Black Sea Journal of Statistics

Volume 1 | Issue 2

ISSN: 3108-4265

KAR-YAY

BLACK SEA JOURNAL OF STATISTICS

(BSJ STAT)

Black Sea Journal of Statistics (BSJ Stat) is an international a peer-reviewed journal published electronically twice a year in June and December. BSJ Stat is an international e-journal that publishes open access. The general aim of BSJ Stat is to provide a platform for statistical scientists and related researchers to publish their research findings on a global scale, to encourage their professional development and to provide new perspectives for further research. BSJ Stat publishes original research articles, reviews and technical notes produced from studies carried out in the fields of theoretical and applied statistics. BSJ Stat is published electronically and all articles are provided to the readers free of charge.

e-ISSN: 3108-4265

Phone: +90 362 408 25 15

Fax: +90 362 408 25 15

Email: bsjstat@blackseapublishers.com

Web site: http://blackseapublishers.online/index.php/stat

Sort of publication: Periodically 2 times (June and December) in a year

Publication date and place: December 15, 2025- Samsun, TÜRKİYE

Publishing kind: Electronically

OWNER

KARYAY Karadeniz Yayımcılık ve Organizasyon Ticaret Limited Company

DIRECTOR IN CHARGE

Prof. Dr. Hasan ÖNDER

EDITOR BOARDS

REVIEWERS OF THE ISSUE*

Prof. Dr. Çiğdem TAKMA, Ege University, Department of Animal Science, Biometry, TÜRKİYE

Prof. Dr. Mehmet TOPAL, Amasya University, Department of Biostatistics, Biostatistics, TÜRKİYE

Assoc. Prof. Dr. Aycan Mutlu YAĞANOĞLU, Atatürk University, Department of Animal Science, Biometry, TÜRKİYE

Assoc. Prof. Dr. Yasin ALTAY, Eskişehir Osmangazi University, Department of Animal Science, Biometry, TÜRKİYE

Assist. Prof. Dr. Esra YAVUZ, Şırnak University, Department of Accounting and Tax, Biometry, TÜRKİYE

Dr. Ebru ERGÜNEŞ BERKİN, Hassa District Directorate of Agriculture and Forestry, Biometry, TÜRKİYE

Dr. Hasan Alp ŞAHİN, Ondokuz Mayıs University, Hemp Research Institute, Biometry, TÜRKİYE

---

* The ranking is arranged alphabetically within the academic title

Volume 1, Issue 2 (2025)

Table of Contents

Black Sea Journal of Statistics

Doi: 10.5281/zenodo.18234659

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 2: 27-30 / December 2025

ESTIMATING THE NONPARAMETRIC CONFIDENCE INTERVAL FOR CORRELATION COEFFICIENT ON ANIMAL DATA

Nursen KURDAL[1], Hasan ÖNDER[2]*

[1]*Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye*

Abstract: Correlation coefficient is widely used in the all areas of science to establish the degree and direction of two variables. Confidence interval is a special form of estimating a certain parameter. With use of this method, a whole interval of acceptable values for the parameter is given instead of a single value, together with a likelihood that the real (unknown) value of the parameter will be in the interval. The confidence interval is based on the observations from a sample, and hence differs from sample to sample. In this study, nonparametric confidence interval estimation for Pearson correlation coefficient were shown using an animal data set.

Keywords: Correlation coefficient, Nonparametric confidence interval, Animal

***Corresponding author:** Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye
E mail: hasanonder@gmail.com (H. ÖNDER)
Nursen KURDAL https://orcid.org/0000-0001-7198-2329
Hasan ÖNDER https://orcid.org/0000-0002-8404-8700

Received: June 26, 2025
Accepted: November 22, 2025
Published: December 15, 2025

Cite as: Kurdal, N., & Önder, H. (2025). Estimating the nonparametric confidence interval for correlation coefficient on animal data. *Black Sea Journal of Statistics, 1*(2), 27–30.

1. Introduction

Correlation is a statistical method that reveals the direction and degree of the relationship between variables. The measure of the relationship between the two variables is called the correlation coefficient (Arkin and Colton, 1939). The correlation coefficient is denoted by the small "*r*" and takes the value between -1 and +1 ($-1 \leq r \leq +1$). If *r* value takes values close to -1, it is determined that there is a negative relationship between variables, and if it takes values close to + 1, there is a positive relationship between variables. If the *r* value is close to 0, it means that there is no relationship between the two variables (Figure 1) (Kurdal and Önder, 2020).

In Figure 1; (a) the decrease in the other depending on the increase of one of the variables is a linear relationship, (b) there is no relationship between the two variables and (c) the increase in one of the variables due to the increase in the other is the linear relationship.

General comments regarding the strength of the correlation coefficient are given below (Köse, 2025):

- 0.00-0.25 Very poor relationship,
- 0.26-0.49 Weak relationship,
- 0.50-0.69 Moderate relationship,
- 0.70-0.89 High relationship,
- 0.90-1.00 Very high relationship.

The Pearson correlation coefficient can be calculated using the estimator (Equation 1):

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \quad (1)$$

Correlation coefficients vary depending on the characteristics of the variables under investigation.

Correlation coefficients used to determine the relationship between classifiable qualitative variables:

- Phi coefficient,
- Cramer V coefficient,
- Ordinary coefficient,
- Lambda coefficient.

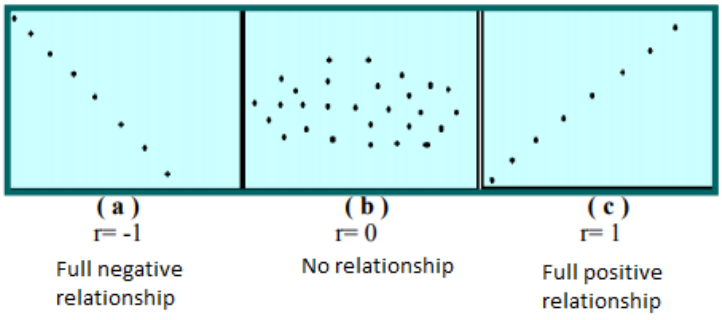


Figure 1. Scatter plots of extreme correlation coefficients.

Correlation coefficients used to determine the relationship between sortable qualitative variables:

- Spearman correlation coefficient,
- Gamma coefficient,
- Kendall's tau-b coefficient,
- Kendal's tau-c coefficient,
- Somer's d coefficient.

Correlation coefficients used in determining the relationship between discrete / continuous qualitative variables:

- Pearson correlation coefficient (if both variables show normal distribution),
- Spearman correlation coefficient (if at least one of the variables is not normally distributed).

Correlation coefficients used to determine the relationship between a classifiable qualitative variable and a discrete / continuous quantitative variable:

- Double series correlation coefficients,
- Point double series correlation coefficients.

Correlation coefficient used to determine the relationship between a sortable qualitative variable and a discrete / continuous quantitative variable:

- Multiple series correlation coefficient (Köse, 2025).

For the animal science as in all the other branches of the science the correlation coefficient (CC) is a basic relation statistics. In many studies confidence interval of the correlation coefficient is ignored. But interval estimation is very essential for statistical prediction.

Recent studies have shown that imprecise estimates of correlation coefficients can result in a series of problems, including an increase in multicollinearity in multivariable regression analyses, as well as overestimating direct effects and increasing noise in path analysis. In this sense, it is essential for experimental planning to ensure sufficient n to estimate correlation coefficients with an acceptable level of precision (Olivoto et al., 2018).

In this study, nonparametric and parametric 95% confidence interval (CI) for correlation coefficient were compared with using goat kid growth data.

2. Materials and Methods

This study was carried out at the private dairy goat farm in Bafra province of Samsun, Turkey (40°31'N, 36°53'E and 650 m above the sea level). Data was collected from 82 Saanen kids from birth (W0) to six month of age (W6).

The well-known parametric confidence interval of Pearson correlation can be calculated as given in Equation 2 (Mudelsee, 2003):

$$CI = t_{n-2,\,\alpha/2}.S_r$$

$$S_r = \sqrt{\frac{1-r^2}{n-2}} \qquad (2)$$

The nonparametric confidence interval of Pearson correlation can be calculated as given in Equation 3 (Olivoto et al., 2018):

$$CI = 0.45304^{r} * 2.25152 * n^{-0.50089} \qquad (3)$$

3. Results and Discussion

From birth (W0) to six month of age (W6) of Saanen kids monthly live weight, the minimum correlation coefficient was obtained as 0.7239 between W0 and W3, and the maximum was obtained as 0.9902 between W4 and W5 (Table 1).

The correlation coefficients showed high relations as expected for live weights. To demonstrate the confidence intervals correlation coefficients was sorted ascending. Correlation coefficients and its nonparametric 95%confidence intervals were given in Figure 2. Correlation coefficients and its parametric 95% confidence intervals were given in Figure 3.

The range of nonparametric confidence intervals was 0.026542 when the range of parametric confidence intervals was 0.102553. It means that nonparametric confidence intervals was so close to each other when parametric confidence intervals was not. With the increasing correlation coefficient parametric confidence intervals decreased but nonparametric confidence intervals. Its underlying reason is the estimator of parametric CI when the CC increases the CI decreases because of the standard error of correlation coefficient as seen in Figure 3.

Table1. Pearson correlation coefficients for monthly live weights of Saanen kids

	W1	W2	W3	W4	W5	W6
W0	0.8527	0.7908	0.7239	0.7558	0.7641	0.7588
W1		0.9705	0.9402	0.9112	0.8867	0.8600
W2			0.9827	0.9599	0.9317	0.8972
W3				0.9671	0.9356	0.8948
W4					0.9902	0.9686
W5						0.9842

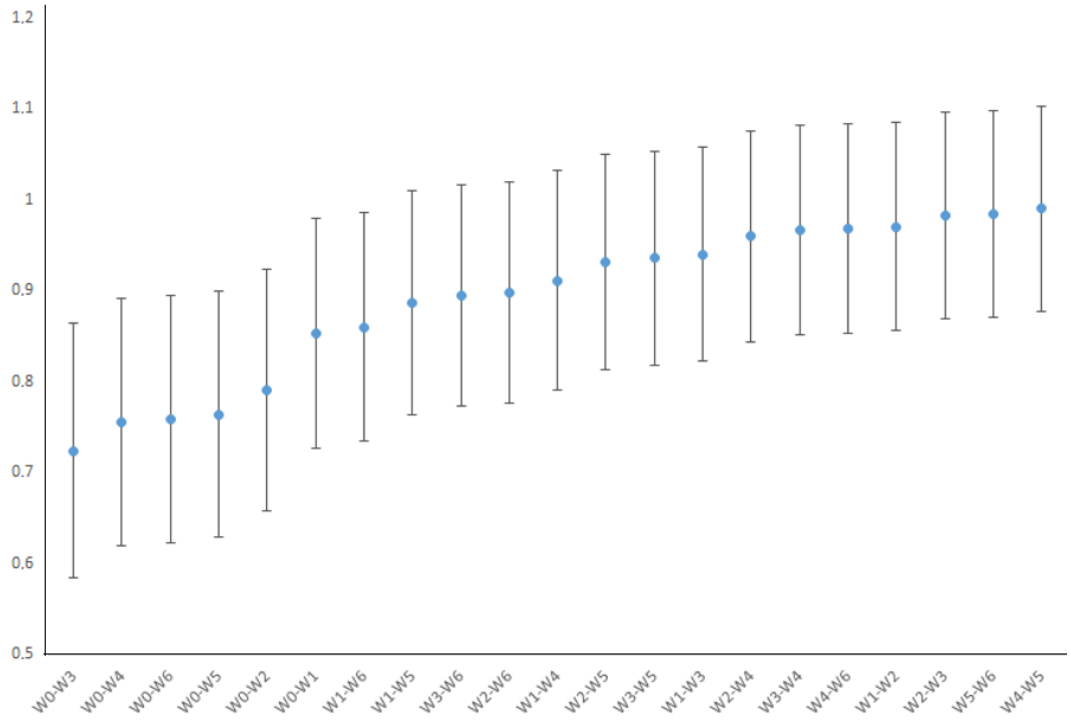


Figure 2. Correlation coefficients and its nonparametric 95%confidence intervals.

