FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 5, May, 2025, pp 110 - 113
DOI: https://doi.org/10.33003/fjs-2025-0905-3656

EGG QUALITY ASSESSMENT: A MODEL COMPARISON APPROACH USING BAYESIAN MIXED LOGIT, MIXED LOGIT, LOGISTIC REGRESSION AND MULTINOMIAL REGRESSION MODELS

*[1]Amalahu, Christian Chinenye, [2]Nwabueze, Joy Chioma and [1]Ekeadinotu, Chibueze Barnabas

[1]Department of Mathematics, University of Agriculture and Environmental Sciences, Umuagwo, Imo State, Nigeria.
[2]Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

*Corresponding authors' email: chinenyechris2017@gmail.com

ABSTRACT

This study compares the performance of Bayesian mixed logit, mixed logit, logistic regression, and multinomial regression models in analyzing egg quality. The results show that the Bayesian mixed logit model outperforms traditional models, with egg weights, shell thickness, and shape index emerging as significant determinants of egg quality. The Bayesian mixed logit model's superior performance is evident in its lower AIC, DIC, RMSE, and MAE values. These findings have implications for the poultry industry, highlighting the importance of considering complex relationships between egg quality traits.

Keywords: Bayesian mixed logit, Egg quality, Mixed logit Model

INTRODUCTION

The quality of eggs is a critical factor in the poultry industry, affecting consumer health, marketability, and profitability Abd El-Azeem, *et al*.,(2023). Egg quality encompasses various characteristics, including shell integrity, yolk color, albumen quality, and internal contaminants (Jones & Musgrove, 2005). Egg quality has become a critical factor, enhancing both the efficiency of production processes and the market value of egg products. Broadly, egg quality is assessed in two categories, internal and external. Internal quality parameters include yolk and albumen height, yolk width, yolk weight, albumen width, and albumen length, all of which reflect the egg's structural and nutritional integrity. External quality parameters, such as egg weight, width, length, and shell thickness, contribute to the physical resilience and marketability of the egg Dilawar *et al*., (2021). These attributes play a pivotal role in optimizing production efficiency and influencing the egg's market appeal.

The poultry industry faces significant challenges in maintaining consistent egg quality standards, resulting in economic losses and food safety concerns. While various studies have applied statistical models to analyze egg quality, including classification and regression trees (CART), Kowalska, *et al*, (2021) and generalized linear mixed models (Jones & Musgrove, 2005), there is a need for a comprehensive comparison of different modeling approaches to identify the most effective method for predicting egg quality.

Recent studies have highlighted the potential of Bayesian modeling approaches in analyzing complex relationships in egg quality data Rosa, *et al*., (2018). However, a comparative analysis of Bayesian mixed logit, mixed logit, logistic regression, and multinomial regression models has not been conducted. This study aims to address this gap by comparing the performance of these four statistical models in analyzing egg quality.

Yildiz, *et al*., (2025) evaluated the effectiveness of machine learning algorithms in predicting quail egg quality based on nine key parameters, including egg weight, egg width, egg length, yolk height, yolk width, yolk weight, albumen height, albumen width, and albumen length. A dataset comprising 350 eggs from 18-week-old Japanese quails was analyzed using Logistic Regression, Naive Bayes, Support Vector Machines, k-Nearest Neighbors, Random Forest, and Gradient Boosting. Their findings showed that models combining internal and external quality parameters achieved significantly higher accuracy compared to models based solely on external attributes. Notably, Random Forest and Gradient Boosting algorithms achieved accuracies exceeding 97%, while predictions based only on external parameters exhibited lower accuracy but presented a promising starting point for non-invasive evaluations. Their study strongly highlights the applicability and flexibility of machine learning in evaluating quail egg quality. Their findings demonstrated that machine learning technologies have the potential to drive innovative approaches in the poultry industry and inspire future research focusing on larger datasets and additional parameters to further enhance accuracy.

Egg weight prediction can be successfully realized from external egg traits using different statistical methods (Orhan *et al*., 2016; Çelik *et al*., 2017). Predictive estimates and evaluation of the relationship between traits of interest are commonly performed using multiple linear regression analysis (MLR); however, these analyzes can be affected by problems of multicollinearity (high correlation between variables), causing errors in the interpretation of the results Shafey *et al*., (2014). In face of this situation, it is important to complement the estimates done using MLR analysis with more efficient statistical procedures that avoid multicollinearity.

