International Journal of Mathematical Research
2013 Vol. 2, No. 5, pp. 37-44
ISSN(e): 2306-2223
ISSN(p): 2311-7427
DOI: 10.18488/journal.24/2013.2.5/24.5.37.44

DISCRIMINATION AND CLASSIFICATION OF POULTRY FEEDS DATA

Olosunde A.A.[1] --- Soyinka A.T[2]

[1]*Department of Mathematics, ObafemiAwolowo University, Ile-Ife*

[2]*Department of Statistics Federal University of Agriculture, Abeokuta*

ABSTRACT

This study is aimed at employing discriminant analysis method and classification for the purpose of achieving the assessment of a discriminant function through which we can discover the reasons of the actual difference between two groups of eggs of which the chicken were fed with different combination of feeds. Fisher's Linear Discriminant Function (LDA) was used as a tool for the Statistical analysis. It was estimated on the basis of a sample of 96 chickens, which were classified into two groups of 48 chickens each. One group was fed with in-organic copper salt combination while the second group with organic copper salt combination. Some important attributes are measured from the eggs produced from these two groups; such as egg's size(g) and cholesterol level(mg).The results obtained assert the efficiency of the discriminant function which we obtained and the possibility of its use for the purpose of discriminating and classifying the eggs of unknown feeds into corresponding group in future.

Keywords: Discrimination, Classification, Multivariate normal, Poultry feeds data, Organic and inorganic.

Mathematics Subject Classification: 60E05, 62H10, 6207.

Received: 26 April 2013/ **Revised:** 16 August 2013/ **Accepted:** 19 August 2013/ **Published:** 22 August 2013

1. INTRODUCTION

The discriminant analysis and classification are considered as multivariate analysis concerning the discrimination of different groups of observations and classifying the new observations into predefined groups. We resort to such method, when it is difficult to understand the causal relation of the difference among the different groups in a sufficient degree [1]. The study aims at using the classical method of discrimination and classification in poultry feeds data based on the cholesterol level and the size of the eggs. The first group of chickens was fed with feeds prepared from in-organic copper-salt and the second group the organic copper-salt. The data on the measured sizes of the eggs as well as the cholesterol level were taken, and then use the discriminant analysis method and classification for achieving an estimation of discriminant function which can be used for separating or discriminating the two groups of eggs produced from the chicken fed with feeds containing different contents. Though the chicken were separated during the process raring them, but they were subjected to the same conditions throughout the process. Their data were taken separately, but the study also aims to identify the arrangement scale of the relative importance of these attributes and to reach a rule which can be used for future discrimination and classification of eggs in one group of both specified groups. It is worth noting that the 96 chicken were randomly selected and also randomly divided into two equal groups of 48 chickens each for the purpose of this study.

2. CLASSES AND FEATURES

Anderson [2] and Hand [3], considers the case of binary classification where (K=2), and we wish to discriminate between two classes Π_1 and Π_2 such as eggs produced from chicken fed with organic copper salt and another group of eggs where chicken were fed with the inorganic copper salt.

2.1 Bayes's Rule Classifier

Let

$$P(\Pi_1) = \pi_i, \qquad i = 1,\ 2, \tag{1}$$

be the prior probabilities that a randomly selected observations $X = x$ belongs to either Π_1 or Π_2. Suppose that the conditional multivariate normal probability density of X for the i th class is

$$P(X = x \mid X \in \Pi_i) = f_i(x)\ i = 1,\ 2. \tag{2}$$

Bayes's theorem yields the posterior probability,

$$P(\Pi_i \mid x) = P(X = x \mid X \in \Pi_i) = \frac{f_i(x)\pi_i}{f_1(x)\pi_1 + f_2(x)\pi_2} \quad , \tag{3}$$

that the observed x belongs to Π_i $i = 1,\ 2$. The technique is to assign x to that class with the higher posterior probability. That is, we assign x to Π_1 if

$$\frac{p(\Pi_1 \mid x)}{p(\Pi_2 \mid x)} \succ 1 \tag{4}$$

and we assign x to Π_2 if otherwise. Now substituting (3) and (4) we have the Bayes's rule classifier assigns x to Π_1 if