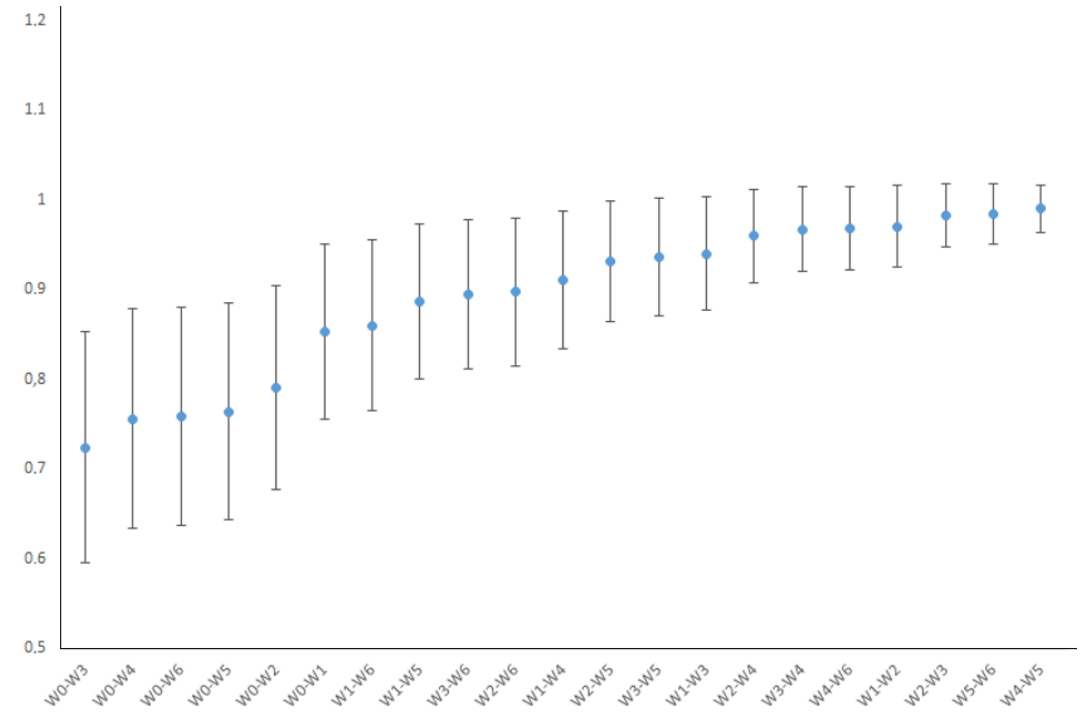


Figure 3. Correlation coefficients and its parametric 95%confidence intervals.

4. Conclusion

The nonparametric confidence interval for Pearson correlation coefficient is more reliable than parametric confidence interval. Use of nonparametric confidence interval can be preferred instead of parametric confidence interval for animal studies.

Author Contributions

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

	N.K.	H.Ö.
C	70	30
D	70	30
S	70	30
DCP	70	30
DAI	70	30
L	70	30
W	70	30
CR	70	30
SR	70	30
PM	70	30
FA	70	30

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 authors declare 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.

Acknowledgements

This study has been published in the proceeding book of the IV. International Congress on Domestic Animal Breeding, Genetics and Husbandry-2020 (ICABGEH-20), Online, August 12-14, 2020.

References

Arkin, H., & Colton, R. R. (1939). *Statistical methods*. Barnes & Noble Inc.

Köse, K. (2025). *Correlation and regression analysis* (Sunum slaytları). Toraks Derneği. Erişim adresi: https://file.toraks.org.tr/TORAKSFD23NJKL4NJ4H3BG3JH/mse-ppt-pdf/Kenan_KOSE3.pdf (Alıntı tarihi: 5 Nisan 2020)

Kurdal, N., & Önder, H. (2020). Estimating the nonparametric confidence interval for correlation coefficient on animal data. In *International Congress on Animal Breeding, Genetics and and Husbandry (ICABGEH-20)* (pp. 89–92).

Mudelsee, M. (2003). Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Mathematical Geology, 35*(6), 651–665.

Olivoto, T., Lúcio, A. D. C., Souza, V. Q., Nardino, M., Diel, M. I., Sari, B. G., Krysczun, D. K., Meira, D., & Meier, C. (2018). Confidence interval width for Pearson's correlation coefficient: a Gaussian-independent estimator based on sample size and strength of association. *Agronomy Journal, 110*(1), 1–8. https://doi.org/10.2134/agronj2017.09.0566

Black Sea Journal of Statistics

Doi: 10.5281/zenodo.18234708

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 2: 31-35 / December 2025

CURVE MODELLING PROCEDURE IN SAS PACKAGE PROGRAM

Mustafa ŞAHİN[1]*, İsmail GÖK[1], Esra YAVUZ[2]

[1]*Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Departman of Agricultural Biotechnology, 46100, Kahramanmaraş, Türkiye*

[2]*Şırnak University, Cizre Vocational School, Department of Accounting and Tax, 73200, Cizre, Şırnak, Türkiye*

Abstract: Growth, lactation, and yield curves are widely used in animal husbandry. In addition to pedigree records, they are widely used as a supporting tool in animal breeding. Curve models based on the herd average provide an indication of the herd's overall structure, while individual models provide insight into individual development. In this context, they are used as a supporting tool, particularly in the selection and culling stages. Although curve models are widely used, the number of programs available in this field is quite limited. Among these limited programs, the SAS statistical package stands out, as it offers sufficient user intervention and is widely used. This study will cover the basic steps of the curve modeling procedure (data definition, data input, model, curve, and curve feature definition). For this purpose, the curve modeling procedure in the SAS package will be discussed in detail using a numerical example related to animal husbandry, ultimately providing an important resource for researchers working in this field.

Keywords: SAS, Curve, Procedure

***Corresponding author:** Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Departman of Agricultural Biotechnology, 46100, Kahramanmaraş, Türkiye

E mail: ms66@ksu.edu.tr (M. ŞAHİN)

Mustafa ŞAHİN https://orcid.org/0000-0003-3622-4543

İsmail GÖK https://orcid.org/0000-0002-0759-1187

Esra YAVUZ https://orcid.org/0000-0002-5589-297X

Received: November 03, 2025

Accepted: November 25, 2025

Published: December 15, 2025

Cite as: Şahin, M., Gök, İ., & Yavuz, E. (2025). Curve modelling procedure in SAS package program. *Black Sea Journal of Statistics, 1*(2), 31–35.

1. Introduction

Curve modeling is a modeling technique that fundamentally enables the mathematical representation of geometric shapes and processes. Widely used in computer-aided design, engineering, and graphics applications, this method has also gained significant ground in the modeling of biological systems and the digitalization of agricultural activities in recent years. As the livestock sector embraces data-driven decision-making processes with the digital transformation, mathematical methods such as curve modeling have become increasingly functional in this field.

In animal husbandry, curve modeling can be used for various purposes, such as generating animal growth curves, analyzing milk yield over time, and examining the relationships between feed intake and weight gain. In this context, parametric curve models (e.g., Gompertz, Logistic, and Richards curves) are widely preferred to mathematically describe the biological development processes of animals. Furthermore, these models offer an important tool for making predictions to increase production efficiency, optimizing feeding strategies, and supporting economic decision-making processes (Bayazıt at al., 2022).

SAS software, widely used in curve modeling, is a powerful statistical analysis platform. In particular, procedures such as PROC NLIN, PROC NLMIXED, PROC GAM, PROC TRANSREG, and PROC SGPLOT can address a wide variety of curve modeling needs, from nonlinear regression analyses to generalized add-on models. Using these SAS procedures, users can define parametric or nonparametric curves, evaluate model fit, and make predictions.

This study will introduce the basic principles of curve modeling in the SAS environment and conduct curve modeling on a sample model using SAS procedures.

2. Materials and Methods

2.1. Materials

Data obtained from the forage legume clover using the in vitro gas production technique were used in the study. Gas production values (ml/1 g DM) were obtained at 3, 6, 12, 24, 48, 72, and 96 hours using (Makkar and Blümmel, 1997). The in vitro gas production method and are given in Table 1. The SAS statistical package program was used in the study (SAS, 1999; Goodwin at al., 2018).

Table 1. Gas production values of clover legume forage plant (ml/1 g DM)

Time	Gas Production
3	118.26
6	180.83
12	241.08
24	304.13
48	356.83
72	386.85
96	397.59

2.2. Methods

In this study, a numerical example from the field of animal nutrition is used to illustrate the curve modeling procedure in the SAS statistical package. The Orskov model (Orskov and McDonald, 1979; dos Santos Cabral at al., 2019; Panik, 2010; Panik, 2013), one of the first equations developed for estimating feed digestibility, is used in the modeling of this numerical example. This model is given in the following equation (1).

$$Y = a + b\,(1 - e^{-ct}) \quad (1)$$

In the equation,

a: represents the initial amount of gas or the amount of digestion,

b: represents the slowly produced gas or the amount of digestion,

c: represents the gas production rate or the rate of digestion,

t: represents the incubation time, and

Y: represents the gas production at the t-th time. Here, "a+b" represents the total amount of gas produced throughout digestion (Orskov and McDonald, 1979; Özkan and Kamalak, 2023).

Implementing any model in the SAS statistical package requires various steps. These include variable definition, data input, statistical model definition, and graphical definition (Beyazıt at al., 2022).

2.2.1. Variable definition

This first stage consists of defining the data set name, data type, data input order, data input method, data precalculation, and data label definition. Variable definition is outlined below.

Data ...;
İnput ... , ... , ... , ...;
Title "……..";
IF/Then/Else;
Drop / Keep;
Rename;
Label;
Cards; or Datalines;

Data: This command is used to name the data set.

Input: This command assigns data to variables. At least one space must be created between variable names. Variables are of two types: numeric and non-numeric. Non-numeric variables must be prefixed with a "$" sign. Variables without this character will be considered numeric variables. Placing the @@ symbol at the end of the input command line indicates that data input is made side by side according to the variable definition.

Title: This command determines the title of the SAS output report.

IF/Then/Else: This command sequence is used to define conditional operations before analysis.

Drop/Keep: This command is used to determine which variables in the data set will be kept or discarded.

Rename: This command is used to change the previously specified variable name.

Label: This command is used to define descriptive labels for variables.

Cards or datalines: This command indicates that the variable definition phase is complete and that data input will begin.

Data input:

This stage consists of input data in accordance with the definitions made in the input variable definition stage. Data input must be carried out in accordance with the variable definition stage. In other words, the data input stage is entirely dependent on the variable definition stage (Korosteleva, 2018).

Statistical model definition:

This stage defines the equations to be used in curve modeling. Symbols such as =, *, /, –, +, and ^ are commonly used in defining equations. Correct model definition is crucial for obtaining statistically significant curves. Furthermore, obtaining accurate outputs such as model parameter estimates and error terms depends on the correct definition of the model. The main headings are briefly outlined below.

Proc nlin; Starts the nonlinear regression procedure. In other words, if the relationship between the dependent and independent variables is nonlinear, that is, if the equation cannot be expressed using linear regression, the "nlin" procedure expresses these relationships using nonlinear mathematical models and performs parameter estimation.

Title "...."; Adds a title to the output.

Parameters a=... b=...; Provides initial estimates for the model's parameters. Appropriate values must be provided for parameter estimation and are mandatory.

Model y = a + b*...; The model's mathematical equation is defined here.

Output out=new p=Model1 r=e; The results are saved to the "new" data set. The estimated "Y" values are placed in "Model1", and the error terms of the estimated model are stored in "e1".

Run; Enables the execution of the procedure described above. Runs the procedure.

