Veterinary and Comparative Biomedical Research
Online ISSN: 3060 – 7663
Original Research
Doi:10.22103/Vcbr.2024.23250.1013

Prediction of chicken gender before putting eggs in incubator using logistic regression model

A. Parchami [1], R. Aghaei Afshar [1], E. Fadaie [2], H. Ghaemi [3], E. Sakhaee [4]*

[1] *Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran*
[2] *Department of Clinical Sciences, Faculty of Veterinary Medicine, Baft branch, Islamic Azad University, Baft, Iran*
[3] *Mahan Chicken Meat Production Complex, Public Joint Stock Company,Mahan, Kerman, Iran*
[4] *Department of Clinical Sciences, Faculty of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman, Iran*

Correspondence:

Author email:
a_parchami@yahoo.com

Article history:

Received: 05 February 2024
Revised: 30 March 2024
Accepted: 10 April 2024
Published: 17 June 2024

Keywords:

Egg gender prediction
Logistic regression
Sharpness index
Statistical prediction

Abstract This paper propose several regression models to predict the probability of hatch a cockerel from egg before it is even putted in an incubator. The prediction of egg's gender is based on the minor and major diameters of eggs as the explanatory variables, which are simply measureable. A binomial logistic regression model is fitted to the observed length of minor and major diameters for 60 eggs to predict their genders. In this study, we achieve a simple statistical classifier model to classify eggs into male and female classes. In other words, we found that the more pointed eggs would be cockerel with high probability and the more spherical eggs would be pullet with high probability. After introducing the thin index for eggs, a simple linear regression model and another binomial logistic regression model are fitted based on the computed thin index data. Based on goodness-of-fit criteria AIC, BIC, chi-square test and so the difference between null deviance and residual deviance, the best model is determined among three fitted regression models. All estimated criteria implied that the proposed logistic regression with considering thin index as the explanatory variable is the best-fitted model to the observed data for the egg gender prediction. Moreover, the regression analyses on the physical shape variables are provided for each three-fitted model. The advantages and merits of the proposed logistic regression model are simplicity, cheep, fast and finally the proposed method has economic benefits for the user.

Introduction

Chicken gender is investigated mostly by large commercial hatcheries to separate female chicks (destined to lay eggs for commercial sale) from the male chicks (which are considered as unwanted byproduct of egg production and are killed and disposed of shortly after birth). Considering the moral importance of the open problem of chicken gender prediction, we start to research on this open problem, since we found that approximately 2 billion unwanted one day-old cockerel chicks killed every year in the world. Recently, Galli et al. have investigate on the chicken gender prediction in Germany. They produce a window in the egg shell to accessed the extra embryonic blood circulation system and to illuminate the flowing blood with a near-infrared laser. They success to determine the gender of eggs in day 3.5 of incubation process by the characterization of the fluorescence spectrum [3], [5], [9]. However, this study aimed to investigate on the prediction of chicken gender before putting egg in incubator. To this aim, three regression models are developed in this study based on the appearance characteristics of eggs in order to predict the chicken gender before putting egg in incubator with an acceptable accuracy and sensitivity. The organization of this paper is as follows: The collected date and some needed methods are briefly stated in Section II. Three regression models are fitted to the available data in Section III. Moreover, comparison between three fitted models is presented in Section III on the basis of several goodness-of-fit criteria. Conclusion part is given in the final section. All calculations and statistical analyses were performed using the R software environment in this paper [9].

Materials and Methods

Data and summary statistics

The main goal of this study is the egg gender prediction regarding to the physical shape of egg before even placed in an incubator. After random selection of sixty fertilized domestic chicken eggs, the eggs were labeled by a marker, see Figure 2. Then, the length of major-diameter (*A*) and the length of minor-diameter (*B*) were measured by caliper and recorded in Table 2 for each egg. The eggs were transferred to the incubator. After nineteen days, before the eggs were opened, each egg was placed in a separate compartment to allow labeling of the chickens' feet according to their respective egg label after the eggs were opened. Then on the 21st day, twenty-eight chicks were born and tagged on their feet. After approximately three months, unfortunately, a number of chicks died and finally only fourteen chicks (four hen and ten rooster) remained. As mentioned earlier, the considered variable to predict the gender of eggs are the length and the diameter of egg which are respectively denoted by *A* and *B* in Figure 1 and their observed values (in term of centimeter) are recorded in Table 2 for 60 eggs. In fact, each egg has a three-dimensional shape and it must has three radiuses r1 > r2 > r3. However, considering the fact that all of vertical egg cuts are circle (and are not elliptic, i.e. r2 □ r3) we can consider only two radiuses, named major-diameter/length and minor-diameter/diameter, to measure the physical shape specification by caliper for each egg. Unfortunately, twenty-eight chicks to be born and approximately three months later only fourteen hens/roosters are finally remained for precise gender determination. Therefore, we miss the gender of 46 sampled eggs, which are denoted by NA in Table 2. First, we are going to fit three different regression models based on 14 complete data, and then predict the genders of 46 non-complete data based on the fitted models. In addition, it must be mentioned that the gender of female and male chicks are shown by 0 and 1 in the Gender column of Table 2, respectively.