$$\frac{f_1(x)}{f_2(x)} \succ \frac{\pi_2}{\pi_1} \qquad (5)$$

and to Π_2 if otherwise. Now if $x \sim N(\mu, \Sigma)$ and make the homogeneity assumption that $\Sigma_1 = \Sigma_2 = \Sigma_{xx}$ Krzanowski [4]. Then we have

$$L(x) = \ln\left[\frac{f_1(x)\pi_1}{f_2(x)\pi_2}\right] = b_0 + b^T X \qquad (6)$$

as a linear function of x, where

$$b = \Sigma_{xx}^{-1}(\mu_1 - \mu_2) \qquad (7)$$

$$b_0 = \frac{-1}{2}\left[\mu_1^T \Sigma_{xx}^{-1}\mu_1 - \mu_2^T \Sigma_{xx}^{-1}\mu_2\right] + \ln\left(\frac{\pi_2}{\pi_1}\right) \qquad (8)$$

Thus if $L(x) \succ 0,\ x \in \pi_1$, otherwise $x \in \pi_2$.

3. APPLICATION TO POULTRY FEEDS DATA

3.1. Motivation

Cholesterol is a waxy substance that comes from two main sources: The liver and food intake,it is well known that that too much cholesterol can clog the arteries, lower blood flow to other tissue in the body, thereby precipitate heart diseases. Eggs have often been frown at because of high cholesterol contents, therefore causing decreased in consumption. But Majumder and Wu [5] note that eggs are inexpensive source of high-quality protein and other nutrients. From these arguments, it simply implies that eggs have some nutrients necessary for daily sustenance, instead of total boycott; something must be done to reduce the risk associated with eggs consumption. It has also widely reported that eggs contain lecithin and phosphor-lipids, integral to the construction of brain cell membrane. In terms of feeding intellect, their value lies mainly in the quality of their proteins. Long used as points of reference when analyzing the quality of other dietary proteins by the Food and Agriculture Organization (FOA), eggs are actually rich in amino acids, essential in the production of the principal neurotransmitters. In order to alleviate the problem associated with consumption, an organic copper-salt combination was introduced to replace inorganic combinations being currently used in preparation of poultry feeds to reduce the risk of cholesterol level in eggs. 96 set of chicken were randomly selected and also randomly divided into

Table-1. Summary of Data in Appendix

Variable	Min	1st Quantile	Median	Mean	3rd Quantile	Maximum
weight (inorganic)	52.43	55.37	58.30	58.35	61.23	65.67
cholesterol (inorganic)	154.80	177.10	196.10	195.70	215.20	231.10
weight (organic)	56.08	57.59	59.12	59.10	60.60	62.69
cholesterol (organic)	60.73	101.75	132.23	131.46	162.69	193.16

two groups comprising 48 each, each population were subjected to the same conditions and treatments, only composition of the feeds are different while, the first group were fed with inorganic copper-salt, the second group were fed with the inorganic copper-salt. Some major important characteristics were measured in the eggs produced from these two groups namely: the weight(g) of the eggs and the cholesterol level(mg/egg).
The data are presented in the Appendix A.

3.2. Analysis of Data

Based on the above facts, it is necessary to have some features which will classified the eggs with increase in weight but lower cholesterol from the eggs with lesser weight but high in cholesterol content. Normality assumptions are common in data analysis, In order to make use of nice properties in the discrimination and classification we need to carry out some important test to ensure proper compliance to the rule guiding the usage. The summary of data in Appendix A are presented in table 1.

4. HYPOTHESIS TESTING

- We need to carry out some normality test on the data to ensure good approximation by the normal distribution.A reasonable approximation by normal distribution is observed in the normal q-q plots in the figure 1.0 below. This is also substantiated by Shapiro-Wilk Normality test, which gives W= 0.945 p-value of 0.09444. Therefore, we do not reject the hypothesis of normality assumption.

