Canonical discriminant analysis of early maturity traits of parent stock layer strains in the tropics

O. M. A. Jesuyon[1†] and C. Isidahomen[2]

[1]*Animal Breeding and Genetics Unit, Department of Animal Production and Health; Federal University Oye - Ekiti, P. M. B. 373 Oye-Ekiti, Ekiti State, Nigeria;* [2]*Ambrose Ali University, Ekpoma, Edo State, Nigeria*

SUMMARY

Stepwise discriminant analysis was conducted to classify pullet strains into Early maturing or Late maturing. Reproductive, Age and body traits examined were Age at first egg (AFE), Hen weight at first egg (HWFE), Hen weight at 20 weeks (H20wks), Egg weight at full maturity (EWT), Hen house production at full maturity (HHP), Fertility of egg set at full maturity (FES), Hatchability of egg set at full maturity (HES) and Pullet day-old chicks hatched at full maturity (PDC). Test of equality of Class means observed significant ($p<0.001$) differences for AFE on the Discriminant Variables (DV) table. Box's M indicated that the assumption of equality of covariance matrices was violated ($p<0.002$), but the log determinants were quite similar (11.554 to 13.544). Two body traits namely AFE and HWFE were entered by the stepwise procedure into the discriminate Function (DF). The DF revealed a significant association between Maturity Classes and all discriminators, accounting for 67.3% of between group variability, although closer analysis of the structure matrix revealed three significant predictors, namely AFE (0.771), HES (-0.314) and PDC (–0.304), while others were poor predictors (0.292 to -0.102). The stepwise procedure however identified the best set of discriminating variables in the DF for maximizing classification as AFE and HWFE. The cross validated classification showed that overall 93.3% of cases were correctly classified. The function thus obtained could be used successfully by farmers to classify any batch of Bovan Nera, Isa Brown or other Parent stock strains in tropical environment for sexual maturity status. The best discriminating pair of traits for ESM in pullets was AFE and HWFE. Results obtained significant differences between the mean of early and late maturing pullets to allow for their classification.

Keywords**:** age at first egg, chicken, classification, early and late maturing, stepwise method

---

[†] Corresponding author e-mail: dr.oluwatosinjesuyon14@gmail.com

INTRODUCTION

The onset of sexual maturity in chicken is a trait of economic importance to farmers and evolutionary significance to Researchers. The time period to sexual maturity in modern domestic layer hens have been reduced by about 20% compared with their wild progenitor chickens (Wright *et al.*; 2012). Sexual maturity status in chicken can be categorized into two namely early sexual maturity (ESM) which takes place at the point of First egg lay (FE) and Full sexual maturity (FSM) which takes place at the peak of egg production (HHP). Both stages can be further sub-categorized or classed into Early maturing (EM) and Late maturing (LM) based on whether the First egg or the Peak is reached before or after the genetically-associated Average Age for the Strain. In order to identify the point of early sexual maturity in a flock of birds, the mean age and body weight of the flock at first egg-lay are usually taken as points of reference (Dunnington and Siegel, 1984). The usual interest of a tropical farmer is the age at which his strain will attain sexual maturity. Farmers today want to know if a strain is early maturing or late maturing. Since there are so many strains competing for patronage, a scientific means of identifying their maturity classes would be beneficial to farmers, managers, researchers and breeders. Through modelling of sexual maturity traits in PS chicken we could classify strains into Early or Late maturing; the discriminant analysis (D A) procedure could be utilized for this purpose. Canonical D A was expected to create an equation which will minimize the possibility of misclassifying strains into their respective sexual maturity classes, groups or categories. Published research findings where canonical discriminant analysis were applied to performance traits of chicken were limited, however Rosario et al., (2008) in his work with broilers earlier suggested analysing performance data sets using canonical D A. The objective of the study was to identify the best combination of traits that could be utilized to classify early sexual maturity in Parent stock (PS) Pullets; construct an equation that will discriminate between strains into Early and Late maturing classes and classify other strains using the discriminant function. Hypothesis tested was that there were no differences in the mean of traits between early maturing and late maturing pullets and therefore no classification differential between layer strains for sexual maturity in the humid tropics.