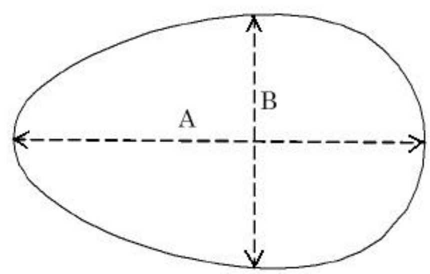


Fig. 1: Collected size data of each egg: length (*A*) and diameter (*B*).

Fig. 2: (Left) Sixty labeled eggs for the experiment, (Right) one-week-old chicks.

Fitting binomial logistic regression

Logistic regression is used when the dependent variable is categorical and it was applied in biological sciences, animal sciences and social sciences during twentieth century. Binomial logistic regression describes the relationship between a dichotomous response variable (*Y* , e.g. dead/alive, failure/success, lose/win, sick/healthy, or specially in this research, female/male of an egg) and a set of explanatory variables $X_1,...,X_m$ by

$$\text{logit}(\hat{p}) = Ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_m x_m, \quad (1)$$

where $\hat{p}$ is the probability of occurrence event *Y* = 1 (e.g., the probability of hatch a cockerel from egg), $\hat{\beta}_0$ is the intercept and $\hat{\beta}_i$ is the coefficient for the independent variables X_i estimated by maximum likelihood for *i* = 1, ..., *m*. Note that Eq. (1) show a linear relation between the natural logarithm of the odds ratio $Ln(\hat{p}/(1-\hat{p}))$ and the independent variables. Hence after fitting the observed data to logistic regression model (1), the probability of alive / success / win / healthy, or specially in this research the probability of hatch a cockerel from egg, can be written as

$$P(Y=1) = \hat{p} = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_m x_m)}}. \quad (2)$$

The interested readers can refer to the references [2], [4], [7] and [8] to see several applications of logistic regression model in biosciences.

Explanatory variable selection by pairwise correlations

Considering the dichotomous response variable gender, let we are going to predict the probability of hatch a cockerel from egg, i.e. $\hat{p} = P(Gender = 1)$, on the basis of the physical shape of egg. We found that the gender variable has a positive correlation with the length variable, and at the same time, has a negative correlation with the diameter variable. Therefore, index $C = A/B$ can be considered as the sharpness index or thin index for each egg based on its major-diameter (*A*) and minor-diameter (*B*) which its values are determined in Table 2 for 60 eggs. To select the most correlated variables with the binary gender variable, all possible pairwise correlations are calculated in Table 1 In addition, several descriptive charts for explanatory variables are shown in Figure 3. In the next section, the association between the sharpness index, the length of minor and major diameters and gender for eggs are determined by two different logistic regression models.

Table 1: Pairwise correlations.

Variables	*A*	*B*	S.I.	G.
Length (*A*)	1.00			
Diameter (B)	0.28	1.00		
Sharpness index (C)	0.85	-0.27	1.00	
Gender	0.34	-0.34	0.52	1.00

A: Length, *B*: Diameter, *S.I.*: Sharpness Index, *G*: Gender

Results and discussion

At the first of this section, three regression models are fit to the available chicks collected data in Table 2 to investigate the relation between the physical shape variables and the gender variable of eggs.

Model 1: The fitted simple linear regression model to predict the probability of hatch a cockerel by the value of the sharpness index is

$$\hat{p} = -4.89 + 4.23C \quad (3)$$

in which the probability of hatch a cockerel from egg is denoted by $\hat{p}$.

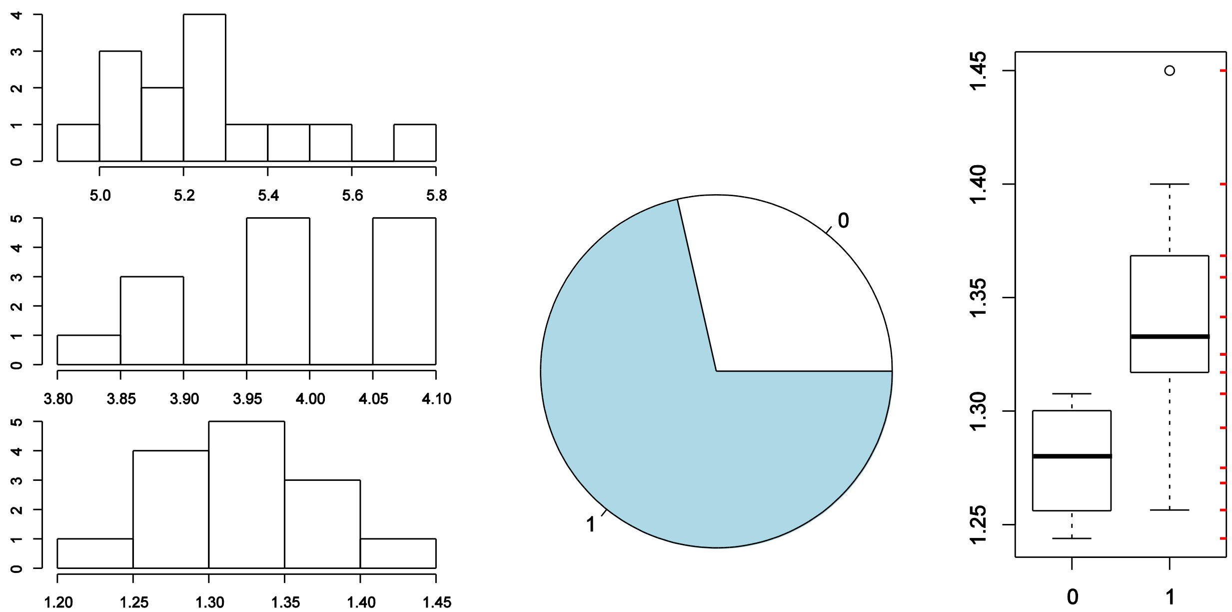
Model 2: The fitted logistic regression model to predict the probability of hatch a cockerel by the value of the sharpness index is

