ISSN PRINT 2319 1775 Online 2320 7876

Research paper Journal Volume 11,Iss 12, 2022

LINEAR MIXED REGRESSION MODELS FOR AGE OF HEN AT OPTIMAL PRODUCTION OF EGGS

Alok Kumar Shukla, Vishwas Tiwari*
Department of Statistics,
D.A-V. College, Kanpur, Uttar Pradesh
drvishwastiwaridav@gmail.com
Surendra Kumar
Department of Mathematics,
Mahamaya Govt. Degree College Mahona, Lucknow, Uttar Pradesh
surendra.kumar776@gmail.com
Gagan Kumar,
Department of Economics,
Govt. Degree College Kuchlai, Sitapur, Uttar Pradesh
gkb.gdc@gmail.com
Priyanka Bharti
Department of Economics,
D.D.U. Govt. PG College Sitapur, Uttar Pradesh
pkb.gdc@gmail.com

*Corresponding Author

ABSTRACT

This research paper introduces a comprehensive method for predicting the age at which hens achieve optimal egg production using mixed linear regression models. By combining two distinct linear regression models and machine learning techniques, we aim to enhance the accuracy and reliability of age prediction for maximizing egg-laying efficiency in hens. The mixed linear regression model is built by integrating parameters estimated through ordinary least squares regression. Various models are derived by incorporating components from these two linear regression models, allowing for exploration of different prediction possibilities. Evaluation measures such as the coefficient of determination and residual sum of squares are utilized to assess model performance and identify the most accurate ones for predicting optimal egg production age. To validate the practical utility of the proposed models, authentic egg production data is used for verification. The predictive capabilities of the mixed linear regression models are extensively assessed against this real dataset, enabling selection of the most suitable model for practical application. In summary, this research demonstrates the effectiveness of the mixed linear regression approach in predicting optimal egg production age in hens, offering valuable insights for enhancing efficiency and productivity in poultry farming.

KEYWORDS: Regression models, Least square method, Fitting measures, Prediction, RMSE.

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11,Iss 12, 2022

1. Introduction

The poultry farming is an economically important for country because the fastest growing segments of agricultural sector in India. Egg production is a result of many genes through biochemical, anatomical, and physiological processes. The amount of egg lied by hens is influenced by many factors, such as breed or strain, age of birds, photo refractoriness, broodiness, moulting, nutrition, and other environmental factors. The hens do not produce eggs daily and have pauses and, therefore, the daily egg production is summarized weekly, biweekly or monthly scale *to* measure their capacity for egg production. When daily egg production is summarized on such a scale of time, egg production curve against time increases rapidly 10 a maximum and decreases subsequently to a minimum value at the end of the year. The number of egg normally increases to reach the peak in certain age, and then decrease to the end of laying periods. Prediction of egg production age as early as possible during the laying cycle using part records facilitates a poultry breeder to select breeding birds early thereby helping in reducing the cost of egg production/day-old chicks. Many statistical models to describe the egg production in laying hens have been published.

Acha (2010) had emphasized the need for more efficient production of birds. Ahmad (2011) and Cason (1990, 1991) had studied the egg production forecasting. Narinc *et al.* (2014) had emphasized the egg production curve analyses in poultry science. Sharifi *et al.* (2022) had study the conducted to evaluate 8 mathematical models for egg production and egg weight. Yakubu *et al.* (2018) had emphasized the modelling egg production by linear, quadratic, artificial neural network and classification regression tree methods. Many more authors including (2006), Misra and Gupta (2008, 2010), Misra *et al.* (2012), Gupta and Yadav (2017, 2018), Kumar *et al.* (2019), Kumar *et al.* (2021), Yadav *et al.* (2017-2019) worked on the estimation of the population parameters.

2. Suggested models

We are motivated by Nwogu & Acha (2014) to use polynomial regression model for optimal production of eggs. We take two linear models were used by Pandey and Kumar (2014) for nonlinear trend in milk data.

$$f(X_i) = A + BX_i + CX_i^2 + D\frac{1}{X_i} \tag{2.1}$$

$$f(X_i) = A + BX_i + C\frac{1}{X_i} + D\frac{1}{X_i^2} \tag{2.2}$$

The linear regression model is of the form

$$Y_i = f(X_i) + e_i \tag{2.3}$$

Y_i is the egg production (in grams per day), X_i is the age of hens (in weeks) and e_i is error.

