant and high, is evidence to suggest that it is a result of multicollinearity [20]. The VIF values belonging to LS and PCR methods are given in *Tablo 2*. VIF values of regression parameters, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ in LS method were found to be higher than 10 of critical value. In the case of multi-relationship, sthandart errors of regression parameters were too high. Likewise, the results in *Table 4* indicated that standard errors of LS regression parameters were too high. Even some parameters which were close to zero were non-significant. These results agreed with multicollinearity problem. The VIF values were less than 10 in PCR values given in *Table 2*. According to these results, standard errors of the parameters were lower than those in LS method. PCR parameter values give in *Table 4* were lower than LS method and even non-significant parameters became significant, which are similar to the results presented by Üçkardeş et al.[20].

In *Table 3* the eigenvalues of the correlation have taken more than one different value and the 4th eigenvalue was found very high when the condition index was examined. Thus it was proved that multicollinearity is present in the values estimated with LS, which is similar to the results presented by Üçkardeş et al.[20].

As seen in *Table 4*, PC4 has a value very near to zero as a result of PCR analysis. Therefore a PCR analysis of PC4 was carried out again. The VIF values belonging to the parameters as a result of PCR analysis are given in *Table 2*. As a remedy for multicollinearity in these values PCR analysis was carried out.

The regression estimation equations, which were received from LS and PCR methods in order to determine inner egg qualities by using outer egg quality parameters, are given in *Table 4*. The regression estimation equations of all inner egg qualities were highly significant (P>0.01). However, the standard deviation of estimated parameters received from LS was higher than that of parameters received from PCR. The measures of goodness of fit of estimation equations received from each method were very near to each other. According to these results in both LS and PCR the highest resulting values were for albumen width (66.72% - 68.25%), while the lowest were for yolk height (29.12% - 29.24%).

Some parameters were non-significant in LS (P>0.05). Particularly the standart deviation for $\hat{\beta}_0$ parameters were very high. These results suggest that it is of high importance to determine inner egg quality traits without breaking the shell. It is obvious that if multicollinearity is present, principal component regression, a biased estimation method, is preferable to LS in order to estimate inner egg quality characteristics because more accurate and more reliable regression estimation equations with lower standard deviation are achieved. The review of the literature showed that there is no study that PCR was not applied in determining inner egg quality traits. As an alternative to avoid multicollinearity Üçkardeş et al.[20] have used Ridge Regression in estimating. It has been observed that their results are in accordance with the results of this study. As a result it can be asserted that in estimating inner egg quality traits by using data related to outer egg quality traits, PCR leads to more accurate and more reliable estimation equations than LS method.

Acknowledgements

The research with the project number of 2005.02.0121.005 was supported by the Akdeniz University Scientific Research Projects Management. The management and handling of the birds were performed according to the regulations required by the European Convention for the Protection of Animals kept for Farming Purposes.

REFERENCES

1. Alkan S, Karabağ K, Galiç A, Karslı T, Balcıoğlu MS: Effects of selection for body weight and egg production on egg quality traits in Japanese quails *(Coturnix coturnix japonica)* of different lines and relationships between these traits. *Kafkas Univ Vet Fak Derg*, 16 (2): 239-244, 2010.

2. Narinç D, Aksoy T, Karaman E: Genetic parameters of growth curve parameters and weekly body weights in Japanese quail *(Coturnix coturnix japonica)*. *J Anim Vet Adv*, 9 (3): 501-507, 2010.

3. Balcıoğlu MS, Yolcu Hİ, Fırat MZ, Karabağ K, Şahin E: Japon bıldırcınlarında canlı ağırlık ve canlı ağırlık artışına ait genetik parametre tahminleri. *Akdeniz Üniv Ziraat Fak Derg*, 17 (1): 81-85, 2005.

4. Narinç D, Karaman E, Aksoy T: Estimation of genetic parameters for carcass traits in Japanese quail using Bayesian methods, *S Afr J Anim Sci*, 40 (4): 342-347, 2010.

5. Narinç D, Karaman E, Firat MZ, Aksoy T: Japon bıldırcınlarında bazı yumurta verim özelliklerine ait varyans unsurlarının farklı tahmin yöntemleri kullanarak elde edilmesiyle çok özellikli genetik parametre ve BLUP tahminleri, *Kafkas Univ Vet Fak Derg*, 17 (1): 117-123, 2011.

6. Alkan S, Karabağ K, Galiç A, Karslı T, Balcıoğlu MS: Determination of body weight and some carcass traits in Japanese quails *(Coturnix coturnix japonica)* of different lines. *Kafkas Univ Vet Fak Derg*, 16 (2): 277-280, 2010.

7. Haugh RR: The Haugh Unit For Measuring Egg Quality. U. S. Egg *Poultry Magazine*. 43, 522-555, 572-573, 1937.

8. Tserveni Gousi AS, Yannakopoulos AL: Carcase characteristics of Japanese quail at 42 days of age. *British Poultry Science*, 27 (1) : 123-127, 1986.

9. Akdeniz F, Çabuk A: Ridge regresyon teorisinde 1970-2001 arasındaki gelişmeler. *V. Ulusal Ekonometri ve İstatistik Sempozyumu 20-22 Eylül, Çukurova Üniversitesi, Adana,* 2001.

10. Ortabaş N: Principal Components in The Problem of Multicollinearity. Dokuz Eylül University, Natural and Applied Science, p. 46, 2001.

11. Albayrak SA: Çoklu bağlantı halinde enküçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama. *Zonguldak Karaelmas Üniv Sosyal Bil Derg*, 1 (1): 105-126, 2005.

12. Marquardt DW: Generalized ınvers, ridge regression, biased linear estimation and non-linear estimation. *Techonometrics*, 12, 591-612, 1970.

13. Vinod HD: Double bootstrap for shrinkage estimators, *J Econometrics*, 68, 287-302, 1995.

14. Pagel MU, Lunneborg CE: Empirical evaluation of ridge regression, *Psychological Bulletin*, 97, 342-355, 1985.

15. Gujarati DN: Basic Econometrics, 3rd ed., McGraw-Hill, New York., 1995.

16. NCSS Inc: Ncss User Guide 2001, Kaysville, NCSS Inc., 2001.

17. Poyraz Ö: Kabuk kalitesi ile ilgili yumurta özellikleri arasındaki fenotipik korelasyonlar. *Lalahan Hayl Araş Enst Derg,* 29, 66-79, 1989.

18. Akbaş Y, Altan O, Koçak C: Effects of hen's age on external and internal egg quality characteristics. *Turk J Vet Anim Sci,* 20, 455-460, 1996.

19. Kul S, Şeker I: Phenotypic correlations between some external and internal egg quality traits in the Japanase quail *(Coturnix coturnix japonica). Int J Poult Sci,* 3 (6): 400-405, 2004.

20. Üçkardeş F, Narinç D, Efe E, Aksoy T: Ridge regresyon yöntemiyle Japon bıldırcını yumurtalarında ak indeksinin tahmin edilmesi. *07-09 Ekim Kümes Hayvanları Kongresi*, *Kayseri,* Bildiri Kitabı, s. 97, 2010.