Thobela and 6enol (2024), investigated egg quality characteristics affecting egg weight of Lohmann Brown Hen with Data Mining Methods. They used Random forest (RF), multivariate adaptive regression spline (MARS), classification and regression trees (CART), bagging MARS, chi-square automatic interaction detector (CHAID), and exhaustive CHAID in egg weight (EW) prediction from selected egg quality characteristics in chicken. A total of 400 egg weight (EW), egg length (EL), egg width (EWD), shell weight (SW), yolk weight (YW), and albumen weight (AW) predictors were turned into account. They used goodness-of-fit criteria to select the best model to estimate Lohman Brown hen egg weight. They separated their data set into train and test for validation through a 10-fold cross-validation. They found the most significant EW predictors were albumen weight, egg width, and egg length. The correlation coefficient (r) value ranged from 0.957 (CHAID) to 0.99999 (MARS and Bagging MARS). The lowest RMSE (0.001) was found for MARS and bagging MARS algorithms and the highest (2.154) was obtained for CHAID. In general, the implemented algorithms excellently predicted the EW of hens. The ascertainment of the egg quality characteristics associated with EW using data mining algorithms were considered an indirect selection criterion for further chicken breeding programs.

Yilmaz and Çelik (2021) modeled the performance of the binary logistics regression model based on the Bayesian estimates. They compared the result of the model based on the ML estimates and used two information criteria such as BIC and AIC for their comparison. Their results showed that for small sample size, that the Bayesian method showed better performance than the Maximum Likelihood method based on the goodness of fit statistics. It is observed that decision-makers' heterogeneity was not explored. They use the following binary logistic regression model given in equation (1) to analyze their data set:

$$log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \varepsilon, \quad i = 1,2,3\ldots 34 \qquad (1)$$

where the dependent variable is European Union (EU) membership (member:1, not member:0) and independent variables are total number of people living, ratio of female parliamentarians $(x_{1,} x_{2,} \ldots, x_8)$. We therefore explicitly improve their work by comparing ours with other models.

MATERIALS AND METHODS

The method adopted in this study is the Bayesian Dynamic Mixed Logistic Regression Model.

Mixed Logit

In Train, (2003), like any random utility model of the discrete choice family of models, we assume that a sampled individual (q=1,…,Q) faces a choice amongst *I* alternatives in each of *T* choice situations. An individual q is assumed to consider the full set of offered alternatives in choice situation *t* and to choose the alternative with the highest utility. The (relative) utility associated with each alternative *i* as evaluated by each individual q in choice situation *t* is represented in a discrete choice model by a utility expression of the general form.

$$U_{itq} = \beta_q X_{itq} + e_{itq} \qquad (2)$$

X_{itq} is a vector of explanatory variables that are observed by the analyst. t, β_q and e_{itq} are not observed by the analyst and are treated as stochastic influences.

We model β_q as a random variable with density $f(\beta/\theta)$where $\theta$are the fixed parameters of the distribution ofβ. If we did knowβ_q , then the model would be a standard logit with the conditional choice probability

$$L_{qi}(\beta_q) = \frac{e^{\beta_q^/ x_{qi}}}{\sum_{j=1}^{J} e^{\beta_q^/ x_{qj}}} \qquad (3)$$

Since β_q is not given, so we have to integrate over the density of the random coefficients to obtain the unconditional choice probability

$$P_{qi} = \int \frac{e^{\beta_q^/ x_{qi}}}{\sum_{j=1}^{J} e^{\beta_q^/ x_{qj}}} f(\beta/\theta) d\beta$$

$$P_{qi} = \int L_{qi}(\beta_q)\, f(\beta/\theta) d\beta \qquad (4)$$

Models of this form are called *mixed logit* because the choice probability $L_{qi}(\beta_q)$ is a mixture of logits with $f(\beta/\theta)$ as the mixing distribution.

The presence of a standard deviation of a β parameter accommodates the presence of preference heterogeneity in the sampled population. This is often referred to as unobserved heterogeneity.

The Bayesian mixed logit model can be specified as:

$$P(Y = 1|X) = \int\int \left[exp(\beta X) / \left(1 + exp(\beta X)\right)\right] p(\beta|\theta) p(\theta) d\theta d\beta$$

where:

$p(\beta|\theta)$ is the conditional distribution of β given θ (this is the mixing distribution), $p(\theta)$ is the prior distribution on θ (this is where the Bayesian part comes in). By incorporating the prior distribution$p(\theta)$, we're adding a Bayesian layer to the model. This allows us to update our beliefs about the model parameters θ using Bayesian inference.

