Black Sea Journal of Statistics
Doi: 10.5281/zenodo.18234774

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 2: 41-46 / December 2025

PREDICTION OF ALBUMEN HEIGHT BASED ON EGG QUALITY TRAITS BY PRINCIPAL COMPONENTS REGRESSION

Seyhan Şerife IŞIK TAÇYILDIZ[1]*

[1]Kahramanmaraş Sütçü İmam University, Elbistan Vocational School, Department of Laboratory and Veterinary Health, 46300, Kahramanmaraş, Türkiye

Abstract: In this study, multiple linear regression analysis and principal component regression (PCR) method were used to predict the albumen height of Atak-S layer hen eggs. Initially, the effects of independent variables on albumen height were examined through multiple linear regression analysis, revealing that egg weight and Haugh unit had statistically significant impacts. Other factors such as egg size, shape index, and shell thickness did not have a significant effect. The explanatory power of the model was found to be high, as it explained 96 % of the total variation in albumen height. Subsequently, PCR was applied to address issues related to multicollinearity problem, and only three principal components (PC1, PC2, and PC3) were included in the model. The effects of these components on albumen height were found to be significant, with particular emphasis on the importance of morphological parameters (egg size and shell structure) in predicting internal egg quality. The PCR model demonstrated high predictive performance, accurately forecasting albumen height. In conclusion, the PCR method used in this study provided a robust model for predicting albumen height and highlighted the critical role of morphological characteristics in determining egg quality. Future studies could test the generalizability of this model using different hen breeds and larger sample sizes, as well as investigate the effects of environmental factors and feeding strategies.

Keywords: Multicollinearity problem, Principal components regression method (PCR), Egg characteristics

***Corresponding author:** Kahramanmaraş Sütçü İmam University, Elbistan Vocational School, Department of Laboratory and Veterinary Health, 46300, Kahramanmaraş, Türkiye
E mail: essi_46@hotmail.com (S. Ş. IŞIK TAÇYILDIZ)
Seyhan Şerife IŞIK TAÇYILDIZ https://orcid.org/0000-0003-3581-6635

Received: November 02, 2025
Accepted: November 28, 2025
Published: December 15, 2025

Cite as: Işık Taçyıldız, S. Ş. (2025). Prediction of albumen height based on egg quality traits by principal components regression. *Black Sea Journal of Statistics, 1*(2), 41–46.

1. Introduction

Egg quality is generally examined under two main headings: external quality (weight, shell thickness, shape index, etc.) and internal quality (album height, yolk index, etc.). Among the internal quality parameters, albumen height, in particular, stands out as a reliable biometric indicator for determining both freshness and overall quality (Silversides and Budgell, 2004).

Quantitative assessment of egg quality offers significant advantages in terms of both increasing production efficiency and ensuring consumer confidence. However, the high correlation of quality parameters often violates the fundamental assumptions of regression analyses and raises the issue of multicollinearity problem. This structural problem weakens the reliability of the estimated model and limits their generalizability (Montgomery et al., 2021). Therefore, the need to use not only observational but also statistically robust and structurally consistent modeling techniques in egg quality analyses arises.

Principal components regression (PCR), one of the method proposed to overcome multicollinearity problem, is an effective approach that increases predictive power by transforming variables through new, independent components in high dimensional and correlated data structures (Jolliffe and Cadima, 2016). This method both maintains explanatory power and strengthens the statistical stability of the model in egg quality modeling.

This study was conducted to evaluate the internal quality parameters of eggs obtained from Atak-S hens, a native breed specific to Türkiye. Such unique biometric studies, which are limited in the literature, contribute to the objective assessment of the performance of domestic genetic resources and to breeding efforts. In the study, multiple linear regression and PCR approaches were comparatively examined, especially based on the albumen height variable, and comprehensive analyses were carried out in the context of the structural reliability, explanatory power and generalizability of the model.

2. Materials and Methods

In this study, 221 egg samples obtained from Atak-S layer hens were analyzed. Measurements were made using a digital caliper with a precision of ±0.01 mm and an electronic scale with a precision of 0.01 g, and the internal structures of the eggs were examined in detail. The measurements obtained were standardized in millimeters and grams. To quantitatively assess egg quality, the following calculations (Equations 1-4) were made (Yannakopoulos and Tserveni-Gousi, 1986; Kaya and Aktan, 2011; Olawumi and Christiana, 2017):

Shape Index = (Egg Width / Egg Length) × 100 (1)

Album Index = (Album Height / Average of Albumen Width and Length) × 100 (2)

Haugh Unit = 100 × log(Album Height + 7.57 – 1.7 × Egg Weight ^ 0.37) (3)

Shell Thickness = Wide End + Narrow End + Middle Section / 3 (4)

As part of the study, a prediction model was developed based on one dependent variable (egg albumen height) and seven independent variables (egg weight, width, length, shape index, Haugh unit shell weight, and thickness). Multiple regression analysis was conducted to assess the success and fit of this model, utilizing the SPSS 26 statistical package.