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -47.93 + 37.44C \quad (4)$$

Model 3: The fitted logistic regression model to predict the probability of hatch a cockerel by the values of the explanatory variables minor and major diameters is as follow

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right)=32.76+9.50\,A-20.30\,B \qquad (5)$$

The explanatory variable in models 1 and 2 is the sharpness index and so we can plot both of them in one figure. Simple linear regression Model 1 versus the logistic regression Model 2 based on the collected sharpness data C is shown in Figure 4.



(a) (b) (c)

Fig. 3: (a) Histograms of the collected major-diameters (*A*), minor-diameters (*B*) and sharpness $C=\frac{A}{B}$ (from top to bottom),
(b) Pie chart for the gender of chicks; 1 for rooster and 0 for hen,
(c) Box plots of the collected thin index *C*; 1 for rooster and 0 for hen.

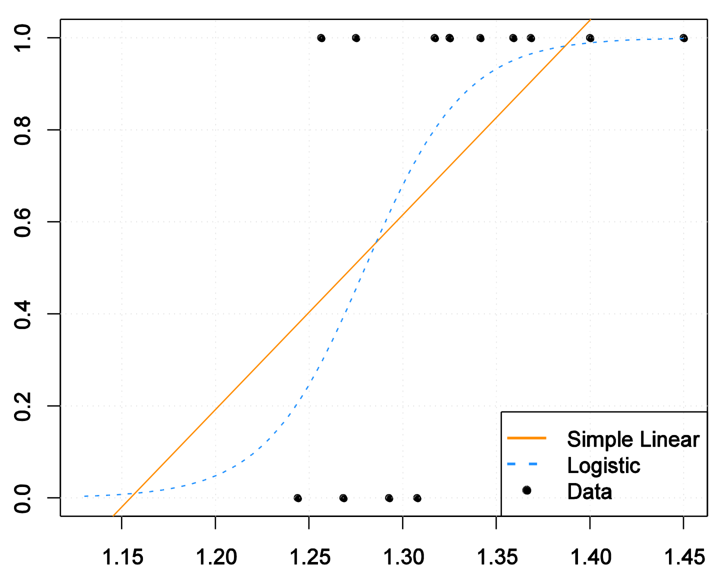


Fig. 4: Simple linear regression versus the logistic regression Model 1 based on the collected sharpness data.

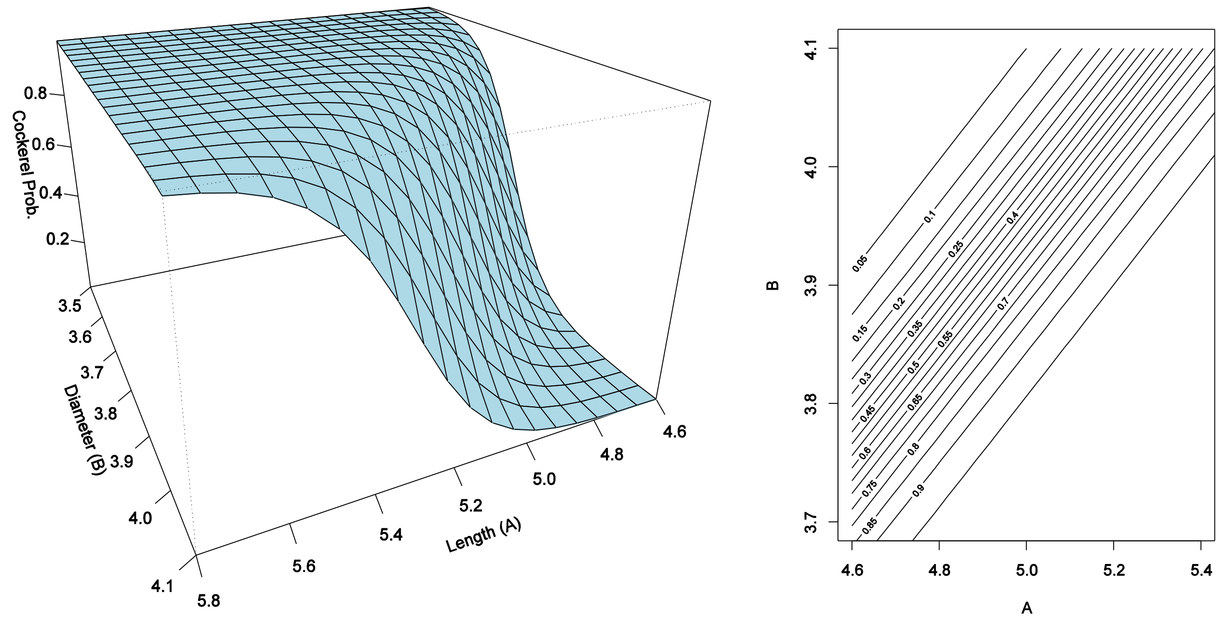


Fig. 5: 3D surface plot (Left) and contour plot (Right) for the probability of cockerel based on the logistic Model 2 in Eq. (5).