RESULTS AND DISCUSSION

This section presents the results of the egg quality data obtained from Chris Farms to examining the performance of the Bayesian Mixed Logit Model, Mixed Logit Model, Logistics Regression Model and Multinomial Model. The Software implementation was Python (using PyMC3).

Table 1: Posterior Distribution of Unknown Parameter θ Using Gamma Prior

Parameter	Mean	Std. Dev.	Equal-tailed	[95% Cred. Interval]
Intercept	-0.02	1.19	-2.39	2.35
Egg weight coefficient	0.038	0.012	0.018	0.061
Egg shell thickness coefficient	0.241	0.045	0.154	0.331
Shape index coefficient	0.551	0.098	0.362	0.743
Variance of farm id	1.06	0.231	0.661	1.503

Table 1 above revealed that for β_1, a unit increase in egg weight is associated with a 0.038 unit increase in quality while for β_2, a unit increase in egg shell thickness is associated with a 0.241 unit increase in quality and for β_3, a unit increase in shape index is associated with a 0.551 unit increase in quality. The estimated log-marginal likelihood of the model of the Bayesian mixed logit model for Egg quality data is -69.51.

Table 2: Mixed Logit Model with Egg Quality Data

Parameter	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Egg weight	0.048	0.016	3.04	0.002	0.017	0.079
Egg shell thickness	0.285	0.058	4.92	0.000	0.172	0.401
Shape index	0.617	0.119	5.19	0.000	0.384	0.853
Cons	-0.023	1.192	-0.02	0.984	-2.391	2.345

The coefficients represent the change in the log-odds of egg quality for a one-unit change in the predictor variable, holding all other variables constant. The standard errors and p-values indicate the precision and significance of the estimates while the 95% confidence intervals provide a range of plausible values for the coefficients and log likelihood of -69.51.

Table 3: Logistic Regression Model with Egg Quality Data with

Parameter	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Egg Weight	1.042	0.013	3.17	0.002	1.016	1.069
Egg Shell Thickness	1.278	0.059	5.33	0.000	1.143	1.421
Shape Index	1.697	0.159	5.63	0.000	1.323	2.176

Log likelihood = -63.21

Table 4: Multinomial Regression Model with Egg Quality Data

Parameter	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
.	Quality 2					
Egg weight	0.053	0.021	2.54	0.011	0.013	0.094
Egg shell thickness	0.317	0.074	4.29	0.000	0.172	0.463
Shape index	0.684	0.151	4.53	0.000	0.388	0.981
cons	-0.235	1.543	-0.15	0.879	-3.261	2.791
	Quality 3					
Egg weight	0.081	0.031	2.63	0.009	0.021	0.142
Egg shell thickness	0.462	0.108	4.28	0.000	0.251	0.674
Shape index	1.039	0.221	4.69	0.000	0.606	1.473
cons	-0.542	2.151	-0.25	0.802	-4.761	3.677

The coefficients represent the change in the log-odds of moving from the base category (quality=1) to the respective category (quality=2 or quality=3) for a one-unit change in the predictor variable, holding all other variables constant.

Table 5: Comparing Bayesian Mixed Logit with Mixed Logit, Logistic Regression and Multinomial Regression Models

Model	Log Likelihood	AIC	BIC	DIC	RMSE	MAE
Bayesian Mixed Logit	-69.51	151.02	173.15	147.91	0.199	0.109
Mixed Logit	-71.23	154.46	178.59	151.31	0.205	0.113
Logistic Regression	-83.15	174.31	198.44	169.19	0.231	0.135
Multinomial Regression	-135.19	283.38	311.19	278.42	0.289	0.173