Multiple regression analysis is a statistical methods that allows a dependent variable to be estimated using multiple explanatory variables. The basic equation used in this analysis is as given in Equation 5:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon \quad (5)$$

In this equation,

Y; represents the dependent variable,

β_0 represents the constant coefficient,

$\beta_1 - \beta_n$ represents the regression coefficients,

X_1–X_n represents the independent variables, and ε represents the error term (Üçkardeş et al., 2012).

To ensure the reliability of the model, it is important to examine the level of relationships between explanatory variables. A high correlation between explanatory variables can lead to the multicollinearity problem (Maxwell, 2000; Montgomery et al., 2001). Commonly used methods to identify this problem are as follows:

Correlation coefficient: A correlation between two variables of 0.75 or above indicates multicollinearity problem.

Variance Influence Factor (VIF): Measures the relationship of each independent variable with the others. A VIF value of 10 or above indicates a high risk of collinearity. Tolerance Value (TV): Defined as $1 - R^2$, small values indicate multicollinearity problem (Albayrak, 2005).

Because each of these methods provides a different perspective, using them together to understand the presence of multicollinearity problem provides more reliable results. Various solutions are proposed in the literature to mitigate the of multicollinearity problem (e.g., variable extraction, combining correlated variables, adding a new variable to the model, PCR, Ridge regression, etc.). In this study, the PCR method was chosen to solve this problem.

The PCR method is an effective approach widely used in the analysis of multidimensional and highly correlated data sets. This methods allows for the analysis of highly correlated variables by converting them into a smaller number of independent and significant new components. This reduces the correlation effect in the data set and increases the statistical accuracy of the model, resulting in more reliable and consistent predictions.

First described by Karl Pearson in 1901 and developed by Harold Hotelling in the 1930s, principal components analysis offers effective solutions, especially in high-dimensional data sets where multicollinearity problem are frequently encountered. The PCR method provides both dimensionality reduction and increases the reliability of statistical analyses in complex data structures where the number of variables is large and interrelated (Jolliffe and Cadima, 2016; Abdi and Williams, 2010). Thanks to the advantages of this methods, it is widely used in many disciplines, including animal husbandry, biology, economics, engineering, environmental sciences, and health. A review of the literature reveals that multicollinearity problem are frequently encountered, particularly in studies conducted in the field of animal husbandry, and the PCR method has been successfully applied to solve these problems. Its mathematical formula is as given in Equation 6.

$$\hat{\beta} = W(T^T T)^{-1} T^t y \quad (6)$$

here;

W: The weight matrix used to obtain the principal components from the X variables.

T=XW: The original data is represented with fewer dimensions and independent principal components.

$(T^T T)^{-1} T^t y$ Calculates the linear relationship (regression) between the principal components and y.

$\hat{\beta}$: Converts the coefficients found on the principal components to the original X variables (Jolliffe and Cadima, 2016).

3. Results and Discussion

The model equation for the multiple linear regression analysis applied to predict the albumen height of Atak-S hen eggs is as given in Equation 7:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 \quad (7)$$

here;

$\hat{Y}$ = Egg albumen height

X_1 = Egg weight

X_2 = Egg width

X_3= Egg length

X_4= Shape index

X_5= Haugh unit

X_6= Shell weight

X_7= Shell thickness

In the dataset used for this model, the overall results of the multiple linear regression analysis indicated a statistically significant probability value (P<0.001). The coefficient estimates, standard errors, t-statistics, and probability values (P) are presented in Table 1, while the correlation coefficient and coefficient of determination are provided in Table 2.

Table 1. Multiple linear regression coefficients, t-statistics, and P-values

Variables	Regression Coefficient (β)	Std. Error of the Estimate	t	P
AH	-28.007	15.93	-1.75	0.08
EW	0.04	0.01	4.53	<0.001
EWI	-0.53	0.35	-1.52	0.12
EL	0.38	0.26	1.42	0.15
SI	0.29	0.20	1.43	0.15
HU	0.13	0.002	72.57	<0.001
SW	0.06	0.04	1.35	0.17
ST	-0.001	0.001	-1.19	0.23

AH: albumen height, EW: egg weight, EWI: egg width, EL: egg length, SI: shape index, HU: Haugh unit, SW: shell weight, ST: shell thickness.