$H_o : \mu_{\text{early maturing}} = \mu_{\text{late maturing}}$

$H_1 : \mu_{\text{early maturing}} \neq \mu_{\text{late maturing}}$

MATERIAL AND METHODS

Data covering ten years consisting of at least 40 batches each of two Parent stock layer hens namely Bovan Nera (BN) and Isa Brown (IB) were obtained from a Farm in Ibadan, Nigeria. Information retrieved from the record books for the study included Age at first egg (AFE), Hen weight at first egg (HWFE), Hen weight at 20 weeks (H20wks), Egg weight at full maturity (EWT), Hen house production at full maturity (HHP), Fertility of egg set at full maturity (FES), Hatchability of egg set at full maturity (HES), and Pullet day-old chicks hatched at full maturity (PDC). These traits were subjected to the Discriminant Analysis (DA) to see which ones will contribute to discrimination between the 2 classes among chickens and across strains. Cases or batches that matured before the mean AFE for the strain were regarded as Early maturing (EM) while those that mature after the mean AFE for the strain were regarded as Late maturing (LM). Thus the classification variable chosen for the DF analysis was AFE. Subsequently, data were grouped into non-overlapping natural maturity groups for the classification – Early and Late maturing – based on the genetically-associated average AFE for each strain. D A involves two processes of (1) testing significance of a set of discriminant functions as in the case of more than 2 classes, and (2) classification. The Stepwise discriminant procedure of SPSS ® (Version 17) was utilized to determine the best set of traits that will discriminate best between classes (independent predictors). The relative importance of predictors in discriminating between the two Early sexual maturity (ESM) classes was assessed using class descriptive statistics, Box's M test of equality of covariance matrices, Stepwise Statistics by Wilks' Lambda, Summary of canonical discriminant functions, and Classification statistics.

Discriminant model

The multivariate model to be used to predict which class a strain belongs to is of the form:

$$D = a + v_1 X_1 + v_2 X_2 + v_3 X_3 + \dots\dots + v_i X_i$$

Where

D = discriminate function or Canonical root. This is a latent variable which is created as a linear combination of discriminating (independent) variables.

a = constant

v = the unstandardized discriminant coefficients or weights for variables which maximizes the distances between the means of the criterion (dependent) variables.

X = discriminating variables

i = the number of classifying variables.

The technique involved finding a linear combination of independent variables that maximized the distance or difference between class memberships (strains) in the categorical dependent variable and come up with an equation that has strong discriminatory power between classes in the classification and dependent variable. The Stepwise method determines the best combination of predictor variables that should be included in the DF. After using experimental data to calculate the discriminant function and classifying cases, any new case or strain could then be classified.

Table 1: Maturity class means and standard deviations of parent stock layers in tropical Nigeria.

Traits	Early maturing	Late maturing	Both Strains
	Mean±SD	Mean±SD	Mean±SD
AFE (days)	110.923±8.098^{b}	127.556±7.650^{a}	117.727±11.394^{ab}
HW20wks (g)	1635.462±85.997	1568.444±91.819	1608.046±92.602
HWFE (g)	1330.346±139.221	1445.188±151.615	1377.327±152.226
EWT (g)	56.408±2.955	56.467±3.541	56.432±3.125
HHP (%)	86.436±5.042	86.861±5.302	86.610±5.028
FES (%)	88.234±4.998	82.807±10.969	86.014±8.220
HES (%)	76.599±10.192	70.630±11.699	74.157±10.978
PDC (%)	36.797±5.267	34.926±8.971	36.031±6.884

Note: Superscripts a, b, indicate significant differences between class means.

Results

Table 1 gave the class means and SD of body parameters estimated for use in the model. There were large differences in means of AFE (16.63 days), HW20wks (67.02 g) and HWFE (114.85 g) between maturity classes respectively. The test of equality of class means was significant ($p<0.0001$) in AFE, and thus indicated strong statistical evidence of significant differences in means between classes. The pooled within-classes matrices showed that there were cases of low inter-correlations (-0.030 to -0.488) between pairs of many independent variables (69% or 44 out of 64 pairs).

The Box M test result elucidated significant (Box M=10.864, F=3.207 and $p>0.022$) values indicating non-similarity and significant differences between classes. It tests the null hypothesis of equal population covariance matrices but the log determinants revealed that class covariance matrices were similar (11.554 - 13.544) between classes formed by the dependent function. The stepwise statistical results showed that two steps only were taken to include AFE and HWFE respectively in the highly significant ($p<0.0001$) discriminant function. The Wilks' lambda statistics dropped from 0.460 in the first to 0.447 in the second step.