Graph (curve and curve properties) definition:

This is the stage in which the curve graph for the estimated model is drawn during the statistical model definition phase. In this stage, the graphical features related to the graph, such as line shape, color, size, axis labels, and positions, are defined. Briefly, it consists of the main headings listed below.

legend1; This is the section where settings related to the legend box within the chart are made. The relevant settings are provided under the following headings.

label = none ;
value = (j = … "Actual data" j = … "Estimated");
mode = protect ; t

position=(right inside middle) ; General graphic settings such as font, symbol, text color, and weight are made in this section. Related settings are provided under the following headings.

cborder=… ;

cshadow=…;
across=…;
shape=symbol(…,….);;
goptions ftext='…..' htext=…. gunit=pct ctext=….. csymbol=; In this section, how the graphic outputs will be drawn (visual settings such as font, size, color) are defined.
symbol1 i=…c=…. v=…; This section defines the symbols and lines for the actual data.
symbol2 i=…c=…v=…l=…; This section defines the symbols and lines for the model prediction curve.
axis1 and axis2 label=(angle=… rotate=… '……..') minor=….; In this section, the necessary definitions for vertical and horizontal axis lines are made.
Proc gplot data=new; Used to display the relationship between variables as a line and scatter plot. For the model defined below, the subcodes that plot the line and scatter plots on top of each other are provided.
plot y*t=1 model1*t=2 / overlay
legend=legend1
vaxis=axis1
haxis=axis2
frame
cframe=white;
run; quit;

3. Results and Discussion

The curve modeling steps and relevant codes of the SAS statistical package program for the gas production values of the legume forage plant in Table 1 are given in Table 2.

Table 2. Curve modeling stages and codes

VD	data; input t y; cards
DI	3 118.26 6 180.83 12 241.08 24 304.13 48 356.83 72 386.85 96 397.59 ;

SMD	proc nlin;
	title "orskov";
	parameters a=82.5 b=305 c=0.05;
	model y=a+b*(1-2.7182**(-c*t));
	output out=new p=MODEL1 r=e1; run;

DF	proc print; run;
	legend1 label=none value=
	(j=left "Gas Production" j=left "Orskov model")
	mode=protect position=(right inside middle)
	cborder=white cshadow=white across=1
	shape=symbol(6,2.5);
	position=center value=(justify=center);
	goptions ftext='Arial' htext=2.5 gunit=pct
	ctext=black csymbol=star;
	symbol1 i=none c=black v=star;
	symbol2 i=spline c=red v=none l=1;
	axis1 label=(angle=90 rotate=0 'Trifolium Gas Production Values') minor=none;
	axis2 label=('Measurement Times') minor=none;
	proc gplot ;plot y*t=1 MODEL1*t=2; /frame cframe=white legend=legend1
	vaxis=axis1 haxis=axis2 overlay ; run;

VD: variable definition, DI: data input, SMD: statistical model definition, DF: defining graphs (curves and curve properties).

When the program codes given in Table 2 were run in the SAS statistical package program, the model variance analysis results (Table 3), model parameter estimates (Table 4) and the Orskov gas production curve of the data set (Figure 1) were obtained as follows. The coefficient estimates obtained through iteration using the Gauss-Newton method (Stroup at al., 2009) are given in Table 5.

Table 3. Model variance analysis results

Source	DF	SS	MS	F	Pr > F
Model	2	68560.5	34280.2	234.46	<.0001
Error	4	584.8	146.2		
Corrected Total	6	69145.3			

SS: Sum of Squares, MS: Mean Square.

Table 4. Model parameter estimates

P	E	AE	ACL
a	82.6964	15.9795	38.3303-127.1
b	308.0	15.4298	265.1-350.8
c	0.0556	0.00770	0.0342-0.0769

P= parameter, E= estimate, AE= approximate standard eror, ACL= approximate 95 % confidence limits.

Table 5. Iterative phase for coefficient estimates (Gauss-Newton)

Iter	a	b	c	SS
0	82.5000	305.0	0.0500	1155.8
1	84.5951	306.6	0.0544	588.2
2	83.0422	307.8	0.0553	584.9
3	82.7581	308.0	0.0555	584.8
4	82.7074	308.0	0.0556	584.8
5	82.6983	308.0	0.0556	584.8
6	82.6967	308.0	0.0556	584.8
7	82.6964	308.0	0.0556	584.8

SS= Sum of Squares, IP= iterative phase.

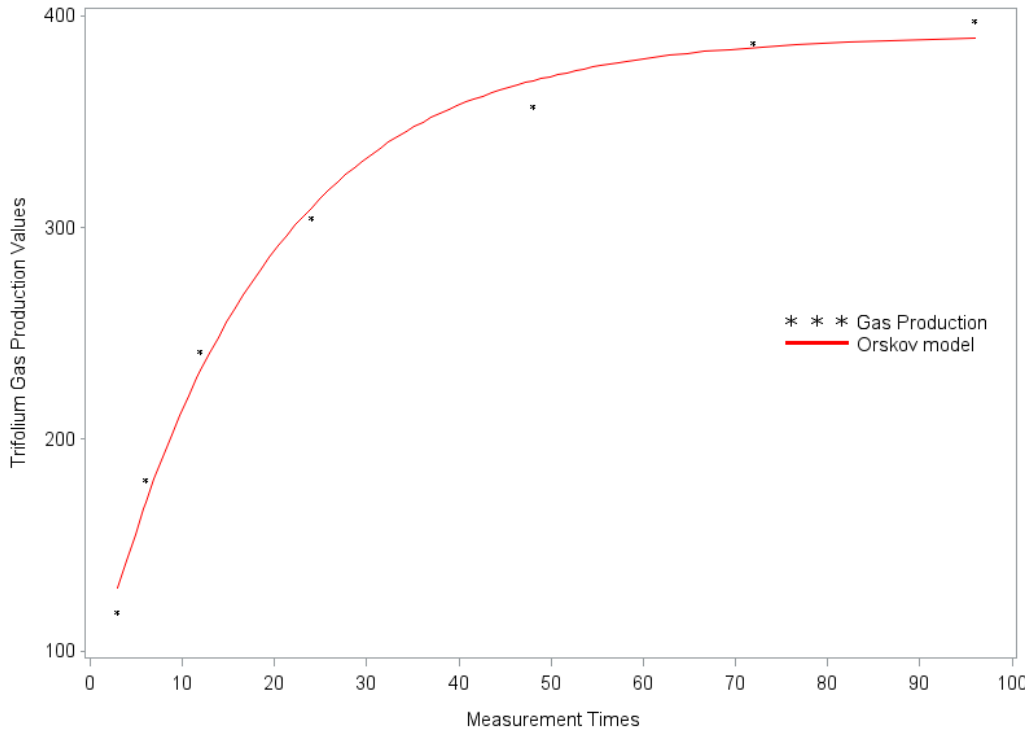


Figure 1. Orskov gas production curve of a forage legume.

When the variance analysis results given in Table 3 are examined, it is seen that the model is found to be statistically significant (P<0.0001). When the results given in Table 4 are examined, it is seen that the amount of gas obtained initially is 82.6964, the amount of gas obtained slowly is 308.0, and the gas production rate or digestion rate is 0.0556. If the initial iteration values for the amount of gas obtained initially, the amount of gas obtained slowly, and the gas production rate are given accurately, the t coefficient estimates can be obtained in the 7th iteration, as seen in Table 5. This clearly demonstrates the importance of the initial iteration values. When Figure 1 is examined, it is seen that the observation values are distributed around the curve obtained with the Orskov model. The coefficient of determination calculated from Table 3 is 0.991 (R^2=1-(584.8/69145.3)). This value is an indication that the created model can represent the point distribution at a high level.

4. Conclusion

In this study, curve modeling analyses based on the Orskov and McDonald model were successfully performed using SAS software. The findings demonstrated that the model is a powerful tool for understanding biological processes and provides a high degree of fit to the data.

The basis for this success lies not only in the selection of the correct model but also in the accurate, careful, and deliberate execution of SAS coding. Especially in nonlinear regression analyses, the correct definition of the model form, the appropriate entry of initial values, the clear definition of parameter constraints, and the correct interpretation of the outputs are crucial. Thanks to the flexible structure offered by SAS, these processes could be carried out systematically and reliable results were achieved.

Accuracy and care are essential in statistical analyses performed with SAS, as coding errors or minor inaccuracies in model definition can completely invalidate the results. Properly written SAS codes not only ensured the accuracy of statistical calculations but also supported the scientific validity of the obtained results.

In conclusion, this study demonstrated that accurate SAS coding is a fundamental cornerstone of the scientific modeling process. Precision in coding is a factor that directly affects the success of the model and plays a critical role in ensuring the reliability and repeatability of the results obtained.

Author Contributions

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

	M.Ş.	İ.G.	E.Y.
C	35	35	30
D	35	35	30
S	35	35	30
DCP	35	35	30
DAI	35	35	30
L	35	35	30
W	35	35	30
CR	35	35	30
SR	35	35	30
PM	35	35	30
FA	35	35	30

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 authors declare 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.

Acknowledgements

This study has been published in the proceeding book of the IX. International Congress on Domestic Animal Breeding, Genetics and Husbandry-2025 (ICABGEH-25), Szeged, Hungary, August 12-14, 2020.

References

Bayazıt, M., Şahin, M., & Tolun, T. (2022). Modeling of in vitro gas production. *Black Sea Journal of Agriculture, 5*(3), 260–264. https://doi.org/10.47115/bsagriculture.1091017

dos Santos Cabral, Í., Azevêdo, J. A. G., Pina, D. dos S., Pereira, L. G. R., Fernandes, H. J., Almeida, F. M., Souza, L. L., de Lima, R. F., & Cirne, L. G. A. (2019). Evaluation of models utilized in in vitro gas production from tropical feedstuffs. *Semina: Ciências Agrárias, 40*(1), 443–456. https://doi.org/10.5433/1679-0359.2019v40n1p443

Goodwin, D. K., Ramsey, A. F., & Chvosta, J. (2018). *Applied econometrics with SAS: Modeling demand, supply, and risk*. SAS Press.

Korosteleva, O. (2018). *Advanced regression models with SAS and R*. Chapman & Hall.

Makkar, H. P. S., & Blümmel, M. (1997). In vitro rumen apparent and true digestibilities of tannin-rich forages. *Animal Feed Science and Technology, 67*(3–4), 245–251.

Orskov, E. R., & McDonald, I. (1979). The estimation of protein degradability in the rumen from incubation measurements weighted according to the rate of passage. *Journal of Agricultural Science, 92*(2), 499–503.

Özkan, Ç. Ö., & Kamalak, A. (2023). İn-vitro gaz üretim tekniğinde kullanılan matematik modeller ve karşılaştırma yöntemleri. İçinde *Tarımsal araştırmalarda istatistiksel analizler* (ss. 1–7). Holistence Publications.

Panik, M. J. (2010). *Regression modeling: Methods, theory, and computation with SAS*. Chapman & Hall.

Panik, M. J. (2013). *Growth curve modeling: Theory and applications*. John Wiley & Sons.

SAS Institute Inc. (1999). *The SAS system for Windows. Release 8.0.1*. SAS Institute Inc.

Stroup, W. W., Milliken, G. A., Claassen, E. A., & Wolfinger, R. D. (2009). *Generalized linear mixed models: Modern concepts, methods and applications*. SAS Press.