The results of this study demonstrate the superiority of the Bayesian mixed logit model in analyzing egg quality. The model's ability to account for heterogeneity and complex relationships between variables makes it well-suited for this type of analysis. The significance of egg weights, shell thickness, and shape index as determinants of egg quality is consistent with previous research (Jones & Musgrove, 2005). Bayesian Mixed Logit model has the best fit and accounts for random effects and heterogeneity while Mixed Logit model is similar and closer to Bayesian mixed logit, but without Bayesian estimation. Logistic Regression model which is simplistic model, assumes binary outcome where Multinomial Regression model accounts for multiple categories, but assumes independence and has a poor fit. The Bayesian mixed logit model outperforms other models in terms of log likelihood, AIC, BIC, DIC, RMSE and MAE. The Bayesian mixed logit model outperforms traditional models, with egg weight, shell thickness, and shape index emerging as significant determinants of egg quality. The model's ability to provide accurate predictions and identify significant determinants of egg quality makes it a valuable tool for the poultry industry. The results of this study have implications for the poultry industry, particularly in terms of breeding and selection programs. By identifying key determinants of egg quality, producers can make informed decisions about breeding and selection strategies to improve egg quality.

CONCLUSION

This study demonstrates the effectiveness of the Bayesian mixed logit model in analyzing egg quality. The model's superior performance and ability to identify significant determinants of egg quality make it a valuable tool for the poultry industry. The findings of this study highlight the importance of considering complex relationships between egg quality traits and demonstrate the potential of Bayesian mixed logit models for improving egg quality analysis.

REFERENCES

Abd El-Azeem, N. A. , Madkour, M., Hashem, N. M., & Alagawany, M. (2023). Early nutrition as a tool to improve the productive performance of broiler chickens. *World's Poultry Science Journal*, 80(1), 171-185. https://doi.org/10.1080/00439339.2023.2262443

Çelik S, Sengül T, IncI H, Söðgüt B, Sengül AY, Kuzu Ç, (2017). Estimation of egg weight from some external and internal quality characteristics in quail by using various data mining algorithms. *Indian Journal of Animal Sciences*. 87:1524-1530. DOI:10.56093/ijans.v87i12.79871

Dilawar MA, Mun HS, Rathnayake D, (2021). Egg quality parameters, production performance and immunity of laying hens supplemented with plant extracts. *Animals*; 11(4):975. https://doi.org/10.3390/ani11040975

Jones, D. R., & Musgrove, M. T. (2005). Effects of Extended Storage on Egg Quality Factors. *Poultry Science* 84(11):1774-7, http://dx.doi.org/10.1093/ps/84.11.1774

Kowalska, E., Kucharska-Gaca, J., Kuźniacka, J. *et al.* Egg quality depending on the diet with different sources of protein and age of the hens. *Sci Rep* 11, 2638 (2021). https://doi.org/10.1038/s41598-021-82313-1

Orhan H, Eyduran E, Tatliyer A, Saygici H.(2016). Prediction of egg weight from egg quality characteristics via ridge regression and regression tree methods. *Revista Brasileira de Zootecnia*. 45:380-385. https://doi.org/10.1590/S1806-92902016000700004

Rosa J.O., Venturini G.C., Chud T.C.S., Pires B.C., Buzanskas M.E., Stafuzza N.B., Furquim G.R., Cruz V.A.R., Schmidt G.S., Figueiredo E.A.P., Lima V.F.M.H., Ledur M.C., Munari D.P. (2018): Bayesian inference of genetic parameters for reproductive and performance traits in White Leghorn hens. *Czech J. Anim. Sci*., 63, 230–236. https://doi.org/10.17221/116/2017-CJAS

Shafey TM, Mahmoud AH, Abouhheif MA.(2014). Dealing with multicollinearity in predicting egg components from egg weight and egg dimensión. *Italian Journal of Animal Science*. 13(4). http://dx.doi.org/10.4081/ijas.2014.3408

Thobela Louis Tyasi1 & 6enol &elik (2024). Investigation of Egg Quality Characteristics Affecting Egg Weight of Lohmann Brown Hen with Data Mining Methods. *Poultry Science Journal* 12(1): 107-117. https://doi.org/10.22069/psj.2024.21337.1934

Train, K., (2003). Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, MA.

Yildiz BI, Eskioğlu K, Özdemir D, Akşit M. (2025). Predicting Quail Egg Quality Using Machine Learning Algorithms. *Braz. J. Poult. Sci*. 27 (1). https://doi.org/10.1590/1806-9061-2024-2037

Yilmaz, A.and1, Eray Çelik, I. E. (2021). A Bayesian Approach to Binary Logistic Regression Model with Application to OECD Data. Yüzüncü Yıl ÜniversitesiFen Bilimleri Enstitüsü Dergisi Cilt 26, 94-101. https://doi.org/10.53433/yyufbed.837533