Table 2. Correlation and determination coefficients obtained by the multiple linear regression method

r	R Square	Std. Error of theEstimate
0.98	0.96	0.24

The multiple linear regression analysis conducted revealed that the model was statistically significant (F=778.76; P<0.001). The determination coefficient, which indicates the explanatory power of the model, was calculated to be 0.96. This result shows that the independent variables explain 96 % of the total variation in albumen height. A high determination coefficient value indicates that the model has a strong explanatory capacity and that the independent variables significantly account for the dependent variable.

According to the analysis findings, egg weight and Haugh unit variables had a statistically significant effect on albumen height (P<0.05). On the other hand, variables such as egg size measurements, shape index, and shell thickness did not show a significant effect on albumen height (P>0.05). This suggests that some of the variables included in the model have a limited effect on albumen height.

The correlation coefficient calculated for the overall fit of the model was found to be 0.98. This value indicates a strong linear relationship between the independent and dependent variables. However, the lack of significance for some variables and the high level of correlation between the independent variables suggest that caution should be exercised regarding the risk of multicollinearity problem. The relationships between the independent variables, as well as the VIF values and TV, were assessed and are presented in Table 3.

The examination of the correlation coefficients indicates a serious issue of multicollinearity problem within the model. In particular, the high positive correlation between egg weight and egg width (r=0.75) and length (r=0.68) demonstrates a significant linear dependency between these variables. Such relationships are considered one of the main reasons for the regression coefficients not being statistically significant.

The presence of multicollinearity problem in the model has been clearly demonstrated at the structural level through the VIF values and TV. The VIF values for variables such as shape index (VIF=845.62), egg length (VIF=787.10), and egg width (VIF=490.86) are significantly higher than the threshold value commonly accepted in the literature (VIF>10). This indicates that there is a high level of shared variance within the model, leading to a substantial reduction in the reliability of parameter estimates. Additionally, the tolerance coefficients for these variables, which are below 0.01, further support the existence of this structural problem.

In this context, to mitigate the effects of multicollinearity problem, reduce dimensionality, and improve the reliability of the regression coefficients, the PCR method was applied. This analysis was conducted using the NCSS12 statistical software. The eigenvalues, eigenvectors, prediction parameters, standard errors, VIF values, and tolerance coefficients for the PCR analysis are presented in Tables 4, 5, and 6.

Table 3. Relationship between independent variables, VIF values, and TV

	Correlation Matrix								
Variables	EW	EWI	EL	SI	HU	SW	ST	TV	VIF
EW	1							0.17	15.73
EWI	0.75**	1						0.003	490.86
EL	0.68**	0.34**	1					0.002	787.10
SI	-0.06	0.44**	-0.69**	1				0.001	845.62
HU	-0.09**	0.002	-0.11	0.11	1			0.95	1.04
SW	0.52**	0.36**	0.27**	0.02	-0.11	1		0.38	12.58
ST	0.13*	0.09	-0.05	0.12	-0.10	0.64**	1	0.50	11.97

EW: egg weight, EWI: egg width, EL: egg length, SI: shape index, HU: Haugh unit, SW: shell weight, ST: shell thickness.

Table 4. Eigenvalues of the correlation matrix obtained by the PCR method

	Eigenvalues of the Correlation Matrix		
PC	Eigenvalue	Incremental Percent	Cumulative Percent
PC1	2.63	37.69	37.69
PC2	1.76	25.27	62.96
PC3	1.29	18.43	81.39
PC4	0.91	13.13	94.52
PC5	0.26	3.82	98.34
PC6	0.11	1.65	99.99
PC7	0.05	0.01	100

Table 5. Eigenvectors obtained by the PCR method

				Eigenvectors			
Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
EW	0.56	-0.05	0.24	0.02	-0.008	0.78	-0.01
EWI	0.46	0.26	0.45	0.12	0.21	-0.46	-0.48
EL	0.43	-0.52	0.06	-0.06	0.20	-0.35	0.60
SI	-0.05	0.69	0.28	0.15	-0.03	0.0002	0.63
HU	-0.10	0.07	0.31	-0.93	0.04	0.01	0.001
SW	0.45	0.20	-0.38	-0.20	-0.71	-0.19	-0.0001
ST	0.23	0.34	-0.62	-0.18	0.62	0.05	0.003

EW: egg weight, EWI: egg width, EL: egg length, SI: shape index, HU: Haugh unit, SW: shell weight, ST: shell thickness.