- Test $H_0 : \Sigma_1 = \Sigma_2 = \Sigma$

  The test statistic= -2ln Λ = 13.054878. Using -2ln $\Lambda \sim X_{12}^{\;2}$, The P - value = 0.3650646. Hence we do not reject H_0 and conclude that the variance-covariance matrices are the same for the two groups.

- Next we test the conditional hypothesis of equality of means, that is $H_0 : \mu_1 = \mu_2 = \mu$, given that $\Sigma_1 = \Sigma_2 = \Sigma$

Table-2. Parameters Estimation for eggs weight in Appendix A

Density	Variables	Mean	Standard deviation	Log-likelihood
normal	Weight (inorganic)	58.35	3.559	-129.055
	Weight (organic)	59.10	1.822	-96.910
	Cholesterol (inorganic)	195.728	21.907	-216.282
	Cholesterol (organic)	131.457	37.232	-241.739

The test statistic = -2ln Λ = 91.70311. The P - value = P [X_6^2 > 91.70311] < 0.0001. Hence we reject H_0 and conclude that they are different.

4.1. Kolmogorov-Smirnov-Lillefort Test for difference in μ

We used Kolmogorov-Smirnov two-sample test for large samples to decide whether there is significant different in the weights and cholesterol levels. When there is a significant difference, then feeds are changed to organic copper-salt from inorganic type. Note that Kolmogorov-Smirnov two-sample test compares the two sample cumulative frequency distributions and determines whether the observed D indicates that which they have been drawn from two populations, one of which is stochastically larger than the other. Let $F_{n1}(X)$ be the cumulative step function of the sample observations for the inorganic copper-salt type and let $F_{n2}(Y)$ be the cumulative step function of the sample observations for organic copper-salt type. We test the null hypothesis that the two samples have been drawn from the same population against the alternative hypothesis that the values of the population from which one the samples was drawn are stochastically larger than the values of the population from which the other sample was drawn. That is the hypothesis of interest is $H_0 : F_{n1}(X) = F_{n2}(Y)$ against the alternative $H_A : F_{n1}(X) \prec F_{n2}(Y)$ for both the egg weights and the cholesterol levels. Define

$$D = \max imum[F_{n1}(X) - F_{n2}(Y)].$$

The sampling distribution of D is said to be generalized exponential power distribution. It has been shown by Goodman that

$X_2{}^2 = 4D^2 \frac{n}{2}$ (For number of observations are the same)

1. For the egg weights D = 0.3125 with corresponding P [$X_2{}^2$ > 9.21] < 0.01. Hence, we reject H_0 and conclude that the weight of the eggs fed with inorganic copper salt is less than the organic type.

Table-3. F – Table for the Discriminant Analysis

F	MS	SS	d.f	variables
7.2791	2.2138	SS_B = 11.0692	4	Between X's
	0.3041	SS_W = 6.6909	92	Within X's
	nil	17.7601	96	Total

2. For the cholesterol level which is the most important, we have: D = 0.6875 with P [$X_2{}^2$ > 45.375] <0.0001. Hence we reject the null hypothesis and conclude that, using the organic copper-salt type the cholesterol level significantly reduced.

The Fisher's (sample) linear discriminant function given by

$$Y = \hat{\alpha}' X = (\overline{X_1} - \overline{X_2})^{-1} S_{pooled} X \qquad (9)$$

was estimated and our estimated equation is y = 35.7632 X_1 + 13.3455 X_2 . The analysis of variance table for the discriminate analysis can be performed to test the following hypothesis:

$H_0 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ Vs. at least one of the four $\alpha' s$ is non-zero and it as follows:

where $SS_B = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)} \left[D^2\right] = 11.0692$ and $SS_D = D^2 = 6.6909$.

Now since F > $F_{4,92}(0.01)$, we reject H_0 at the 10 percent level of significance. This means the estimated discriminant function has a big advantage in discriminating or classifying the eggs under study. The mid-point (Z) between the groups is given by $\hat{Z}$ = 221.1849. Therefore, the classification rule will be as follows:

allocate X_0 to π_1 if $Y_0 = \hat{\alpha}' X \geq$ 221.1849

allocate X_0 to π_2 if $Y_0 = \hat{\alpha}' X \prec$ 221.1849

Now for the purpose of evaluating the performance of the adopted classification method, we calculate Apparent Error Rate (APER)=4/96=0.0417 and Actual Error Rate(AER)=0.00412 from the confusion matrix. The value of the error rate in both cases does not exceed 10 percent. This value means that the optimal classification rule in use resulted in placing less than 10 percent of the eggs in a population which they do not belong to.