ISSN PRINT 2319 1775 Online 2320 7876

2.1. Fitting of models

The parameters of the proposed models (2.1) and (2.2) can be directly obtained by the application of the famous least square method of estimation. For detailed description of estimation of parameters, the reference can be made of Draper & Smith (1998). The model (2.3) admits an additive error (e_i) term if error terms are independently and identically distributed random variable following normal distributions, the maximum likelihood estimators of the parameters can also be obtained.

3. Model Selection

Model selection is based Gujarati and Sangeetha (2007) and Montgomery et al. (2012) on the analysis of residuals has also been performed for checking model adequacy. Residuals are assumed to have zero mean, constant variance, no correlation and normal distribution or they are very near to zero mean almost a constant variance and negligible correlation and very near to normal distribution.

3.1 Goodness of fit of different models

Coefficient of Determination – R^2

The coefficient of determination R^2, which tells how much variation out of total variation is explained by the regression model. Thus R^2 expresses the how much of the part of the total sum of squares fall into the sum of squares due to regression.

The formula for R^2 is given by,

$$R^2 = 100\left(1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}\right)\%$$

Adjusted Coefficient of Determination – R^2_{adj} A more reliable measure of goodness of fit than the R^2 is the Adjusted Coefficient of Determination R^2_{adj}, since R^2 always increases as the number of terms or variables increases whether the variable is irrelevant while R^2_{adj} tells up to which extent the variables or the parameters should be used.

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11,Iss 12, 2022

$$R^2_{adj} = 100\left(1 - \left(\frac{n-1}{n-p}\right)\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}\right)\%$$

Where, n is the number of observations and p is the number of parameters of the model.

Mean Squared Error (MSE):

The residual mean square is defined as

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}$$

where n is the number of observations and p is the number of parameters of the model. y_i is the observed value of Y, is the predicted value from the fitted model $\hat{y}_i$. A small value of MSE reflects the appropriateness of the fitted model.

Akaike Information Criterion (AIC):

The Akaike Information Criterion (AIC) is calculated from

$$AIC = 2In(RMSE) + \frac{2p}{n}$$

where $RMSE$ is the root mean squared error during the estimation period, p is the number of estimated coefficients in the fitted model, and n is the sample size used to fit the model.

Hannan-Quinn Criterion (HQC):

The Hannan Quinn Criterion (HQC) is calculated from

$$HQC = 2\,In(RMSE) + \frac{2p\,In(In(n))}{n}$$

This criterion uses a different penalty for the number of estimated parameters.

Schwarz-Bayesian Information Criterion (SBIC):

The Schwarz-Bayesian Information Criterion (SBIC) is calculated from

$$SBIC = 2\,In(RMSE) + \frac{p\,In(n)}{n}$$

Again, the penalty for the number of estimated parameters is different than for the other criteria.

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11,Iss 12, 2022

Mallows' C_p statistic:

We calculated from

$$C_P = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{MSE(full)} - (n - 2p)$$

where *MSE (full)* is the mean squared error of the model when all independent variables are included in the fit. If a fitted model has little bias, C_p should be close to p. It is desirable to have a small C_p as long as the value is not much greater than p.

3.2 Model selection by significance of parameter

In this section first we calculate p-value of each parameter, then after remove higher p-value parameter form model. Fit regression line by rest of variable. Then we find best regression model.

3.3 Examination of Residuals

Analysis of the residuals (errors) is strongly recommended to decide about the suitability of a model by Draper and Smith (1998). Three important assumptions of the model are:

(i) Errors are not auto correlated.

(ii) Errors are independent.

(iii)Errors are normally distributed.

The assumptions can be verified by examining the residuals.

Test for auto correlation of errors (Durbin-Watson Test)

We test H_0 : Errors are not auto correlated (if DW test values > d_U)

Against H_1 : Errors are auto correlated (if DW test values < d_L)

where d_L and d_U are given in Draper and Smith (1998). DW Test values greater than 1.72 times d_U confirm that there is no problem of auto-correlation.

Test for independence of errors (Run Test)

We test H_0 : Errors are independent.

Against H_1 : Errors are not independent.

Test for normality (Shapiro-Wilk Test, n <50)

We test H_0 : Errors are normally distributed.

Against H_1 : Errors are not normally distributed.

4. Analysis of Models

We are analysis the models by publish data, given by Nwogu & Acha (2014).