Table 6. Prediction parameters, standard errors, VIF values, and TV obtained by the PCR method

	Regression Coefficient (β)	Std. Error of the Estimate	t	VIF	TV
C	-4.94	-	-	-	-
PC1	0.05	0.01	0.15	1.01	0.99
PC2	-0.02	0.02	-0.02	1.01	0.99
PC3	-0.007	0.01	-0.01	1.17	0.85
PC4	-0.004	0.004	-0.008	1.12	0.89
PC5	0.13	0.001	0.98	1.01	0.99
PC6	0.06	0.04	0.02	1.01	0.99
PC7	-0.001	0.0009	-0.02	1.01	0.99

In the PCR analysis, the eigenvalues of the correlation matrix were examined, and it was observed that only the first three components had eigenvalues greater than 1. This finding indicates that it is appropriate to include only PC1, PC2, and PC3 in the model. These three components explain 81% of the total variance and represent a significant portion of the information contained in the dataset. Specifically, PC1 explains 37% of the total variance, PC2 explains 25%, and PC3 explains 18%.

The content analysis of the components was based on the eigenvector loadings. PC1 is strongly associated with variables such as egg weight (EW), egg width (EWI), shell weight (SW) and egg length (EL). The PC2 component shows strong relationships with shape index (SI), egg length (EL), shell thickness (ST) and egg width (EWI). PC3 is highly correlated with shell thickness (ST), egg width (EWI), shell weight (SW), and Haugh unit (HU).

The VIF values for all components remain well below 2, confirming that there are no issues with multicollinearity problem. Additionally, the TV further support the reliability of the model.

As a result, the model formed by using only the components PC1, PC2, and PC3, which have eigenvalues greater than 1 and high statistical significance, will be both statistically robust and consistent in terms of interpretability. The model formulated with these components can be expressed as given in Equation 6.

$$Y= -4.94+0.05PC1-0.02PC2-0.007PC3 \quad (8)$$

Upon examining Tables 1 and 2 it was observed that most of the parameters in the predicted regression model were not statistically significant (P>0.05). This finding is consistent with studies conducted by Aktan (2004) and Akçay and Sarıözkan (2015) in the literature.

The data presented in Table 3 reveal important findings regarding the relationships between the independent variables. The relationship between egg weight and both egg length and width was found to be positively significant, with correlations of 75% and 68%, respectively. Additionally, a negative and significant relationship of 69% was observed between egg length and shape index (P<0.05). These findings align with the studies of Duman et al. (2016), Okur et al. (2018), Uçar and Kahya (2020), Vekić et al. (2022), and Kurşun et al. (2024).

In Tables 4, 5, and 6, the findings related to the PCR method, used to address the issue of multicollinearity problem, are consistent with the works of Adenaike et al. (2015), Shafey et al. (2015), Sarı et al. (2016), Çankaya et al. (2019), Tırınk et al. (2020), Kebede et al. (2022), Gök and Kurşun (2025), and Kurşun and Gök (2025).

4. Conclusion

The multiple linear regression analysis and PCR method applied in this study to predict the albumen height of Atak-S hen eggs demonstrated that the model is robust and reliable. Multiple linear regression analysis revealed that egg weight and the Haugh unit had significant effects on albumen height, while variables such as egg size, shape index, and shell thickness did not have a significant impact. In the PCR analysis, only the components PC1, PC2, and PC3 were included in the model, and the effects of these components on albumen height were found to be significant. Notably, the relationship between PC1 and PC2 with morphological characteristics plays an important role in predicting egg albumen height. The predicted albumen height value from the model is consistent with the sample mean and the confidence interval, indicating high predictive accuracy.

These findings highlight that, in predicting albumen height, not only internal quality indicators but also morphological parameters such as the egg's physical dimensions and shell structure must be considered. Future studies can test the generalizability of PCR based modeling by applying the method to different chicken breeds and larger sample sizes. Furthermore, investigating the effects of environmental factors and feeding strategies on egg quality could contribute to the more effective optimization of production strategies. This study provides valuable contributions to the literature in terms of both methodological approach and quality parameters.

Author Contributions

The percentages of author' contributions are presented below. The author reviewed and approved the final version of the manuscript.

	S.Ş.I.T.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declares that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433–459.

Adenaike, A. S., Akpan, U., & Ikeobi, C. O. N. (2015). Principal component regression of body measurements in five strains of locally adapted chickens in Nigeria. *Thai Journal of Agricultural Science, 48*(4), 217–225.

Akçay, A., & Sarıözkan, S. (2015). Estimation of income in egg production using Ridge Regression analysis. *Ankara Üniversitesi Veteriner Fakültesi Dergisi, 62*(1), 69–74.

Aktan, S. (2004). Determining some exterior and interior quality traits of quail eggs and phenotypic correlations by digital image analysis. *Journal of Animal Production, 45*(1), 7–13.