5 Concluding Remarks

We have proposed a linear discriminant function to classify eggs based on the size and cholesterol level. The function gives a good prediction this is shown from the APER and AER calculated.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no conflict of interests.

Contributors/Acknowledgement: All authors contributed equally to the conception and design of the study.

REFERENCE

[1] M. A. Ali, A. Sadoon, and V. K. Iryna, "Discrimination and classification levels of students as per their major specialization," *Journal of Applied Sci.*, vol. 6, pp. 1845-1853, 2006.

[2] T. W. Anderson, *An introduction to multivariate statistical methods*. New York: John Wiley, 1984.

[3] D. J. Hand, *Discrimination and classification*. New York: John Wiley, 1981.

[4] W. J. Krzanowski, "The performance of fisher's line discriminant function under non-optimal condition," *Technometrics*, vol. 19, pp. 191-200, 1977.

[5] K. Majumder and J. Wu, "Angiotensin i converting enzyme inhibitory peptides from simulated in vitro gastrointestinal digestion of cooked eggs," *J. Agric. Food Chem.*, vol. 52, pp. 471-477, 2009.

Figure-1. Appendix A: Observed data from Inorganic and Organic Copper salt

I no rgani c co p pe r-s al t		O rg an ic co p pe r-s al t	
weigh t	ch o les tero l	w e igh t	c h olesterol
52 .67	164 .23	56 . 08	56 .08
53 .17	167 .42	56 . 34	56 .34
53 .67	170. 6	56 . 61	56 .61
54 .17	173 .78	56 . 87	56 .87
54 .67	176 .96	57 . 13 57 .13	
55 .17	180 .14	57 . 39	57 .39
55 .67	183 .32	57 . 65 57 .65	
56 .17	186 .51	57 . 92	57 .92
56 .67	189 .69	58 . 18	58 .18
57 .17	192 .87	58 . 44	58 .44
57 .67	196 .05	58 .7	58.7
58 .17	199 .24	58 . 96	58 .96

58 .67	202 .42		59 . 23 59 .23	
59 .17	205.6		59 . 45	59 .45
59 .67	208 .78		59 . 75	59 .75
60 .17	211 .96		60 . 01 60 .01	
60 .67	215 .14		60 . 27	60 .27
61 .17	218 .33	60. 54	60 .54	
61 .67	221 .52	60 .8	60. 8	
62 .17	224 .69	61 . 06	61 .06	
62 .67	224 .85	61 . 32	61 .32	
63 .17	227 .88	61 . 58	61 .58	
63 .43	228 .03		61 . 85	61 .85
65 .67	231 .06		62 . 34	62 .34
65 .15	228 .0	62 . 11	62 .11	
63 .43	224 .8	61 . 85	61 .85	
62 .93	221 .6	61 . 58	61 .58	
62 .43	218 .46	61 . 32	61 .32	
61 .93	215 .28	61 . 06	61 .06	
61 .43	212.1		60 .8	60. 8
60 .93	208 .92		60 . 54	60 .54
60 .43	205 .74	60 . 27	60 .27	
59 .93	202 .56		60 . 01	60 .01
59 .43	199 .37	59 . 75	59 .75	
58 .93	196 .19		59 . 49	59 .49
58 .43	193 .01		59 . 23	59 .23
57 .93	189 .83		59	59
57 .43	186 .65		58 .7	58. 7
56 .93	183 .46		58 . 44 58 .44	
56 .43	180 .28		58 . 1858 .18	
55 .93	177. 1		57 . 92	57 .92
55 .43	173 .72		57 . 6557 .65	
54 .93	170 .74		57 . 3957 .39	
54 .43	167 .55		57 . 1357 .13	
53 .93	164 .37		56 . 8756 .87	