Both models 1 and 2 are shown when *C* increases the probability of hatch a cockerel from egg increases and obviously the probability of pullet decreases. Moreover, a three-dimensional graph for the probability of cockerel, with its contour plot are shown in Figure 5 using the fitted logistic Model 3 in Eq. (5). This three-dimensional surface shown that increasing the major diameter of egg can cause increasing the probability of hatch a cockerel, and also, decreasing the minor diameter can cause increasing the probability of hatch a cockerel.

Predicted probabilities from the fitted models.

As presented in Section 2, we miss the gender of 46 sampled eggs, which are denoted by NA in Table 2. Now, we are going to predict the probability of hatch a cockerel from 46 missed eggs by equations

$$\hat{p} = \begin{cases} 1 & \text{if} \quad C > 1.392 \\ -4.89 + 4.23C & \text{if} \quad 1.156 \le C \le 1.392 \\ 0 & \text{if} \quad C < 1.156 \end{cases} \quad (6)$$

$$\hat{p} = \frac{1}{1 + e^{47.93 - 37.44C}} \quad (7)$$

and

$$\hat{p} = \frac{1}{1 + e^{-32.76 - 9.5A + 20.3B}} \quad (8)$$

for three fitted models 1, 2 and 3, respectively. The prediction does not need to the recorded genders and therefore the predicted probabilities are computed for all 60 eggs in Table 2.

It is obvious that the probability of correct diagnosis is equal to $\hat{p}$ if the egg has a cockerel (Gender = 1) and it is equal to $1 - \hat{p}$ if it has a pullet (Gender = 0). So, for 14 alive chicks the probability of correct diagnosis are calculated in Table 2 for which their genders are known. The means of these calculated probabilities are computed at the bottom of 2 which can be used as a new criteria to compare the fitted models. Therefore, according to these new goodness-of-fit criteria, Model 3 can be considered as the best among three presented models in Eqs. (3) - (5).

Figure 6 shown the predicted probability of hatch pullet from eggs based on the fitted logistic Model 2 in which the number of eggs with specified gender are labeled. Based on three fitted regression models and using the computed information in Table 2, the following box plots has been drawn in Figure 7: (1) box plots of the predicted probabilities of hatch pullet from eggs (2) box plots of the probabilities of correct diagnosis, and (3) box plots of residuals. In addition, the mean value of each variable is shown by a red point in Figure 7 for all plots.

Tables 3, 4 and 5 are contains the summary for regression analyses for physical shape variables predicting gender based on three fitted models 1, 2 and 3, respectively.

Goodness-of-fit criteria for model selection

R-squared is the percentage of the dependent variable variation that a linear model explains. The R-squared for simple linear Model 1 is equal to 0.269 which reveals that only approximately 27% of the data fit the regression model. This low percentage of R-squared indicates that the linear Model 1 is not statistically significant.

Akaike information criterion (AIC) is a fined technique based on in-sample fit to estimate the likelihood of a fitted model to predict the future values. Bayesian information criterion (BIC) is another criteria for model selection that measures the trade-off between fitted model and its complexity. The AIC [1] and BIC [10] are formally defined by $AIC = -2\ln(\hat{L}) + 2k$ and $BIC = -2\ln(\hat{L}) + \ln(n)\,k$, where $\hat{L}$ is the maximized value of the likelihood function of the model, n is the sample size and k is the number of estimated parameters by the model. A good model is one that has minimum *AIC* or *BIC* among