Albayrak, S. A. (2005). Biased estimation techniques as an alternative to the least squares method in the presence of multicollinearity and an application. *International Journal of Management and Economics Business, 1*(1), 105–126.

Çankaya, S., Eker, S., & Abacı, Ş. H. (2019). Comparison of least squares, ridge regression and principal component approaches in the presence of multicollinearity in regression analysis. *Turkish Journal of Agricultural Food Science and Technology, 7*(8), 1166–1172.

Duman, M., Şekeroğlu, A., Yıldırım, A., & Eleroğlu, H. (2016). Relation between egg shape index and egg quality characteristics. *European Poultry Science, 80*, 1–9.

Gök, İ., & Kurşun, K. (2025). Prediction model of albumen index and height in Japanese quail eggs via external quality characteristics. *International Journal of Agriculture, Environment and Food Sciences, 9*(2), 493–501.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202.

Kaya, E., & Aktan, S. (2011). Flock age and hatching egg storage duration in Japanese quails: 1. Effects on dark eggshell traits. *Süleyman Demirel Üniversitesi Ziraat Fakültesi Dergisi, 6*(2), 30–38.

Kebede, K., Getachew, A., & Urge, M. (2022). Principal components regression of internal egg quality traits in two exotic chicken breeds in Haramaya. *Journal of Food Chemistry and Nanotechnology, 8*(3), 102–107.

Kurşun, K., & Gök, İ. (2025). Prediction model of albumen index in duck eggs via external egg quality characteristics in case of multicollinearity. *Black Sea Journal of Agriculture, 8*(4), 525–532. https://doi.org/10.47115/bsagriculture.1655436

Kurşun, K., Çelik Güney, M., & Baylan, M. (2024). Comparison of internal and external egg quality traits in domestic and foreign layer chicken hybrids (Atabey, Decalp, and Nick). *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 7*(1), 141–149.

Maxwell, S. E. (2000). Sample size in multiple regression analysis. *Psychological Methods, 5*(4), 434–458. https://doi.org/10.1037/1082-989X.5.4.434

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis* (3rd ed.). John Wiley & Sons.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Wiley.

Okur, N., Eratalar, S. A., & Yaman, A. (2018). Relationships among some quality characteristics in broiler hatching eggs. *Turkish Journal of Agricultural and Natural Sciences, 5*(3), 298–302. https://doi.org/10.30910/turkjans.448368

Olawumi, S., & Chiristiana, B. (2017). Phenotypic correlations between external and internal egg quality traits of *Coturnix* quails are under intensive housing system. *Journal of Applied Life Sciences International, 12*(3), 1–6.

Sarı, M., Tilki, M., & Saatçi, M. (2016). Genetic parameters of egg quality traits in long-term pedigree recorded Japanese quail. *Poultry Science, 95*(8), 1743–1749. https://doi.org/10.3382/ps/pew118

Shafey, T. M., Elsayed, O. S. H., Mahmoud, A. H., & Abouheif, M. A. (2015). Managing colllinearity in modeling the effect of age in the prediction of egg components of laying hens using stepwise and Ridge regression analysis. *Brazilian Journal of Poultry Science, 17*(4), 473–482.

Silversides, F. G., & Budgell, K. (2004). The relationships among measures of egg albumen height, pH, and whipping volume. *Poultry Science, 83*(10), 1619–1623. https://doi.org/10.1093/ps/83.10.1619

Tırınk, C., Abacı, Ş. H., & Önder, H. (2020). Comparison of Ridge regression and Least Squares Methods in the presence of multicollinearity for body measurements in Saanen kids. *Journal of the Institute of Science and Technology, 10*(8), 1429–1437. https://doi.org/10.21597/jist.671662

Uçar, A., & Kahya, Y. (2020). The comparison of weight and shape related traits in eggs from different chicken genotypes. *Turkish Journal of Agriculture Food Science and Technology, 12*(12), 2571–2578.

Üçkardeş, F., Efe, E., Narinç, D., & Aksoy, T. (2012). Estimation of albumen index in Japanese quails using the Ridge regression method. *Academic Journal of Agriculture, 1*(1), 11–20.

Vekić, M., Savić, Đ., & Jotanović, S. (2022). Phenotypic correlations between egg quality traits amid the laying phase of broiler breeder hens. *Contemporary Agriculture, 71*(1–2), 13–19. https://doi.org/10.2478/contagri-2022-0003

Yannakopoulos, A. L., & Tserveni-Gousi, A. S. (1986). Quality characteristics of quail eggs. *British Poultry Science, 27*(2), 171–176.