Black Sea Journal of Statistics

Doi: 10.5281/zenodo.18234746

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 2: 36-40/ December 2025

COMPARATIVE ANALYSIS OF INDIVIDUAL GROWTH CURVES IN BROILER CHICKENS

İsmail GÖK[1]*

[1]*Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Department of Agricultural Biotechnology, 46100, Kahramanmaraş, Türkiye*

Abstract: In this study, growth curves for 62 individual broiler chickens were obtained by evaluating weekly live weight data collected over a six-week period, using several growth models, including Gompertz, Gamma, Logistic, Richards, Bertalanffy, Cubic, Cubic Piecewise, Wilmink, Wood, Exponential, Monomolecular, and McNally. In the comparison of these growth curves, various statistical criteria, such as mean square error, adjusted coefficient of determination, accuracy factor, bias factor, Durbin-Watson statistic, Akaike information criterion, adjusted Akaike information criterion, and Bayesian information criterion, were employed. The Gompertz model provided the best fit across all criteria, accurately representing the growth curve of broiler chickens ($\bar{R}^2 = 0.99$, AIC= 62.42, CAIC: 76.42, BIC: 57.50). Similar to other studies in the literature, this model has been shown to produce reliable results under varying environmental and genetic conditions. The obtained data provide significant contributions to decision support systems in growth monitoring, genetic analysis, and the development of production strategies.

Keywords: Growth curves, Broiler chickens, SAS

***Corresponding author:** Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Department of Agricultural Biotechnology, 46100, Kahramanmaraş, Türkiye
E mail: gkisoo1995@gmail.com (İ. GÖK)
İsmail GÖK https://orcid.org/0000-0002-0759-1187

Received: November 01, 2025
Accepted: November 28, 2025
Published: December 15, 2025

Cite as: Gök, İ. (2025). Comparative analysis of individual growth curves in broiler chickens. *Black Sea Journal of Statistics, 1*(2), 36–40.

1. Introduction

Poultry farming plays a critical role in meeting the increasing global demand for animal protein. Broiler chickens, in particular, have become one of the most preferred species in industrial production systems due to their rapid growth capacity, short production cycles, and high feed conversion ratios (Bessei, 2006; Akinsola et al., 2021). In this context, accurately monitoring and modeling the growth performance of broiler chickens is crucial not only for enhancing economic efficiency but also for improving genetic selection and feeding programs (Silambarasan et al., 2012).

Animal growth modeling is typically carried out through mathematical growth curves that describe the changes in live weight over time. These models define the biological stages of growth (initial, rapid growth, plateau phase) while also quantifying individual performance differences (Knízetová et al., 1991). However, most growth modeling is often done using aggregate data, overlooking the need to model individual weight changes. Individual modeling provides more accurate insights into the growth rate, genetic potential, and environmental influences on each bird, enabling more personalized and precise predictions. Therefore, individual-based modeling studies are critical for optimizing production systems and making more informed decisions (Buzala and Janicki, 2016; Akinsola et al., 2021).

This study presents a comparative analysis of twelve different growth models (Gompertz, Gamma, Logistic, Richards, Bertalanffy, Cubic, Piecewise Cubic, Wilmink, Wood, Exponential, Monomolecular, McNally) using individual live weight data to gain a more detailed understanding of broiler chicken growth. Each of these widely used models offers distinct parametric structures for different growth phases, and their performance in modeling individual data sets has been evaluated. The findings will facilitate more accurate modeling of the growth process in broiler chickens and demonstrate how individual based modeling can contribute to decision support processes in broiler chickens production.

2. Materials and Methods

The material of this study consists of live weight data of broiler chickens raised at the Poultry Farming Unit of the Research and Application Farm, Faculty of Agriculture, Çukurova University. A total of 62 broiler chickens were included in the study, with the live weight of each bird measured individually from week 0 (hatch) to week 6 on a weekly basis. The mean and standard deviation values of the obtained data set are presented in Table 1.

In the modeling of the growth process, twelve growth models, which are frequently used in the literature and possess different structural characteristics, were evaluated. These models include Gompertz, Gamma, Logistic, Richards, Bertalanffy, Cubic, Piecewise Cubic,

Wilmink, Wood, Exponential, Monomolecular, and McNally models. The mathematical formulas for each model are provided in Table 2.

Table 1. Mean and standard deviation values for the broiler chickens

	Mean and Standard Deviation
1.Week	42.02±2.10
2.Week	223.91±18.21
3.Week	548.8±722.53
4.Week	1151.64±74.30
5.Week	1509.23±137.49
6.Week	2063.3±5124.35

Table 2. Equations used in modeling growth curves

Model Names	Equality
Gompertz	$Y_t = \beta_0 \exp(-\beta_1 \exp(-\beta_2 t))$
Gamma	$Y_t = \beta_0^{\beta_1}(e^{-\beta_2 t})$
Logistic	$Y_t = \beta_0(1 + \beta_1 \exp(-\beta_2 t))^{-1}$
Richard	$Y_t = \beta_0(1 + \beta_1 \exp(-\beta_2 t))^{\beta_3}$
Bertalanffy	$Y_t = \beta_0(1 - \beta_1 \exp(-\beta_2 t))^3$
Cubic	$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$
Cubic Piecewise	$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-a)^3$
Wilmink	$Y_t = \beta_0 + \beta_1 t + \beta_2 e^{-0.05t}$
Wood	$Y_t = \beta_0 t^{\beta_1} * e^{-\beta_2 t}$
Exponential	$Y_t = \beta_0 * e^{\beta_1 t}$
Monomolecular	$Y_t = \beta_0(1 - e^{-\beta_1 t})$
McNally	$Y_t = \beta_0 t^{\beta_1} \exp(-\beta_2 t + \beta_3 t^{1/2})$

In the equations given in Table 2, β_0, β_1, β_2, β_3, and β_4, represent the parameters in the equations, t represents the time and Y_t represents the live weight gains at the t-th time (Şahin and Efe 2010; Yavuz et al., 2019; Yalçınöz and Şahin; 2020).

Table 3. Model comparison criterions

Criterion	Equality
Mean Square Error	$MSE = ESS/EDF$
Adjusted Coefficient of Determination	$\bar{R}^2 = 1 - (1 - R^2)(n - 1/(n - p - 1))$
Accuracy Factor	$AF = 10^{\sum_{i=1}^{n} \lvert \log(\hat{Y}_i/Y_i) \rvert / n}$
Bias Factor	$BF = 10^{\sum_{i=1}^{n} \log(\hat{Y}_i/) / n}$
Durbin-Watson Value	$DW = \frac{\sum_{i=2}^{n}(e_1 - e_2)^2}{\sum_{i=1}^{n} e_i^2}$
Akaike Information Criterion	$AIC = n x \ln\left(\frac{MSE}{n}\right) + 2k$
Corrected Akaike Information Criterion	$CAIC = n x \ln\left(\frac{MSE}{n}\right) + \left(\frac{n(n+p)}{n-p-2}\right)$
Bayesian Information Criterion	$BIC = n x \ln\left(\frac{MSE}{n}\right) + k \ln(n)$

MSE: Mean Square Error, ESS: Error Sum of Squares, EDF: Error Degrees of Freedom, n: sample size, p: number of independent variables, $\hat{Y}_i$: estimated value, Y_i: observation value, e_i: error term and k: number of parameters.

The goodness of fit of each model to the observed data was evaluated using various statistical comparison criterions, including mean square error, adjusted coefficient of determination, accuracy factor, bias factor, Durbin-Watson statistic, Akaike information criterion, adjusted Akaike information criterion, and Bayesian information criterion. Details of these comparison criterions are provided in Table 3 (Tolun et al., 2023; Tolun et al., 2024; Çetenak et al., 2024).

3. Results and Discussion

In this study, the arithmetic means and standard errors of the model comparison criteria are presented in Table 4, and the parameter estimate values are provided in Table 5. The growth predictions made by each model based on the mean values are graphically shown in Figure 1, allowing for a visual assessment of the models' fit to the observed data.

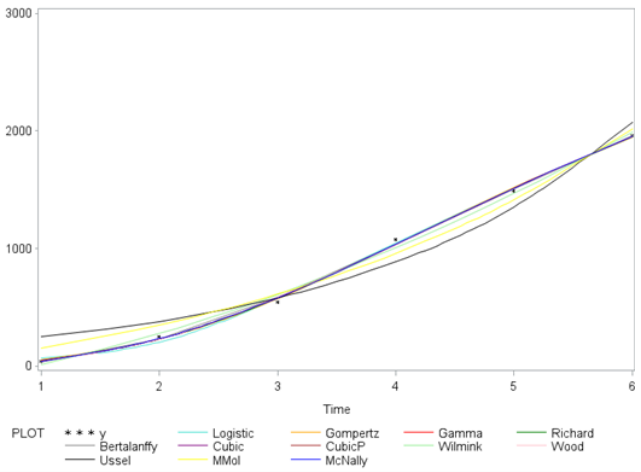


Figure 1. Growth curves obtained for all models.

Table 4. Model comparison and evaluation criterions

Models	MSE	CCD	AIC	CAIC
Gompertz	5533.27±1084.80	0.99±0.01	62.42±2.74	76.42±2.74
Gamma	14689.11±3535.81	0.99±0.001	63.41±2.11	77.41±2.11
Logistic	14094.11±3594.93	0.99±0.003	63.24±2.05	77.24±2.05
Richard	20282.83±5284.30	0.99±0.004	65.15± 2.09	111.82±2.09
Bertalanffy	19808.81±6831.76	0.99±0.002	66.28±2.26	112.95±2.26
Cubic	51392.75±31097.3	0.90±90.02	63.56±2.50	77.56 ±2.50
Cubic Piecewise	18059.42±4613.17	0.99±0.003	64.74±1.94	111.41±1.94
Wilmink	16782.94±2720.14	0.99±0.004	69.86±1.14	116.53±1.14
Wood	12409.94±3246.48	0.99±0.003	63.81± 1.97	77.81±1.97
Exponential	70797.88±18939.0	0.99±0.01	79.42±0.97	93.42±0.97
Monomolecular	33752.85±7410.72	0.99±0.003	74.50±0.98	88.50±0.98
McNally	172174.8±55500.3	0.99±0.04	73.07±3.21	87.07±3.21
	BIC	DW	AF	BF
Gompertz	57.50±2.11	2.01±0.1	1.03±0.08	1.05±0.04
Gamma	58.20±2.05	2.28±0.5	1.03±0.01	1.08±0.02
Logistic	58.02±2.03	1.98±0.4	1.03±0.01	1.09±0.01
Richard	57.91±2.14	1.88±0.4	1.03±0.02	1.01±0.02
Bertalanffy	59.04±2.26	2.12±0.2	1.05±0.10	1.06±0.05
Cubic	58.65±2.48	2.22±0.3	1.13±0.04	1.11±0.02
Cubic Piecewise	58.33±1.91	2.46±0.6	1.12±0.04	1.11±0.02
Wilmink	62.62±1.11	1.65±0.4	1.08±0.04	1.01±0.07
Wood	58.90±1.90	2.58±0.7	1.04±0.01	1.05±0.02
Exponential	74.51±0.94	2.10±0.2	1.08±0.01	1.06±0.01
Monomolecular	69.59±0.99	2.17±0.3	1.06±0.01	1.04±0.01
McNally	68.16±3.14	2.06±0.1	1.04±0.04	1.02±0.02

MSE: Mean Square Error, **CCD:** Corrected Coefficient of Determination, **AIC:** Akaike Information Criterion, C**AIC:** Corrected Akaike Information Criterion, **BIC:** Bayesian Information Criterion, **DW:** Durbin-Watson, **AF:** Accuracy Factor, **BF:** Bias Factor.

Table 5. Parameters estimates of growth curves

Models	a	b	c	d	e
Gompertz	3141.62±110.18	6.54±0.27	0.42±0.02	-	-
Gamma	12.20±3.26	1.31±0.35	0.17±0.04	-	-
Logistic	52.14±4.98	150.68±20.46	0.13±0.01	-	-
Richard	3156.92±114.53	0.28±0.13	0.45±0.03	0.15±0.04	-
Bertalanffy	3613.18±112.81	1.08±0.03	0.26±0.01	-	-
Cubic	98.27±47.79	-235.87±50.98	144.27±24.88	-11.61±2.12	-
Cubic Piecewise	96.12±45.1	-212.14±54.17	140.19±26.11	-10.13±2.15	0.001±0.008
Wilmink	2684.00± 7717.51	405.32±108.32	2923.40±781.31	-	-
Wood	48.19±2.55	3.02±0.208	0.31±0.03	-	-
Exponential	160.62±3.38	0.42±0.004	-	-	-
Monomolecular	-451.40±18.75	-0.28±0.007	-	-	-
McNally	120.55±31.06	2.64±0.51	0.48±0.13	1.59±1.17	-

According to the comparison criterions provided in Table 4, the models with the lowest mean square error values are Gompertz, Logistic, and Gamma, in that order. This result indicates that these models have higher predictive power compared to the others.