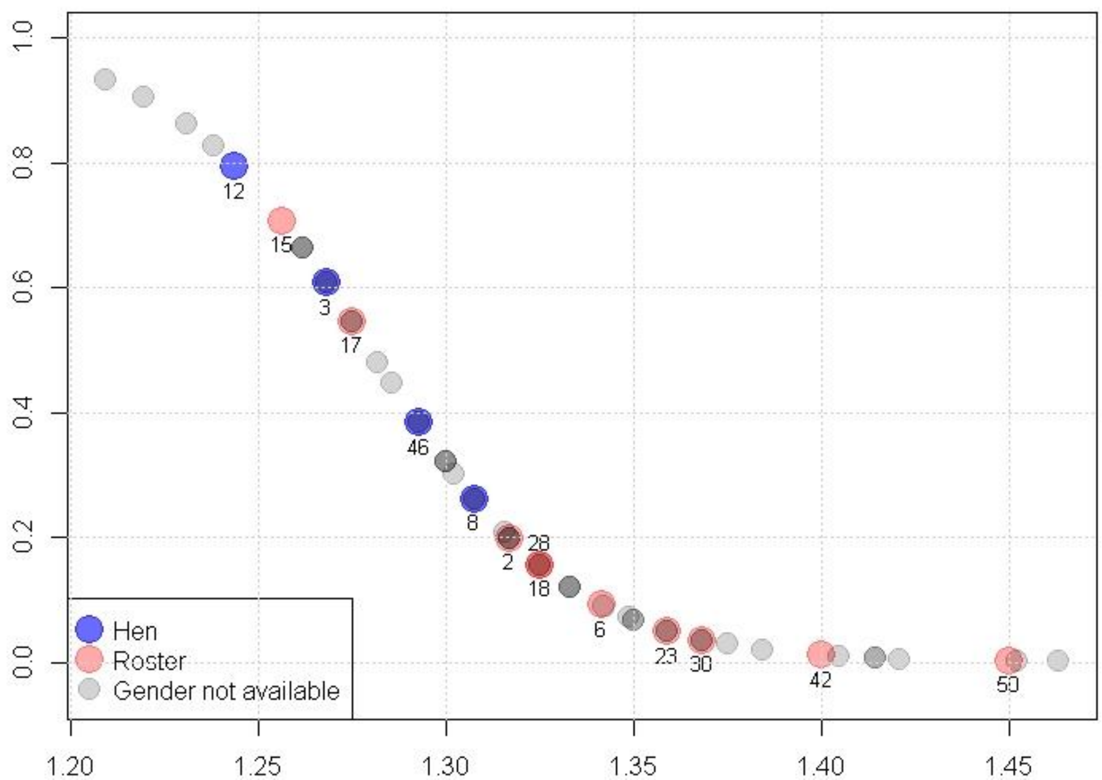


Fig. 6: The predicted probability of hatch pullet from eggs based on the fitted logistic Model 2.

Table 2: Recorder values of explanatory variables, the probability of hatch a cockerel, the probability of correct diagnosis and residuals under three proposed models in Eqs. (3)-(5) for 60 eggs.