4.1 Analysis of Models by Goodness of fit

Regression Model (2.1) written as

$$Y_i = A + B X_i + C X_i^2 + D\frac{1}{X_i}$$

Above model rewritten as $Y = A + B Z_1 + C Z_2 + D Z_3$

ISSN PRINT 2319 1775 Online 2320 7876

Journal Volume 11,Iss 12, 2022

Where $Y = Y_i$, $Z_1 = X_i$, $Z_2 = X_i^2$, $Z_3 = \frac{1}{X_i}$

Table 4.1: Model Results

R^2	R^2_{adj}	MSE	AIC	HQC	SBIC	Cp	Included Variables
18.8243	16.3644	8.8503	2.2947	2.3254	2.3836	261.668	Z_1
31.7913	29.7244	7.4366	2.1207	2.1513	2.2095	214.917	Z_2
2.94118	0.0	10.582	2.4991	2.5298	2.5880	328.052	Z_3
83.5269	82.4973	1.8521	0.7877	0.8337	0.9210	30.3917	Z_1, Z_2
91.1422	**90.5886**	**0.9959**	**0.1673**	**0.2133**	**0.3006**	**2.9354**	**Z_1, Z_3**
90.9727	90.4085	1.0149	0.1862	0.2323	0.3196	3.5466	Z_2, Z_3
91.4017	90.5696	0.9979	0.2264	0.2878	0.4042	4.0	Z_1, Z_2, Z_3

Regression Model (2.2) written as

$$Y_i = A + B X_i + C\frac{1}{X_i} + D\frac{1}{X_i^2}$$

Above model rewritten as $Y = A + B\, Z_1 + C\, Z_2 + D Z_3$

Where $Y = Y_i$, $Z_1 = X_i$, $Z_2 = X_i^2$, $Z_3 = \frac{1}{X_i}$

Table 4.2: Model Results

R^2	Adj. R^2	MSE	AIC	HQC	SBIC	Cp	Included Variables
18.8243	16.3644	8.8503	2.2947	2.3254	2.3836	253.245	Z_1
2.94118	0.0	10.5821	2.4991	2.5298	2.5880	317.718	Z_2
2.94118	0.0	10.5821	2.4913	2.5220	2.5802	315.012	Z_3
91.1422	**90.5886**	**0.9959**	**0.1673**	**0.2133**	**0.3006**	**2.0163**	**Z_1, Z_2**
89.6848	89.0401	1.1597	0.3196	0.3656	0.4529	7.1196	Z_1, Z_3
77.4309	76.0203	2.5375	1.1026	1.1486	1.2359	50.028	Z_2, Z_3
91.1469	90.2901	1.0275	0.2557	0.3170	0.4334	4.0	Z_1, Z_2, Z_3

4.2 Model selection by significance of parameter

(4.2.1). Again written as regression model of equation (2.1)

$$Y_i = A + B X_i + C X_i^2 + D\frac{1}{X_i}$$

Above model rewritten as $Y = A + B\, Z_1 + C\, Z_2 + D Z_3$

IJFANS
International Journal of
Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11,Iss 12, 2022

Table 4.3: Parameter Estimate

Parameter	Estimate	S E	T	P
CONSTANT	35.8074	7.6585	4.6754	0.0001
Z_1	-0.2168	0.1743	-1.2436	0.2229
Z_2	-0.0015	0.0012	-0.9672	0.3409
Z_3	-533.106	100.05	-5.3283	0.0000

Variable Z_2 remove from above model, because P-value is 0.3409, which is highest P-value.

(4.2.2) Again written as regression model of equation (2.2)

$$Y_i = A + B X_i + C \frac{1}{X_i} + D \frac{1}{X_i^2}$$

Above model rewritten as $Y = A + B Z_1 + C Z_2 + D Z_3$

Table 4.4: Parameter Estimate

Parameter	Estimate	S E	T	P
CONSTANT	43.8789	7.4202	5.9134	0.0000
Z_1	-0.3908	0.0563	-6.9302	0.0000
Z_2	-659.287	291.378	-2.2626	0.0308
Z_3	435.99	3413.28	0.1277	0.8992

Variable Z_3 remove from above model, because P-value is 0.8992, which is highest P-value.

5. Final appropriate regression model

We find appropriate regression model by using above methods for selection of good model.

$$Y_i = A + B X_i + C \frac{1}{X_i} \qquad (5.1)$$

This model is also used by Misra et al. (2009) in sampling theory.