When evaluating the adjusted coefficient of determination, all models produced results at the 0.99 level. This suggests that the models have high explanatory power and exhibit a strong fit to the observed data.

The Durbin-Watson statistic yielded values close to 2 for all models. This finding indicates that there is no significant autocorrelation among the error terms, and the models meet the assumption of independent errors.

The accuracy and bias factors of the models being close to 1 further support that the predictions are reliable and unbiased.

According to the evaluations of Akaike information criterion, adjusted Akaike information criterion, and Bayesian information criterion, the Gompertz, Logistic, and Gamma models obtained lower values compared to the other models.

When all metrics are considered together, it can be concluded that the Gompertz model best represents the live weight data of broiler chickens.

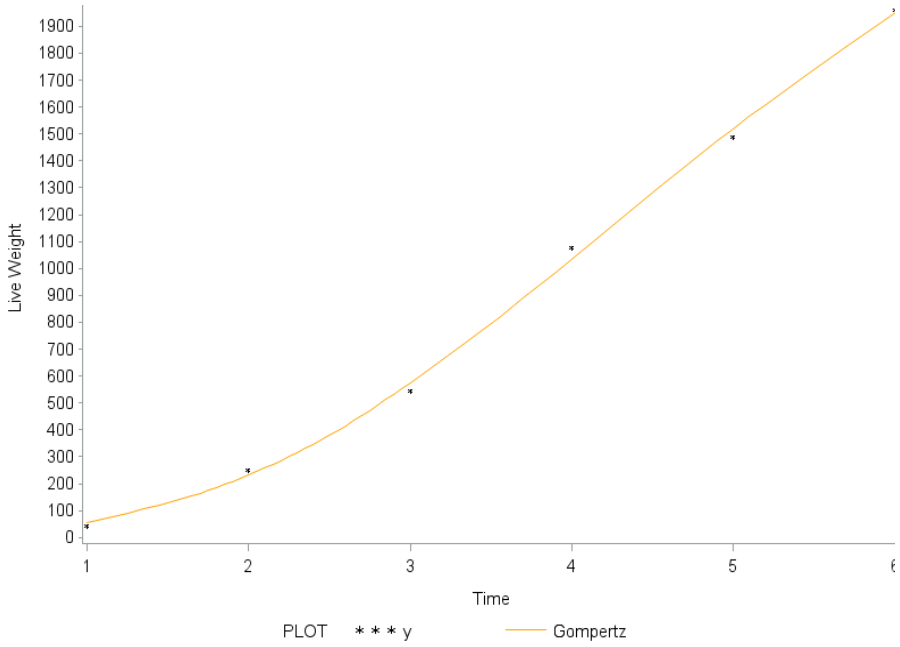


Figure 2. Growth curve obtained for the Gompertz model.

The predicted growth curve for this model is shown in Figure 2, and its high fit to the observed data is visually confirmed.

The results obtained indicate that the Gompertz model demonstrated stronger performance in terms of statistical fit compared to the other models. This is attributed to the model's flexible structure, which allows it to effectively represent the different growth phases (initial, rapid growth, and plateau phase).

The use of various statistical criteria, including mean square error, adjusted coefficient of determination, Durbin-Watson statistic, accuracy and bias factors, as well as Akaike, adjusted Akaike, and Bayesian information criteria, has enabled a comprehensive evaluation of both the statistical adequacy and predictive power of the models.

The Gompertz model emerging as the most suitable model in this study is consistent with findings from previous research. Ricklefs (1967), Tzeng and Becker (1981), Anthony et al. (1986), Anthony et al. (1991), Barbato (1991), Marayuma (1998), Şengül and Kiraz (2005), Nooris et al. (2007), Demuner et al. (2017), and Şengül et al. (2024) have all reported that the Gompertz model is frequently used to represent poultry growth data.

4. Conclusion

The findings of this study reveal that the Gompertz model outperforms the other models in terms of overall fit and predictive success. In criteria such as mean square error, adjusted coefficient of determination, accuracy factor, bias factor, Durbin-Watson statistic, and information criteria, the Gompertz model was identified as the statistically most suitable model. Its low error rate and high explanatory power indicate that it effectively represents the growth process of broiler chickens.

The results emphasize that model selection is of great importance not only for biological relevance but also for statistical validity. The Gompertz model's ability to balance different growth phases, its simple structure, and high predictive accuracy make it a valuable tool for applied animal science research and production planning. In this context, its use as a decision support tool in areas such as feeding strategies, determining slaughter times, and economic efficiency analysis is considered to be beneficial.

Future research should focus on testing the performance of the Gompertz model in different genetic lines, environmental conditions, and production systems to assess its generalizability. Additionally, comparing parametric models with modern approaches such as machine learning and Bayesian methods could contribute to the development of more flexible and effective solutions in growth modeling.

Author Contributions

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

	İ.G.
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 declare 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

Akinsola, O., Sonaiya, E. B., Bamidele, O., Hassan, W. A. A., Yakubu, A., Asayi, F., Uduak, O., Olayınka, A., & Adebambo, O. A. (2021). Comparison of five mathematical models that describe growth in tropically adapted dual-purpose breeds of chicken. *Journal of Applied Animal Research, 49*(1), 158–166. https://doi.org/10.1080/09712119.2021.1915792

Anthony, N. B., Emmerson, D. A., & Nestor, K. E. (1991). Research note: Influence of body weight selection on the growth curve of turkeys. *Poultry Science, 70*(1), 192–194. https://doi.org/10.3382/ps.0700192

Anthony, N. B., Nestor, K. E., & Bacon, W. L. (1986). Growth curves of Japanese quail as modified by divergent selection for 4-week body weight. *Poultry Science, 65*(10), 1825–1833. https://doi.org/10.3382/ps.0651825

Barbato, G. F. (1991). Genetic architecture of growth curve parameters in chickens. *Theoretical and Applied Genetics, 83*(1), 24–32.

Bessei, W. (2006). Welfare of broilers: A review. *World's Poultry Science Journal, 62*(3), 455–466.

Buzala, M., & Janicki, B. (2016). Review: Effects of different growth rates in broiler breeder and layer hens on some productive traits. *Poultry Science, 95*(9), 2151–2159. https://doi.org/10.3382/ps/pew173

Çetenak, T., Gök, İ., Yavuz, E., & Şahin, M. (2024). Statistical models and evaluation criteria used in poultry farming. *Black Sea Journal of Agriculture, 7*(6), 710–719. https://doi.org/10.47115/bsagriculture.1532659

Demuner, L. F., Suckeveris, D., Muñoz, J. A., Caetano, V. C., Lima, C. G., de Faria Filho, D. E., & de Faria, D. E. (2017). Adjustment of growth models in broiler chickens. *Pesquisa Agropecuária Brasileira, 52*(12), 1241–1252. https://doi.org/10.1590/S0100-204X2017001200013

Knízetová, H., Hyánek, J., Kníze, B., & Roubícek, J. (1991). Analysis of growth curves of fowl. I. Chickens. *British Poultry Science, 32*(5), 1027–1038.

Maruyama, K., Potts, W. J. E., Bacon, W. L., & Nestor, K. E. (1998). Modeling turkey growth with the relative growth rate. *Growth, Development & Aging, 62*(4), 123–139.

Norris, D., Ngambi, J. W., Benyi, K., Makgahlela, M. L., Shimelis, H. A., & Nesamvuni, E. A. (2007). Analysis of growth curves of indigenous male Venda and Naked Neck chickens. *South African Journal of Animal Science, 37*(1), 21–26. https://doi.org/10.4314/sajas.v37i1.4021

Ricklefs, R. E. (1967). Relative growth, body constituents, and energy content of nestling barn swallows and red-winged blackbirds. *The Auk, 84*(4), 560–570. https://doi.org/10.2307/4083336

Şahin, M., & Efe, E. (2010). Use of cubic spline regressions in modeling lactation curves in dairy cattle. *Kahramanmaraş Sütçü İmam Üniversitesi Journal of Natural Sciences, 13*(2), 17–22.

Şengül, T., Çelik, Ş., Şengül, A. Y., İnci, H., & Şengül, Ö. (2024). Investigation of growth curves with different nonlinear models and MARS algorithm in broiler chickens. *PLoS ONE, 19*(11), e0307037. https://doi.org/10.1371/journal.pone.0307037

Şengul, T., & Kiraz, S. (2005). Non-linear models of growth curves in large white turkeys. *Turkish Journal of Veterinary and Animal Sciences, 29*(2), 331–337.

Silambarasan, P., Samanta, R., & Das, T. K. (2012). Production performance of broiler chickens influenced by feed restriction systems. *Indian Journal of Animal Sciences, 82*(11), 1451–1455.

Tolun, T., Gök, İ., & Şahin, M. (2024). Modeling of some egg characteristics in henna partridges. *Black Sea Journal of Agriculture, 7*(6), 729–742. https://doi.org/10.47115/bsagriculture.1555738

Tolun, T., Yavuz, E., Şahin, M., & Gök, İ. (2023). Modeling egg curves in partridges. *Black Sea Journal of Agriculture, 6*(1), 21–25. https://doi.org/10.47115/bsagriculture.1139272

Tzeng, R. Y., & Becker, W. A. (1981). Growth patterns of body and abdominal fat weight in male broiler chickens. *Poultry Science, 60*(6), 1101–1106. https://doi.org/10.3382/ps.0601101

Yalçınöz, E., & Şahin, M. (2020). Modeling of egg production curves in laying hens. *Kahramanmaraş Sütçü İmam Üniversitesi Journal of Agriculture and Natural Sciences, 23*(5), 1373–1378. https://doi.org/10.18016/ksutarimdoga.vi.691069

Yavuz, E., Önem, A. B., Kaya, F., Çanga, D., & Şahin, M. (2019). Modeling of individual growth curves in Japanese quails. *Black Sea Journal of Engineering Science, 2*(1), 11–15.

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.

Black Sea Journal of Statistics

Doi: 10.5281/zenodo.18234811

Open Access Journal
e-ISSN: 3108-4265

Derleme (Review)
Cilt 1 - Sayı 2: 47-53 / Aralık 2025
(Volume 1 - Issue 2: 47-53/ December 2025)

GENOMİK DAMZILIK DEĞER TAHMİNİ

Ebru ŞİRİN[1], Serdar GENÇ[2]*

[1]*Kırşehir Ahi Evran University, Agricultural Faculty, Department of Agricultural Biotechnology, 40100, Kırşehir, Türkiye*
[2]*Kırşehir Ahi Evran University, Agricultural Faculty, Department of Animal Science, 40100, Kırşehir, Türkiye*

Özet: Hayvan yetiştiriciliği, yaklaşık 10 bin yıl önce ilk ıslah çalışması olarak bilinen evciltme ile başlamaktadır. Evciltmeden bugüne, etkili seleksiyon ve çiftleştirme programları ile birçok karakter bakımından çeşitli hayvan gruplarında önemli genetik ilerlemeler sağlanmıştır. Damızlık değerinin tahminiyle yapılan fenotipe dayalı seleksiyon uygulamaları ile seleksiyon yoğunluğu artırılmış ve generasyon aralıkları kısaltılmıştır. Yetiştiricilik bakımından genetik ıslahın amacı, mevcut çevre şartlarında gelecek generasyonda istenen karakterlerin iyileştirilmesini sağlamaktır. Arzu edilen özelliklere ulaşılabilmesi için çevre faktörleri ile birlikte genotipin de geliştirilmesi gerekmektedir. Bu amaçla REML, BLUP gibi klasik ıslah yöntemlerinin kullanılabilmekle beraber günümüz teknolojisinde meydana gelen gelişmeler genomik bilginin de kullanılabilmesini sağlamıştır. Bu çalışmada markör destekli seleksiyon ve genomik seleksiyon açıklanmıştır.