	Explanatory variable				Probability of cockerel			Probability of correct diagnosis			Residual		
Egg #	*A*	*B*	*C*	Gender	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1	5.8	4.1	1.415	NA	1.000	0.994	0.990	NA	NA	NA	NA	NA	NA
2	5.4	4.1	1.317	1	0.687	0.801	0.693	0.687	0.801	0.693	0.313	1.249	1.442
3	5.2	4.1	1.268	0	0.480	0.392	0.253	0.520	0.608	0.747	-0.480	-1.646	-1.338
4	5.6	4.2	1.333	NA	0.756	0.881	0.665	NA	NA	NA	NA	NA	NA
5	5.3	4.1	1.293	NA	0.584	0.617	0.467	NA	NA	NA	NA	NA	NA
6	5.5	4.1	1.341	1	0.790	0.909	0.854	0.790	0.909	0.854	0.210	1.100	1.171
7	5.1	3.9	1.308	NA	0.647	0.739	0.884	NA	NA	NA	NA	NA	NA
8	5.1	3.9	1.308	0	0.647	0.739	0.884	0.353	0.261	0.116	-0.647	-3.824	-8.590
9	5.1	3.8	1.342	NA	0.793	0.911	0.983	NA	NA	NA	NA	NA	NA
10	5.4	4.1	1.317	NA	0.687	0.801	0.693	NA	NA	NA	NA	NA	NA
11	5.3	4.2	1.262	NA	0.453	0.337	0.103	NA	NA	NA	NA	NA	NA
12	5.1	4.1	1.244	0	0.377	0.206	0.116	0.623	0.794	0.884	-0.377	-1.259	-1.131
13	5.0	4.1	1.220	NA	0.274	0.094	0.048	NA	NA	NA	NA	NA	NA
14	5.2	3.9	1.333	NA	0.756	0.881	0.952	NA	NA	NA	NA	NA	NA
15	4.9	3.9	1.256	1	0.430	0.293	0.532	0.430	0.293	0.532	0.570	3.416	1.881
16	5.1	4.0	1.275	NA	0.509	0.454	0.499	NA	NA	NA	NA	NA	NA
17	5.1	4.0	1.275	1	0.509	0.454	0.499	0.509	0.454	0.499	0.491	2.204	2.004
18	5.3	4.0	1.325	1	0.720	0.844	0.870	0.720	0.844	0.870	0.280	1.185	1.150
19	5.4	4.0	1.350	NA	0.826	0.932	0.945	NA	NA	NA	NA	NA	NA
20	5.2	4.0	1.300	NA	0.615	0.679	0.720	NA	NA	NA	NA	NA	NA
21	5.4	3.9	1.385	NA	0.973	0.981	0.992	NA	NA	NA	NA	NA	NA
22	4.8	3.9	1.231	NA	0.322	0.137	0.305	NA	NA	NA	NA	NA	NA
23	5.3	3.9	1.359	1	0.864	0.951	0.981	0.864	0.951	0.981	0.136	1.052	1.020
24	5.5	4.0	1.375	NA	0.932	0.972	0.978	NA	NA	NA	NA	NA	NA
25	5.3	3.9	1.359	NA	0.864	0.951	0.981	NA	NA	NA	NA	NA	NA
26	5.2	4.0	1.300	NA	0.615	0.679	0.720	NA	NA	NA	NA	NA	NA
27	5.2	3.8	1.368	NA	0.904	0.965	0.993	NA	NA	NA	NA	NA	NA
28	5.3	4.0	1.325	1	0.720	0.844	0.870	0.720	0.844	0.870	0.280	1.185	1.150
29	5.3	3.9	1.359	NA	0.864	0.951	0.981	NA	NA	NA	NA	NA	NA
30	5.2	3.8	1.368	1	0.904	0.965	0.993	0.904	0.965	0.993	0.096	1.036	1.007
31	5.1	3.9	1.308	NA	0.647	0.739	0.884	NA	NA	NA	NA	NA	NA
32	5.4	4.2	1.286	NA	0.554	0.554	0.229	NA	NA	NA	NA	NA	NA
33	5.3	4.0	1.325	NA	0.720	0.844	0.870	NA	NA	NA	NA	NA	NA
34	5.0	3.9	1.282	NA	0.539	0.520	0.746	NA	NA	NA	NA	NA	NA
35	5.2	4.1	1.268	NA	0.480	0.392	0.253	NA	NA	NA	NA	NA	NA
36	5.8	4.3	1.349	NA	0.821	0.929	0.636	NA	NA	NA	NA	NA	NA
37	5.4	4.0	1.350	NA	0.826	0.932	0.945	NA	NA	NA	NA	NA	NA
38	5.2	3.9	1.333	NA	0.756	0.881	0.952	NA	NA	NA	NA	NA	NA
39	5.3	4.0	1.325	NA	0.720	0.844	0.870	NA	NA	NA	NA	NA	NA
40	5.2	3.8	1.368	NA	0.904	0.965	0.993	NA	NA	NA	NA	NA	NA
41	6.1	4.2	1.452	NA	1.000	0.998	0.996	NA	NA	NA	NA	NA	NA
42	5.6	4.0	1.400	1	1.000	0.989	0.991	1.000	0.989	0.991	-0.038	1.011	1.009
43	5.3	4.2	1.262	NA	0.453	0.337	0.103	NA	NA	NA	NA	NA	NA
44	5.4	3.8	1.421	NA	1.000	0.995	0.999	NA	NA	NA	NA	NA	NA
45	5.9	4.2	1.405	NA	1.000	0.991	0.972	NA	NA	NA	NA	NA	NA
46	5.3	4.1	1.293	0	0.584	0.617	0.467	0.416	0.383	0.533	-0.584	-2.610	-1.875
47	5.1	4.0	1.275	NA	0.509	0.454	0.499	NA	NA	NA	NA	NA	NA
48	5.6	4.3	1.302	NA	0.624	0.698	0.207	NA	NA	NA	NA	NA	NA
49	5.3	4.1	1.293	NA	0.584	0.617	0.467	NA	NA	NA	NA	NA	NA
50	5.8	4.0	1.450	1	1.000	0.998	0.999	1.000	0.998	0.999	-0.249	1.002	1.001
51	5.8	4.1	1.415	NA	1.000	0.994	0.990	NA	NA	NA	NA	NA	NA
52	5.2	4.3	1.209	NA	0.231	0.066	0.006	NA	NA	NA	NA	NA	NA
53	6.0	4.1	1.463	NA	1.000	0.999	0.999	NA	NA	NA	NA	NA	NA
54	5.4	4.1	1.317	NA	0.687	0.801	0.693	NA	NA	NA	NA	NA	NA
55	5.2	4.2	1.238	NA	0.353	0.173	0.043	NA	NA	NA	NA	NA	NA
56	5.2	4.0	1.300	NA	0.615	0.679	0.720	NA	NA	NA	NA	NA	NA
57	5.4	4.1	1.317	NA	0.687	0.801	0.693	NA	NA	NA	NA	NA	NA
58	5.3	4.2	1.262	NA	0.453	0.337	0.103	NA	NA	NA	NA	NA	NA
59	5.0	3.8	1.316	NA	0.681	0.793	0.957	NA	NA	NA	NA	NA	NA
60	5.2	4.1	1.268	NA	0.480	0.392	0.253	NA	NA	NA	NA	NA	NA
Mean	5.327	4.028	1.323		0.690	0.711	0.683	0.681	0.721	0.754	0	0.364	-0.007
SD	0.261	0.134	0.058		0.207	0.274	0.325	0.215	0.268	0.256	0.401	1.965	2.747