Table 5.1: Parameter Estimate

Parameter	Estimate	S E	T	P
CONSTANT	42.9661	1.9661	21.8532	0.0000
X	-0.3842	0.0212	-18.1046	0.0000
$1/X$	-622.405	38.5068	-16.1635	0.0000

ISSN PRINT 2319 1775 Online 2320 7876

 Journal Volume 11,Iss 12, 2022

Table 5.2: Analysis of Variance

Source	S S	DF	M S	F	P
Model	327.921	2	163.96	164.63	0.0000
Residual	31.8693	32	0.9959		
Total	359.79	34			

Table 5.3: Examination of Residuals for best model

Test	Durbin-Watson Test	Run Test	Shapiro-Wilk Test
Test Values	1.08277 (0.0005)	18.49 (0.0000)	0.980068 (0.8174)

The equation of the fitted model is

$$\hat{Y}_i = 42.9661 - 0.3842\, X_i - 622.405 \frac{1}{X_i} \qquad (5.2)$$

However, to assess the adequacy of the fitted model for the study data, the basic statistics, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the residuals

$$e_i = Y_i - \hat{Y}_i \qquad (5.3)$$

Since the PACF ($\hat{\phi}_{kk}$) cuts after lag 1, the autoregressive process of order 1, AR (1),

$$e_i = \phi_0 + \phi_1\, e_{i-1} + a_i \qquad (5.4)$$

Table 5.4: Basic statistics of the residuals from fitted models

Residuals (e_i) from Eqn (5.3)					Residuals (a_i) from Eqn (5.4)				
Statistics		Lag k	ACF (r_k)	PACF ($\hat{\phi}_{kk}$)	Statistics		Lag k	ACF (r_k)	PACF ($\hat{\phi}_{kk}$)
Mean	0.0000	1	0.422	0.422	Mean	0.0000	1	-0.016	-0.016
Median	-0.0406	2	0.200	0.027	Median	0.0319	2	0.122	0.121
Skewness	-0.2302	3	-0.013	-0.129	Skewness	-0.5005	3	-0.097	-0.094
Kurtosis	-0.0752	4	-0.035	0.010	Kurtosis	0.7202	4	-0.003	-0.020
Std dev	0.9681	5	-0.116	-0.097	Std dev	0.8804	5	-0.078	-0.057
		6	-0.164	-0.105			6	-0.080	-0.090
		7	-0.179	-0.067			7	-0.117	-0.108
		8	-0.137	-0.033			8	0.014	0.018

ISSN PRINT 2319 1775 Online 2320 7876

Table 5.5: Estimates of Parameters of autoregressive model (5.4)

Parameter	Estimate	S E	T	P
Constant (ϕ_0)	-0.0319	0.1535	-0.2082	0.8364
ϕ_1	0.4542	0.1642	2.7648	0.0094

From Table 5.5 it is clear from the value that the constant (ϕ_0) is not significantly different from zero, while the coefficient of ϕ_1 is significant. Therefore, the final estimate of model is

$$e_i = 0.4542\, e_{i-1} + a_i$$

From equations (5.2) and (5.4), the model which describes the pattern of egg production as a function of age of hens becomes

$$\hat{Y}_i = 42.9661 - 0.3842\, X_i - 622.405 \frac{1}{X_i} + 0.4542\, e_{i-1}$$

6. Recommendation and Conclusion

The fitted model shows that the age of the hens at maximum is about 40.50 weeks. The corresponding egg production at this is about 12.04 grams per day. The hens were also found to be at their best, in terms of production of eggs when they are aged 32.5 weeks to 50.5 weeks, when their output is at least 11.72 grams per day. Hence, for optimal production of eggs, it is recommended that hens are not kept beyond 50.5 weeks.

References

[1] Acha, C.K. (2010). Impact of Housing and Age of Birds on Egg production. International Journal of Numerical Mathematics, 5 (2), 275 – 298.

[2] Ahmad, H. A. (2011). Egg Production Forecasting: Determining Efficient Modeling Approaches, J. Appl. Poult. Res. 20,463–473.

[3] Cason, J. A. (1990). Comparison of Linear and Curvilinear Decreasing Terms in Logistic Flock Egg Production Models. *Poultry Science* 69, 1467-1470.

[4] Cason, J. A. (1991). Egg Production Models for Molted Flocks. Poultry Science, 70, 2232-2236.