Anahtar kelimeler: Damızlık değer tahmini, Seleksiyon, Genetik ıslah

Genomic Breeding Value Estimation

Abstract: Animal breeding began approximately 10,000 years ago with domestication, known as the first breeding effort. From domestication to the present day, effective selection and breeding programs have led to significant genetic advances in various animal groups across many traits. Phenotype-based selection, based on breeding indicator prediction, has increased selection intensity and shortened breeding intervals. The goal of genetic improvement in breeding is to enable the modification of desired traits in future generations under current environmental conditions. To obtain improved traits, the genotype must be incorporated along with environmental factors. For this purpose, advances in current technology have enabled the use of genomic methods, enabling the use of classical breeding methods such as REML and BLUP. These traits are described in marker-assisted selection and genomic selection.

Keywords: Breeding value estimation, Selection, Genetic breeding

***Corresponding author:** Kırşehir Ahi Evran University, Agricultural Faculty, Department of Animal Science, 40100, Kırşehir, Türkiye
E mail: serdargenc1983@gmail.com (S. GENÇ)
Ebru ŞİRİN https://orcid.org/0000-0002-7416-6367
Serdar GENÇ https://orcid.org/0000-0002-1512-9072

Gönderi: 21 Kasım 2025 **Received:** November 21, 2025
Kabul: 04 Aralık 2025 **Accepted:** December 04, 2025
Yayınlanma: 15 Aralık 2025 **Published:** December 15, 2025

Cite as: Şirin, E., & Genç, S. (2025). Genomic breeding value estimation. *Black Sea Journal of Statistics, 1*(2), 47–53.

1. Giriş

Hayvan yetiştiriciliği, yaklaşık 10 bin yıl önce ilk ıslah çalışması olarak bilinen evciltme ile başlamaktadır (Anderson, 2001; Mc Gill ve Lievaart, 2011; Eggen, 2012). Evciltmeden bugüne, etkili seleksiyon ve çiftleştirme programları ile birçok karakter bakımından çeşitli hayvan gruplarında önemli genetik ilerlemeler sağlanmıştır. Yetiştirme değerinin tahminiyle yapılan fenotipe dayalı seleksiyon uygulamaları ile seleksiyon yoğunluğu artırılmış ve generasyon aralıkları kısaltılmıştır (Mc Gill ve Lievaart, 2011).

Son 30 yıllık periyotta inek başına yıllık süt üretimi 40-80 kg artış (yaklaşık %1) göstermiş, tavuklarda ise daha az besleme ile ticari tüketim ağırlığına ulaşma süresi 50 yıl içinde yaklaşık % 65 oranında azaltılmıştır (Hayes, 2007). Çiftlik hayvanlarında karakterler kalitatif ve kantitatif karakterler olarak gruplandırılır. Kalitatif karakterler bir ya da birkaç gen çifti tarafından kontrol edilir ve çevre faktörlerinden hemen hiç etkilenmezler. Bu nedenle, bu karakterler için istenilen genetik yapıların kısa sürede ve kolayca oluşturulabilmesi mümkündür. Ancak, verimle ilgili olan karakterler genellikle kantitatif karakterlerdir. Kantitatif karakterler çok sayıda gen tarafından kontrol edilir ve çevreden etkilenirler. Böyle karakterler bakımından bireyin fenotipi değerlendirilerek genotipi hakkında her zaman isabetli sonuçlar elde etmek mümkün olmamaktadır (Daş, 2015).

Yetiştiricilik bakımından genetik ıslahın amacı, bir sürü ya da popülasyondaki hayvanlarda belirli verimleri artırmaktır. Verimleri artırmak için çevre faktörleri ile birlikte genotipin de geliştirilmesi gerekmektedir. Genotipin geliştirilmesinde ise çeşitli seleksiyon yöntemleri kullanılmaktadır (Mc Gill ve Lievaart, 2011; Gürses ve Bayraktar 2014).

2. Modern Islah Yöntemleri

2.1. Markör Destekli Seleksiyon

Yetiştiricilik bakımından genetik ıslahın amacı, bir sürü ya da popülasyondaki hayvanlarda belirli verimleri artırmaktır. Verimleri artırmak için çevre faktörleri ile birlikte genotipin de geliştirilmesi gerekmektedir. Ancak, verimle ilgili olan karakterler genellikle kantitatif

karakterlerdir. Kantitatif karakterler çok sayıda gen tarafından kontrol edilir ve çevreden etkilenirler. Böyle karakterler bakımından bireyin fenotipi değerlendirilerek genotipi hakkında her zaman isabetli sonuçlar elde etmek mümkün olmamaktadır (Daş, 2015).

Seleksiyon uygulamalarında hedef, genotipin doğru tahminini sağlayabilmenin yanında istenilen genotipe sahip olan ve bunu fenotipe yansıtabilen hayvanları kısa sürede ve düşük maliyetlerle elde edebilmektir (Özdemir ve Doğru, 2008; Daş, 2015).

Seleksiyonda moleküler yöntemler üzerine yapılan çalışmalar artmıştır. 1990'lı yıllara kadar çiftlik hayvanlarında bazı kantitatif karakterler ile protein polimorfizmleri arasındaki ilişkiler araştırılmıştır. Sığırlarda kan grupları ile süt verimi arasında bağlantı kurulmaya çalışılmış, süt ve kan proteinlerindeki polimorfizmlerle çeşitli verimler arasındaki ilişkiler incelenmiştir. Ancak proteinler üzerine yapılan bu çalışmalar, karakterlerle yeterince ilişkilendirilememiştir (Özşensoy ve Kurar, 2013; Bal ve Akyüz, 2014).

DNA markörlerine dayalı seleksiyon fikri yaklaşık 40 yıl öncesinden beri var iken, kantitatif karakterlerin oluşumunda rol alan genlerin çok sayıda oluşu ve bu genler arasındaki ilişkilerin yeterince bilinmeyişi nedeniyle kullanılabilirliği 2000'li yılların başından itibaren mümkün olabilmiştir (Şekil 1) (Hayes vd, 2009).

Şekil 1. Moleküler markerların kullanımı.

Kantitatif karakterlerin oluşumunda etkili olan genetik varyasyonu açıklamak için "Infinitesimal model" ve "The finite loci model" olmak üzere 2 model önerilmiştir. Hayvan yetiştiriciliğinde uzun süre oldukça öneme sahip olan Infinitesimal model'de, bir karakterin oluşumunda birbirinden bağımsız çok sayıda genin az ancak toplamalı etkisinin olduğu önerilmiştir. The finite loci model ile ise genomun 20000 civarında gen veya sınırlı sayıda lokustan oluştuğu belirlenmiş, kantitatif karakterlerdeki varyasyonda sınırlı sayıda lokusun etkili olduğu anlaşılmıştır.

Herhangi bir karakterin oluşumundan önemli derecede sorumlu olan majör genler ve çok sayıda lokusun etkileri 2000'li yıllarda araştırılmış ve lokuslardaki (Quantitative Trait Loci, QTL: Kantitatif özelliklerin belirlendiği lokuslar) varyasyonların karakterler üzerine önemli etkilerinin olduğuna yönelik kanıtlar artmıştır. QTL bölgelerinin belirlenmesi için aday gen yaklaşımı ve QTL haritalama olmak üzere 2 yaklaşım bulunmaktadır (Hayes, 2007).

TGDD'nin seleksiyonda kullanılması genomik seleksiyon olarak adlandırılmıştır. Genomik seleksiyon, belirteçlerin (SNP'ler) tüm popülasyondaki QTL'lerle bağlantı dengesizliği (BD) içinde olmasını gerektirir. BD, iki lokusun alelleri arasındaki (örneğin, bir belirtecin ve bir QTL'nin alelleri arasındaki) rastgele olmayan ilişki olarak tanımlanabilir. Aynı kromozom üzerinde bir belirteç lokusu A (alelleri A_1, A_2) ve bir QTL lokusu B (alelleri B_1 ve B_2) göz önüne alındığında, LD, belirteç ile QTL arasındaki korelasyon (r^2) olarak Eşitlik 1 ve 2'de verildiği şekilde ölçülebilir:

$$D = frek.(A_1B_1)*frek.(A_2B_2) - frek.(A_1B_2)*frek.(A_2B_1) \quad (1)$$

$$r^2 = D^2/[frek.(A_1)*frek.(A_2)*frek.(B_1)*frek.(B_2)] \quad (2)$$

Belirteç ile QTL arasındaki r2, belirteçde açıklanabilen QTL'ye ait varyansın oranını gösterir.

Aday gen yaklaşımında tek bir karakterin oluşumunda büyük etkiye sahip majör genlerdeki varyasyonların, karakterlerdeki varyasyonlardan sorumlu olabileceği düşünülmüştür. Birçok gen ve gen bölümü çok sayıda çiftlik hayvanında incelenerek karakterlerdeki varyasyonlarla ilişkilendirilmiştir. Koyunlarda; boorola genindeki varyasyonla bir batında çok sayıda yavru elde etme ve östrojen reseptör geni (ESR) ile bir batında doğan yavru sayısı gibi ilişkiler ortaya konulmuştur. Bu çalışmalara benzer çok sayıda aday gen çalışması birçok tür için yapılmış ve yapılmaya devam etmektedir. Fakat aday gen analizinde; çoğunlukla bir karakteri birden çok genin etkiliyor oluşu ve birçok hayvanda çok sayıda genin dizilenmesinin gerekliliği, dolayısıyla ortaya çıkan yüksek maliyet bu yöntemin dezavantajlarını oluşturmuştur. Ayrıca aday gen yaklaşımında bir gende tespit edilen varyasyonun fenotipe yansıması ile ilgili yapılan çıkarımlar da yanıltıcı olabilmektedir (Hayes, 2007; Komisarek ve Doynek, 2009).

İkinci yaklaşımda herhangi bir karakterdeki varyasyonla ilişkisi olan kromozom bölgeleri tanımlanmakta ve incelenmektedir. QTL haritalama olarak bilinen bu yaklaşımda karakteri etkileyen gen ya da genler bilinmemektedir. Bunun yerine gen ya da genlerle ilişkili DNA markörleri kullanılmakta, markörler ile varyasyonlar arasındaki ilişkiler incelenmektedir. Markörler, kromozomlarda fonksiyonları bilinmeyen fakat kalıtımı izlenilebilen sınırlı bölgelerdir. Bağlantı haritalama (Linkage Mapping) sayesinde belirli miktarda markör ile QTL bölgeleri arasındaki ilişkiler populasyon çapında incelenmiş ve birçok karakter için bağlantı analizine dayalı QTL haritalama çalışmaları neredeyse bütün çiftlik hayvanlarında yapılmıştır. RFLP (Restriction Fragment Length Polymorphism), SSCP (Single Strand Conformation

Polymorphism), RAPD (Random Amplified Polymorphic DNA), AFLP(Amplified Fragment Length Polymorphism), SNP (Single Nucleotide Polymorphism) gibi çok sayıda DNA temelli genetik markör olmakla birlikte daha çok yüksek derecede değişken mikrosatellit markör kullanılarak yakın akrabalı yetiştirmelerin olduğu sürülerde bağlantı analizi uygulamaları yapılmış ve çeşitli QTL bölgeleri haritalanmıştır (Özşensoy ve Kurar, 2013; Sharma vd., 2015).