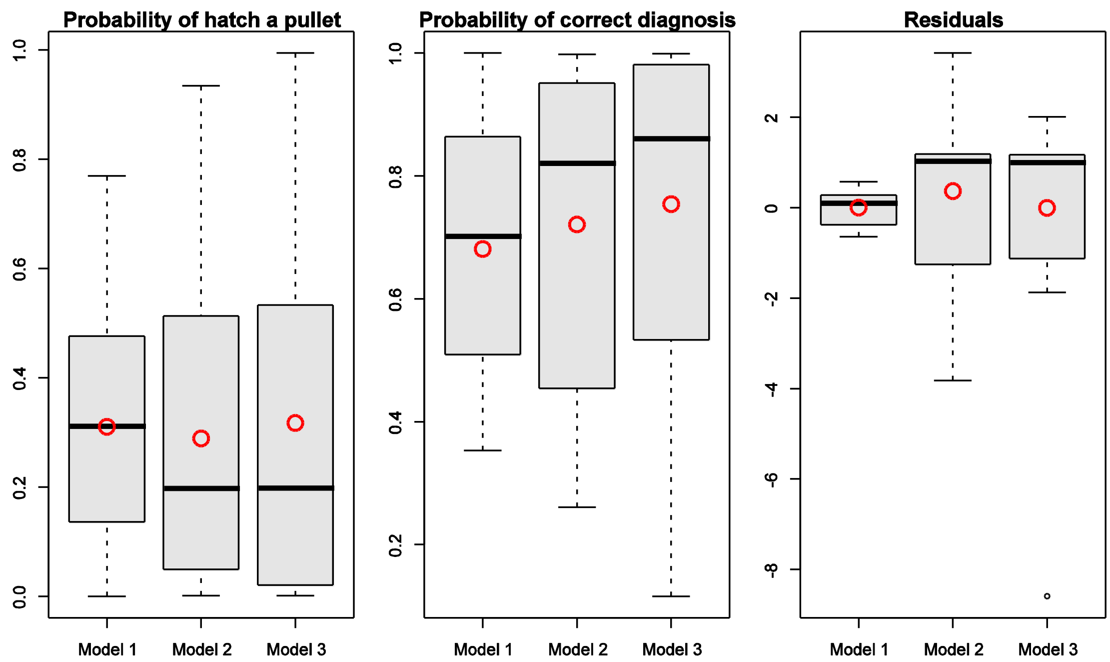


Fig. 7: Box plots of the predicted probability of hatch pullet from eggs, the probability of correct diagnosis and residuals under three fitted models in Eqs. (3)-(5).

Table 3: Regression analysis summary for gender predicting based on simple linear Model 1 presented in Eq

Summary of Fit						
R-squared	0.269		Adjusted R-squared	0.208		
Mean of Response	0.714		AIC	19.09		
Root Mean Squared Error	0.386		BIC	21.01		
Analysis of Variance						
Source	**DF**	**Sum of Squares**	**Mean Squares**	**F value**	**Prob > |F|**	
Model	1	0.769	0.769	4.418	0.057	
Error	12	2.089	0.174			
Total	13	2.857				
Parameter Estimates						
Term	**Estimate**	**Std Error**	**t value**	**Prob > |t|**	**95% CI**	**99% CI**
Intercept	-4.887	2.667	-1.832	0.092	(-10.7, 0.9)	(-13.0, 3.2)
Sharpness (*C*)	4.232	2.013	2.102	0.057	(-0.2, 8.6)	(-1.9, 10.4)

Table 4: Regression analysis summary for gender predicting based on logistic regression Model 2 presented in Eq. (4).

Summary of Fit						
Null deviance	16.752 on 13 DF		Residual deviance	11.610 on 12 DF		
Mean of Response	0.714		AIC	15.61		
Root Mean S. E.	0.380004		BIC	16.89		
Parameter Estimates						
Term	**Estimate**	**S. E.**	**z value**	**Prob > |z|**	**95% CI**	**99% CI**
Intercept	-47.93	28.50	-1.682	0.093	(-121.6, -4.3)	(-153.8, 5.9)
Sharpness (*C*)	37.44	22.03	1.699	0.089	(3.9, 94.6)	(-3.8, 119.7)

Table 5: Regression analysis summary for gender predicting based on logistic regression Model 3 presented in Eq. (3)

Summary of fit

Null deviance	16.752 on 13 DF		Residual deviance	10.720 on 11 DF		
Mean of Response	0.714		AIC	16.72		
Root Mean Squared Error	0.348		BIC	18.64		
Parameter Estimates						
Term	**Estimate**	**Std. Error**	**z value**	**Prob. > |z|**	**95% CI**	**99% CI**
Intercept	32.756	38.281	0.856	0.392	(-37.5 124.3)	(-60.7, 163.1)
Length (*A*)	9.501	6.118	1.553	0.120	(0.3, 25.8)	(-1.8, 33.4)
Diameter (*B*)	-20.304	13.267	-1.530	0.126	(-56.9, -0.7)	(-74.1, 3.9)

all other models. Hence, considering Tables 3, 4 and 5, one can claim that the proposed logistic regression with explanatory variable *C*, i.e. Mode 2, is the best-fitted model among three presented models in this study.

The null deviance shows how well the response is predicted by the model with nothing but an intercept. The residual deviance shows how well the response is predicted by the model when the predictors are included. The difference between null deviance and residual deviance show that the fitted logistic regressions in 4 and 5 are suitable, since greater the difference imply to a better fitted model. In addition, we can test whether the logistic regression model provides an adequate fit for the data (null hypothesis) by a chi-square test using p-value= P (χ^2_{DF} > deviance) in which χ^2_{DF} is the chi-square random variable with *DF* degrees of freedom. Therefore, considering Tables 4 and 5, one can compute p-values to test the goodness-of-fit for logistic models 2 and 3 as 0.477 and 0.467, respectively, which imply to high significance level for both proposed logistic regression models.

In follow, the presented logistic Model 2 in Eq. (4) is briefly interpreted as the best-fitted model. Based on Model 2, each one change in the sharpness index C will increase the natural logarithm of the odds ratio of getting admit by 37.44, and the p-value indicates that it is significant in determining the admit.