[5] Draper, N. R. and Smith, H., (1998). Applied Regression Analysis, 3rd Ed., John Wiley, New York

[6] Gujarati, D. N. and Sangeetha, (2007). Basic Econometrics, 4th Ed, Tata Mc Graw-Hill.

[7] Gupta, R.K. (2006). Estimation of Population Variance Using Ratio Type Estimator, Indian Journal of Mathematics and Mathematical Sciences, 2, Issue 2, 169-176.

[8] Gupta, R.K. and Yadav, S.K. (2017). New Efficient Estimators of Population Mean Using Non-Traditional Measures of Dispersion, Open Journal of Statistics, 7, 394-404.

ISSN PRINT 2319 1775 Online 2320 7876

[9] Gupta, R.K. and Yadav, S.K. (2018). Improved Estimation of Population Mean Using Information on Size of the Sample, American Journal of Mathematics and Statistics, 8(2), 27-35.

[10] Kumar, G., Bharti, P., Yadav, S.K. & Kumar, S. (2019). Improvement for estimating population mean in simple random sampling scheme, Journal of Reliability and Statistical Studies, 12 (1), 41-49.

[11] Kumar, S., Kumar, G., Saha, S., Doley, P. & Singh, L. (2021). Generalized Class of Population mean in Agriculture surveys using known Auxiliary Parameters and Sample Size, International Journal of Agricultural and Statistical Sciences, 17 (2), 617-623.

[12] Misra, S. & Gupta, R.K. (2008). Almost Unbiased Jackknifed Ratio type estimator of Population variance, International Journal of Agricultural and Statistical Sciences, 4, Issue 2, 345-350.

[13] Misra, G. C., Shukla, A.K. and Yadav, S.K. (2009). A Comparison of Regression Methods for Improved Estimation in Sampling. Journal of reliability and statistical studies, Vol. 2, Issue 2, 85-90.

[14] Misra, S. & Gupta, R.K. (2010). Estimation of Geometric Mean of Skewed Populations with and without using Auxiliary Information, Journal of Reliability and Statistical Studies, 3, Issue 2, 35-46.

[15] Misra, S., Gupta, R.K. and Shukla, A.K. (2012). Generalized Class of Estimators for Estimation of finite Population Variance, International Journal of Agricultural and Statistical Sciences, 8, No. 2, 447-458.

[16] Montgomery, D. C., Peck, E. A. and Vining, G. C., (2003). Introduction to Linear Regression Analysis, Wiley. 5th Ed., Wiley, 2012.

[17] Narinc, D., Uckardes, F. and Aslan, E.(2014). Egg Production Curve Analyses in Poultry Science. *World's Poultry Science Journal*, 70(12), 817-828

[18] Nwogu, E. C., and Acha, C. K. (2014). Age of Birds at Optimal Production of Eggs: A Polynomial Regression Analysis. Pakistan journal of statistics and operational research, Vol. 10, Issue 2, 205-215.

[19] Pandey, T. K. and Kumar, V. (2014). Systematic Sampling For Non-Linear Trend In Milk Yield Data, Journal of Reliability and Statistical Studies,7(1), 157-168.

[20] Sharifi, Mohd. A., Patil, Ch. S., Yadav, A. S. and Bangar, Y. Ch.(2022) Mathematical Modeling For Egg Production and Egg Weight Curves in A Synthetic White Leghorn. Poultry Science, 101,1-6.

[21] Yakubu, A., Oluremi1, O. I. A. and Ibrahim, Z. N. (2018). Modelling egg production in Sasso dual-purpose birds using linear, quadratic, artificial neural network and classification regression tree methods in the tropics. Livestock Research for Rural Development, Vol. 30, Issue 10, 205-215.

[22] Yadav, S.K., Singh, L., Mishra, S.S., Mishra, P.P., & Kumar, S. (2017). A Median Based Regression Type Estimator of the finite Population Mean, International Journal of Agricultural and Statistical Sciences 13 (1), 265-271.

ISSN PRINT 2319 1775 Online 2320 7876

[23] Yadav, S.S., Mishra, S.S., Kumar, S. & Kumar, R. (2018). New Efficient Predictive Estimator of Population Mean using Median of Study Variable, Jnanabha, 48 (1), 17-25.

[24] Yadav, S.K., Sharma, D., Yadav, A., Kumar, S. (2019), Optimal Searls Estimation of Population Variance under Systematic Sampling Scheme: A Simulation Study, International Journal of Computing Science and Mathematics, 16 (3), 239-251.