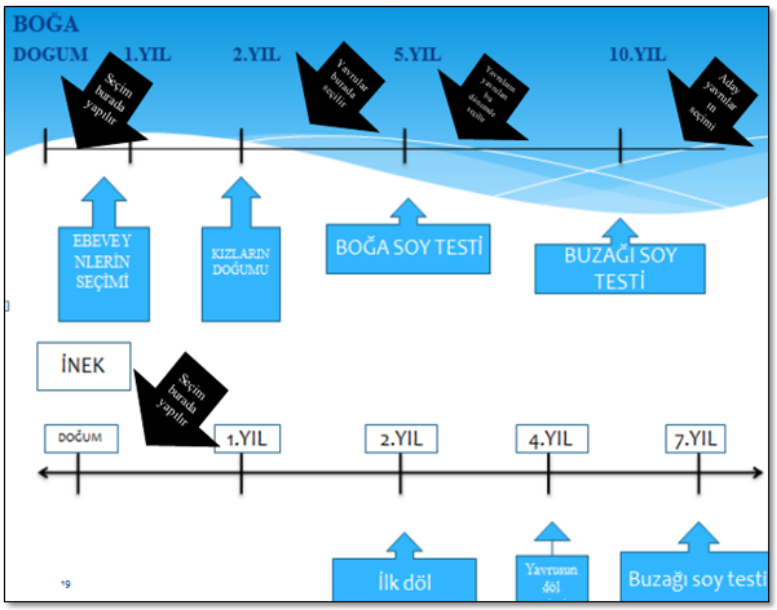
Ancak bağlantı analizleri ile haritalanan markör ve QTL bölgeleri arasındaki uzaklıklar fazla olduğu için markör kullanımı ve varyasyon tespiti hayvan yetiştirme programlarında yeterince etkili olamamıştır. Bağlantı analizi ile yapılan haritalama çalışmalarında çok sayıda akraba içeren birden fazla sürü kullanılması gerekmektedir. Ayrıca bu analiz yöntemi ile bir popülasyonda herhangi bir karakterle ilişkili markör başka popülasyonlarda geçersiz olabilmektedir. Bağlantı analizi daha çok sürü bazında çalışılabilmekte ve büyük popülasyonlarda yeterince güvenilir olmamaktadır. Genomda markör ve QTL bölgeleri arasındaki mesafe fazla olduğu için generasyonlar boyunca bağlantı analizine dayalı çalışmalar etkili değildir (Dekkers, 2004).

QTL tespiti, Linkage disequlibrium (Bağlantı eşitsizliği, LD)'dan yararlanılarak mümkün olmuştur. LD bir popülasyondaki bireylerin genomlarının bazı bölgelerinde beklenilen rekombinasyonun şekillenmemesidir. Bir QTL bölgesi ile LD halinde olan yani beraber kalıtılan bölgelerin markör olarak kullanımıyla QTL bölgeleri tespit edilebilmiştir (Goddard ve Hayes, 2009; Özbeyaz ve Kocakaya, 2011). ABD Tarım Bakanlığı tarafından çiftlik hayvanlarına ait QTL haritaları geniş bir veritabanı olarak saklanmaktadır (http://www.animalgenome. org/QTLdb).

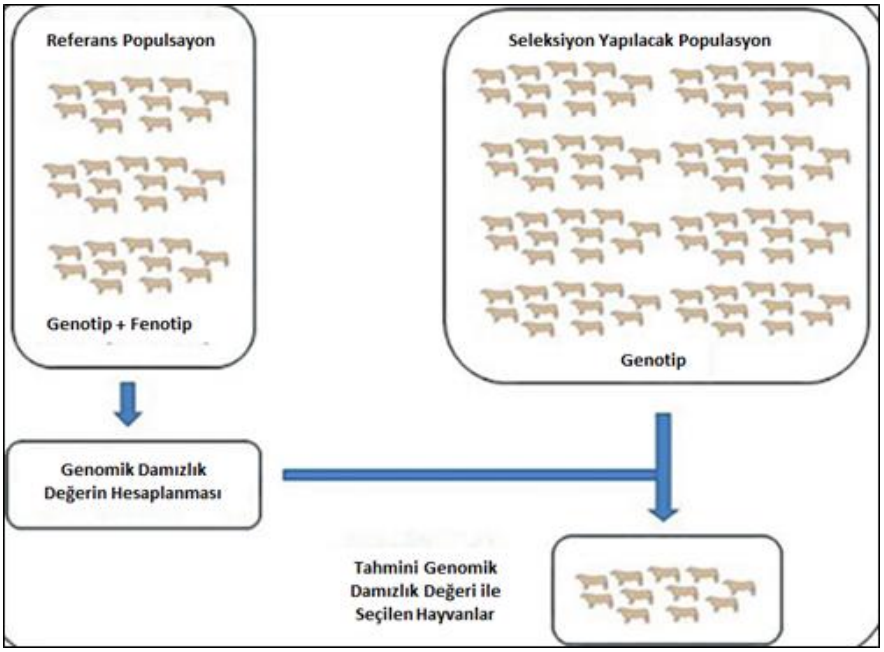
2.2. Genomik Seleksiyon

Meuwissen vd. (2001) yılında çok sayıda LD markör ile QTL bölgelerinin genomdaki konumları kesin bir şekilde bilinmeden hayvanların tahmini yetiştirme değerlerinin ortaya konulabileceğini bildirmiştir. GS, bireyler arasında herhangi bir karakterde varyasyona sebep olan genetik varyasyonu çok sayıda markörün, QTL bölgeleri ile arasındaki LD durumunu kullanarak tespit etme prensibine dayanmaktadır. Yöntem, tüm genomu kapsayan yoğun markör kullanımıyla bir populasyonda bulunan adayların, yetiştirme değerlerinin istatistiksel metotlar kullanılarak tahmin edildiği bir çeşit MAS yöntemidir (Hayes vd., 2013; Cole ve Silva, 2016).

Genomik seleksiyonda Genomik Damızlık Değeri'nin (GDD) belirlenebilmesi için referans populasyonlar kullanılmaktadır. Referans populasyondaki hayvanların genotip ve fenotip bilgilerinin kombinasyonuyla hesaplanan Tahmini Genomik Damızlık Değeri (GEBV), fenotip bilgisi olmayan ve sadece markörlerle genotiplendirilmiş olan aday populasyondaki hayvanların GDD'nin hesaplanmasında kullanılır. Bu hesaplamalar ile aday populasyonda üstün olduğu tespit edilen hayvanlar gelecek generasyon için ebeveyn olarak seçilir (Şekil 2 ve 3).



Şekil 2. Genomik seleksiyonda boğa ve ineğin seçimi.

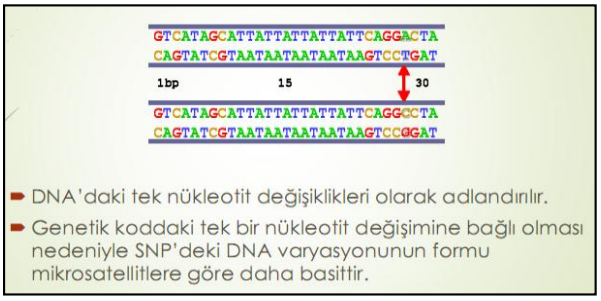


Şekil 3. Fenotipi bilinen referans bir populasyonun genotiplendirilmesi ile elde edilen GYD'nin sadece genotiplendirilmiş aday populasyonda kullanımı ile aday populasyondaki hayvanların TGDD'nin belirlenmesi ve en iyilerin seçimi.

2.3. Genomik Seleksiyon Uygulamalarının Güvenilirliği

Referans populasyonun büyüklüğüne, hedef populasyon ile referans populasyon arasındaki genetik ilişkiye, markör yoğunluğuna, hesaplama metotlarına ve karakterlerin kalıtım derecelerine bağlıdır. Kalıtım derecesi düşük karakterlerde GDD doğruluğunun yüksek olması için referans populasyondaki hayvan sayısının fazla olması gerekmektedir. Genomik seleksiyon uygulamalarında GDD'yi hesaplamak için Least Squares (En Küçük Kareler Yöntemi), BLUP (Best Linear Unbiased Prediction) ve Bayes (Bayesian Estimation) yöntemleri kullanılmaktadır. GDD'nin güvenilir olabilmesi için çok sayıda genom dizisi ve SNP (Şekil 4) verisi gerekmektedir. Her geçen gün önemli çiftlik hayvanı türlerinde çeşitli karakterlerin varyasyonlarından sorumlu olduğu tespit edilen spesifik SNP'ler tanımlanmaktadır. Sığırlarda şimdiye kadar 3 milyon SNP tanımlanmıştır. Tanımlanmış SNP'ler birçok yöntemle belirlenebilmekle birlikte günümüzde en yaygın uygulama alanı bulan yöntem çip teknolojisidir. Bu teknoloji ile on binlerce SNP'in populasyonlardaki durumları tespit edilebilmektedir. SNP çip, DNA bağlayan binlerce küçük noktadan oluşur ve her nokta spesifik bir SNP'den sorumludur. Küçük bir plastik ya da cam ile DNA'nın incelenmesine olanak veren bu teknoloji sayesinde SNP'ler ile karakterler ilişkilendirilmekte ve seleksiyonda önemli bir parametre olarak değerlendirilmektedir.

Sığırlarda, yaşadığı süre içerisinde normalde verebileceği yavru sayısından 3-4 kat fazla sayıda yavru almak süperovulasyon ve embriyo transferi gibi uygulamalar ile mümkün olmaktadır. Bir boğa ise doğal yöntemlerle ömrü boyunca 10-100 baş buzağı verebilmektedir. Suni tohumlama teknolojisi sayesinde bir boğa bir sürüdeki genetik ilerlemenin % 90'ından daha fazlasını etkileyebilir ve sürüde binlerce yavru ile temsil edilebilir. Bu nedenle, seleksiyon uygulamalarında genellikle baba hattının belirlenmesi amacıyla Progeny Testing (döl kontrolü) yapılmaktadır (Kappes, 1999; Koning, 2008; Bouquet ve Juga, 2013). PT uygulamasının sağlayacağı genetik ilerleme, yüksek genetik değere sahip boğaların doğru bir şekilde seçilebilmesi ve PT'nin etkili bir şekilde yapılabilmesine bağlıdır (Vishwanath, 2003).



Şekil 4. SNP: tek nükleotit polimorfizmi.

Üç SNP ile TGDD (Tahmini Genomik Damızlık Değer) hesabının basit bir örneği Tablo 1'de verilmiştir.

Tablo 1. Üç SNP ile TGDD (Tahmini Genomik Damızlık Değer) hesabının basit bir örneği (G: Genotip, D: Değer), (Jonas ve de Koning, 2015).

HAYVAN	SNP1		SNP2		SNP3		TYGD
	G	D	G	D	G	D	
1	AA	8	BB	-4	AA	2	6
2	AA	8	AA	4	BB	-2	10
3	AB	0	AB	0	AA	2	2
4	BB	-8	AA	4	AA	2	-2

Süt verimi ve kalitesi gibi ekonomik öneme sahip karakterlerin çoğu sadece dişi hayvanlarda ölçülebilmekte, seleksiyonda hedeflenen başarıya ulaşmada PT uygulamaları özellikle böyle karakterler için önemli avantaj sağlamaktadır (Schefers ve Weigel, 2012). PT uygulamaları ülkeden ülkeye, yetiştiriciden yetiştiriciye değişiklik göstermekle birlikte benzer prensiplerle uygulanmaktadır. Hedef, sağlıklı, en üstün verim kapasitesine sahip ve bu özellikleri en uzun süre sürdürebilen hayvanları doğru bir şekilde tespit edebilmektir. Uygulamada çok sayıda damızlık adayı boğa kullanılır. Kullanılan her bir boğanın kızlarının verim ortalamaları, teste giren bütün boğaların kızlarının ortalamalarıyla kıyaslanır (Schaeffer,2006; Özyurt, 2009; Tırpan ve Tekin, 2014).