The proposed Model 3 in Eq. (5) can be interpret as follow. If *B* considered a fix value, then each unit increase in the major-diameter (*A*) increases the log odds of getting admit by 9.5 and p-value indicates that it is somewhat significant in determining the admit. Similarly, if A fixed, then each unit increase in the minor-diameter (*B*) decreases the log odds of getting admit by 20.30 and p-value indicates that it is somewhat significant in determining the admit.

Conclusion

Several regression models are propose in this paper to predict the probability of hatch a male/female chicken from egg before it is even placed in an incubator. The prediction of egg's gender is based on the minor and major diameters of eggs which are simply measureable by the caliper. A binomial logistic regression model is fitted to collected data to predict the the probability of cockerel by considering two diameters of egg as the explanatory variables. Moreover, we introduce the sharpness/thin index for egg, as a function of its physical shape. Then, a simple linear regression model and another binomial logistic regression model are fitted on the basis of calculated measures for the sharpness index. The sharpness index may be define by another formula in future research works, but it must be noted that the logistic regression analysis under any linear transformation of the presented index as a new sharpness index, e.g. $\frac{A-B}{B}$, is not meaningful and will have a same result.

After analysis and interpret three fitted regression models, the following goodness-of-fit criteria are estimated in this research to determine the best regression model: (1) Akaike information criterion, (2) Bayesian information criterion, (3) the difference between null deviance and residual deviance and (4) chi-square test. All estimated criteria implied that Model 2, i.e., the proposed logistic regression with explanatory variable C, is the best fitted model among three fitted models in this paper. Advantages and merits of the proposed method in this research are: (I) the proposed method is simple and inexpensive because only laboratory equipment needed is a caliper and so it can be simply implemented in any small poultry farm, (II) high speed of prediction process, (III) eggs remain perfectly healthy and safe after the gender prediction, so the proposed method could have both financial and economic benefits for the user, and (IV) regarding to the previous item, the proposed method could lead to greater respect for animal rights in the egg industry to avoid to killing the unwanted male chicks.

It should be emphasized that the proposed regression models in this research are calculated based on a special kind of egg in Kerman, Iran. Therefore, to predict the gender of different breeds of eggs, the presented models must be up-to-date based on the new measured local data. To increase the accuracy of the proposed logistic regression model, one can import more related physical explanatory variables (e.g. color and the degree of the eggshell darkness) to the regression model in future works. Moreover, one can consider chemical explanatory variables in the logistic regression model to increase the accuracy of model in the prediction. As presented earlier, the most of the chicks unfortunately died at the first of this study and finally fourteen chicks remained. Hence, it must be mention that the small sample size can be reduces the validity of the proposed models in this paper and large sample sizes is proposed for future works.

Acknowledgements

The authors would like to express their deepest appreciation to Mahan Chicken Production Complex of Kerman, one of the largest producer one-day-old broiler chickens in Iran. The research would not have been possible without the support of Dr. Hasan Ghaemi, health manager of the company.

Conflict of interest

The authors declare that they have no competing interests.

Ethical approval

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

.

References

1. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 1974, 19:716-723.
2. Berkson J. Application of the logistic function to bio-assay. JASA, 1944, 39(227), 357-365.
3. Galli R, Preusse G, Uckermann O, Bartels T, Krautwald-Junghanns ME, Koch E, Steiner G. In Ovo Sexing of Domestic Chicken Eggs by Raman Spectroscopy. Anal Chem, 2016, 88(17), 8657-8663.
4. Greenhouse J B, Bromberg J A, Fromm D. An introduction to logistic regression with an application to the analysis of language recovery following a stroke. J Commun Disord, 1995, 28(3), 229-246.
5. Galli R, Preusse G, Uckermann O, Bartels T, Krautwald-Junghanns ME, Koch E, Steiner G. In ovo sexing of chicken eggs by fluorescence spectroscopy. Anal and Bioanal Chem, 2016,409(5), 1185-1194.
6. Galli R, Preusse G, Schnabel C, Bartels T, Cramer K, Krautwald-Junghanns ME, Koch E. Steiner G. Sexing of chicken eggs by fluorescence and Raman spectroscopy through the shell membrane. PLOS ONE, 2018, 13(2), e0192554.
7. Nottingham Q J, Birch J B. A note on the small sample behavior of logistic regression in a bioassay setting. J Biopharm Stat, 1998, 8(4), 565-576.
8. Myers W R, Myers R H, Carter W H, White K L. Two-stage designs for the logistic regression model in single-agent bioassays. J Biopharm Stat, 1996, 6(3), 283-301.

9. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL https://www.R-project.org/.
10. Stone M. Comments on model selection criteria of Akaike and Schwarz. (Series B). J R Stat, 1979, 41, 276-278.

How to cite this article:
Parchami, A., Aghaei Afshar, R., Fadaie, E., Ghaemi, H., Sakhaee, E., Prediction of chicken gender before putting eggs in incubator using logistic regression model. Veterinary and Comparative Biomedical Research, 2024, 1(2): 8 – 18. http://doi.org/10.22103/Vcbr.2024.23250.1013.