Kanada'da Holstein ırkı bir sığır populasyonunda kullanılan PT yönteminde; bir boğanın bakım, besleme, sperma stoklanması gibi masrafları dahil maliyeti yaklaşık olarak 50000 dolar olmaktadır. Her yıl 500 boğa teste tabi tutulmakta, dolayısıyla yıllık 25 milyon dolar harcanmaktadır. Bir boğa için testin tamamlanma süresi 64 ayı bulmaktadır. 500 aday boğadan sadece en iyi 20 tanesi testi geçmekte, böylece bu 20 boğanın her birinin maliyeti 1,25 milyon dolar olmaktadır.

2.4. Genomik Damızlık Değer Hesaplama

Genomik damızlık değer hesaplama ile ilgili formüller eşitlik (3) ve (4) de verilmiştir. Matris gösterimi eşitlik (5) de verilmiştir.

$$y = Xb + \sum_{i}^{m} M_i g_i + e \tag{3}$$

$$y = Xb + \sum_{i}^{m} M_i g_i + W_u + e \tag{4}$$

Xb= sabit etki, M_i ve g_i= genomik etki, Wu=pedigri, e= hata.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}X & Z'R^{-1}Z \\ W'R^{-1}X & W'R^{-1}X & WR^{-1}W + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{bmatrix} \tag{5}$$

Örnek veri setleri Şekil 5'de gösterilmiştir. Eşitlik 6'daki matris kullanılarak elde edilen her bir hayvana ait GEBV değerleri matrisi Şekil 6'da te gösterilmiştir. Sonuç olarak hesaplanan damızlık değerleri Şekil 7'de gösterilmiştir.

Animal	Sire	Dam	Mean	EDC	Fat DYD	SNP Genotype								
13	0	0	1	558	9.0	2	0	1	1	0	0	0	2	1	2
14	0	0	1	722	13.4	1	0	0	0	0	2	0	2	1	0
15	13	4	1	300	12.7	1	1	2	1	1	0	0	2	1	2
16	15	2	1	73	15.4	0	0	2	1	0	1	0	2	2	1
17	15	5	1	52	5.9	0	1	1	2	0	0	0	2	1	2
18	14	6	1	87	7.7	1	1	0	1	0	2	0	2	2	1
19	14	9	1	64	10.2	0	0	1	1	0	2	0	2	2	0
20	14	9	1	103	4.8	0	1	1	0	0	1	0	2	2	0
21	1	3	1	13	7.6	2	0	0	0	0	1	2	2	1	2
22	14	8	1	125	8.8	0	0	0	1	1	2	0	2	0	0
23	14	11	1	93	9.8	0	1	1	0	0	1	0	2	2	1

Şekil 5. Örnek veri seti.

$$y = Xb + Zg + Wu + e \tag{6}$$

$$Z = \begin{bmatrix} 1{,}357 & -0{,}357 & 0{,}286 \\ 0{,}357 & -0{,}357 & -0{,}714 \\ 0{,}357 & 0{,}643 & 1{,}286 \\ -0{,}643 & -0{,}357 & 1{,}286 \\ -0{,}643 & 0{,}643 & 0{,}286 \\ 0{,}357 & 0{,}643 & -0{,}714 \\ -0{,}643 & -0{,}357 & 0{,}286 \\ -0{,}643 & 0{,}643 & 0{,}286 \end{bmatrix}$$

Şekil 6. Elde edilen her bir hayvana ait GEBV değerleri.

Mean effects	
	6,895
EBVs for animals with records	
13	3,114
14	1,697
15	4,200
16	3,842
17	2,861
GEBV for genotyped animals	
18	1,477
19	1,410
20	0,572
21	0,691
22	1,526
23	0,036
24	0,564
25	1,765
26	0,527

Şekil 7. Hesaplanan damızlık değerleri.

2.5. Genomik Seleksiyonun Avantajları

- Genetik kazancı arttırır.
- Seçim doğruluğunu arttırır.
- Generasyonlar arası süreyi kısaltır.
- Verim çağından önce hayvanlar seçilir.
- Döl Kontrolü ihtiyacını azaltır.

2.6. Genomik Seleksiyonun Dezavantajları

- Doğru marker tahminler için yeterince büyük bir hayvan grubunun fenotiplendirilmesi gerekmektedir.
- Daha az kalıtsallık, daha fazla kayıt gerektirir.
- Daha pahalıdır.
- Bazı türlerin haritaları tamamlanmamıştır.
- Generasyon aralığı zaten düşük olduğundan genetik seçimden dolayı genetik kazanç daha az olacaktır.

2.7. Başka Neler Yapılabilir?

- Genetik varyasyonun daha net resmini elde etmek

için daha fazla SNP gereklidir.
- Genotip için daha fazla hayvan kaydı gereklidir.
- Tahmin yöntemleri hassaslaştırılmalıdır.
- BLUP ve Bayesin ne zaman kullanılacağı belirlenmelidir.

3. Sonuç

GEBV Daha yüksek doğruluk ve daha kısa üretim aralığıyla daha hızlı genetik ilerleme mümkündür. Birçok ülkede uygulanmaktadır. Klasik ve genetik seleksiyon birleştirilmelidir. Son zamanlarda DNA çip teknolojisindeki bu ilerlemeler sayesinde moleküler genetik yöntemlerin hayvan genetiği ve ıslahı çalışmalarında kullanımı da artmıştır. Damızlık değerlendirmenin belirlenmesindeki isabet genomik değerlendirmeyle çok yükselmiştir. Hayvan populasyonlarındaki genetik varyasyonları belirlemek için DNA düzeyinde çalışmak çok daha isabetli sonuçlar vermektedir. Bu amaçla DNA polimorfizmlerini belirlemek için birçok yöntem kullanılmaktadır. Bunlar; SNP, Mikrosatelliter, RFLP, RAPD gibi yöntemlerdir. Bu teknolojiler gen fonksiyonlarının belirlenmesi, ebeveyn tayini ebeveyn doğrulama çalışmaları, çeşitli karakterlerle ilgili markerlerin tespiti, genom haritalarının çıkarılması gibi birçok konuda yaygın olarak kullanılmaktadır. Bu yöntemler içerisinde genetik ilerlemenin sağlanmasında SNP yöntemi diğer yöntemlere oranla daha fazla potansiyele sahiptir. SNP yöntemiyle damızlık değerlendirme sistemlerine moleküler düzeyde katkı sağlanmıştır. Geleneksel damızlık seçimi ile genomik seleksiyonun kombine edilmesi ve böylelikle genetik ilerleme hızının arttırılmasına yönelik çalışmalar büyük hız kazanmıştır. Damızlık seçiminde gelecekte SNP teknolojisinin tek alternatif olması, sürpriz bir gelişme olmayabilir.

Katkı Oranı Beyanı

Yazarların katkı yüzdeleri aşağıda verilmiştir. Yazarlar makaleyi incelemiş ve onaylamıştır.

	E.Ş.	S.G.
K	50	50
T	50	50
Y	50	50
VTI	50	50
VAY	50	50
KT	50	50
YZ	50	50
GR	50	50

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, GR= gönderim ve revizyon.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Kaynaklar

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics, 2*(2), 130–138.

Bal, O., & Akyüz, B. (2014). Halk elinde yetiştirilen holştayn, doğu anadolu kırmızısı ve yerli kara sığır ırklarında diacylglycerol O-Acyltransferase 1 (DGAT 1) gen polimorfizminin PCR-RFLP yöntemi ile belirlenmesi. *Erciyes Üniversitesi Veteriner Fakültesi Dergisi, 11*(1), 7–13.

Bouquet, A., & Juga, J. (2013). Integrating genomic selection into dairy cattle breeding programmes: a review. *Animal, 7*(5), 705–713.

Cole, J. B., & Silva, M. V. G. B. (2016). Genomic selection in multi-breed dairy cattle populations. *Revista Brasileira de Zootecnia, 45*(4), 195–202.

Daş, H. (2015). QTL tespiti için hayvanlarda kullanılan populasyonlar ve istatistiksel metotlar. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi, 4*(2), 27–297.

Dekkers, J. C. M. (2004). Commercial application of marker- and geneassisted selection in livestock: Strategies and lessons. *Journal of Animal Science, 82*(13, Ek), 313–328. https://doi.org/10.2527/2004.8213_supplE313x

Eggen, A. (2012). The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers, 2*(1), 10–15. https://doi.org/10.2527/af.2011-0027

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics, 10*(6), 381–391.

Gürses, M., & Bayraktar, M. (2014). Moleküler markerlerin hayvan yetiştiriciliği ve genetiğinde kullanımı. *Fırat Üniversitesi Sağlık Bilimleri Tıp Dergisi, 28*(2), 99–106.

Hayes, B. J. (2007). QTL mapping, MAS, and genomic selection. İçinde *A short-course organized by Animal breeding & Genetics* (ss. Sayfa aralığı eksik). Department of Animal Science, Iowa State University.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science, 92*(2), 433–443.

Hayes, B. J., Lewin, H. A., & Goddard, M. E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics, 29*(4), 206–214.

Jonas, E., & de Koning, D. J. (2015). Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. *Frontiers in Genetics, 6*, 49.

Kappes, S. M. (1999). Utilization of gene mapping information in livestock animals. *Theriogenology, 51*(1), 135–147.

Komisarek, J., & Dorynek, Z. (2009). Effect of ABCG2, PPARGC1A, OLR1 and SCD1 gene polymorphism on estimated breeding values for functional and production traits in polish holstein-friesian bulls. *Journal of Applied Genetics, 50*(2), 125–132.

Koning, D. J. D. (2008). Application of molecular information in sustainable animal breeding. *Revista Brasileira de Zootecnia, 37*, 122–126.

McGill, D., & Lievaart, J. J. (2011). Genomic selection in dairy cattle: A review and discussion on some possible applications. İçinde *Dairy Research Foundation Symposium* (ss. 117–122).

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819–1829.

Özbeyaz, C., & Kocakaya, A. (2011). Süt sığırlarında genomik değerlendirme (derleme). *Lalahan Hayvan Araştırma Enstitüsü Dergisi, 51*(2), 93–104.

Özdemir, M., & Doğru, Ü. (2008). Sığırların verim özellikleri üzerine etkili önemli moleküler markörler. *Atatürk Üniversitesi*

Ziraat Fakültesi Dergisi, 39(1), 127–135.

Özşensoy, Y., & Kurar, E. (2013). Genetik bağlantı analizi ve uygulama alanları. *Erciyes Üniversitesi Veteriner Fakültesi Dergisi, 10*(1), 53–62.

Özyurt, A. (2009). TİGEM Esmer popülasyonunda döl kontrolü (progeny testing) ve uygulama olanakları. *Tekirdağ Ziraat Fakültesi Dergisi, 6*(3), 257–264.

SAS Institute Inc. (2007). *QTL mapping, MAS, and genomic selection: A short-course organized by Animal breeding & Genetics* (B. J. Hayes). Department of Animal Science, Iowa State University.

Schaeffer, L. R. (2006). Strategy for applying Genome-wide selection dairy cattle. *Journal of Animal Breeding and Genetics, 123*(4), 218–223.

Schefers, J. M., & Weigel, K. A. (2012). Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers, 2*(1), 4–9.

Sharma, A., Lee, J. S., Dang, C. G., Sudrajad, P., Kim, H. C., Yeon, S. Y., Kang, H. S., & Lee, S. H. (2015). Stories and challenges of genome wide association studies in livestock- a review. *Asian-Australasian Journal of Animal Sciences, 28*(10), 1371. https://doi.org/10.5713/ajas.14.0715

Tırpan, M. B., & Tekin, N. (2014). Türkiye'deki progeny test çalışmalarına genel bir bakış. *Erciyes Üniversitesi Veteriner Fakültesi Dergisi, 11*(3), 197–203.