Table 2: Eigenvalue, canonical correlation, Wilks' lambda and significance of discriminate function

Eigen and canonical correlation values					Test of functions			
Function	Eigenvalue	% of variance	Cumulative %	Canonical correlation	Wilks' Lambda	Chi-square	df	*P*
1	1.970	100.00	100.00	0.814	0.337	20.681	2	0.0001

Table 2 revealed the summary of the discriminant function statistics. It gave an eigenvalue higher than unity and a canonical correlation value of 1.970 and 0.814 respectively. The test of the function gave Wilks' lambda statistics of 0.337 and a highly significant ($p<0.0001$) function. Those variables with low (<0.3) correlation values were regarded non-significant and were dropped, but three variables had values above 0.3 namely AFE (0.771), HES (-0.314) and PDC (-0.304).

Table 3: Canonical discriminant function coefficients and maturity class centroids

Canonical discriminant function coefficients		Class centroids	
Trait	Function	Maturity class	Function
AFE	0.219	Early maturing	-1.113
HWFE	-0.008	Late maturing	1.608
Constant	-14.752		

Table 3 displayed the canonical discriminant function coefficients and the class centroids. The class centroids or class means of all predictor variables in the discriminant equation were given as -1.113 and 1.608. The equation created was:

Maturity Class = -14.752 + (0.219*AFE) + (-0.008*HWFE).

The results of ANOVA on the discriminant function scores which was another overall test of the DA model was highly significant ($p<0.0001$). The discriminant scores were plotted on histogram (Figure 1) and this also demonstrated the effectiveness of the discriminant function for classification of cases.

Discussion

The results of class means suggested that the variables could be useful and be good discriminators, because the differences or separators were large. Early maturing pullets demonstrated better performance in HW20wks, FES, HES and PDC; while the late maturing pullets gave better mean results in HWFE and marginal increases above the other class in EWT and HHP. The test of equality of class means was highly significant, and gave strong statistical evidence of important difference in AFE between classes. The low inter-correlations result of the pooled within-classes matrices also supported the use of the independent variables (VIs) for discriminant analysis.

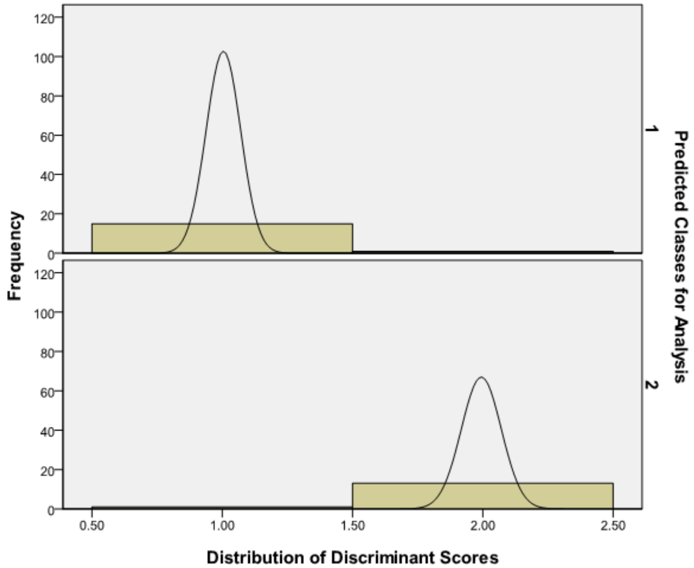

Figure 1: Histogram showing the distribution of discriminant scores for Early maturing and Late maturing classes of parent stock pullet strains in tropical Nigeria.

Box M tests for equality of covariance matrix and the hypothesis that covariance matrices do not differ between maturity classes formed by the dependent function. Researchers want this test not to be significant so that the null hypothesis that the groups do not differ can be retained. For this assumption to hold, the log determinants should be equal. When tested by Box's M, we looked for a non-significant M to show similarity and lack of significant differences and in this case the log determinants appear similar. However, with large samples, a significant Box M result is not regarded as too important (Garson, 2012; Sage, 2014). The stepwise procedure included only variables minimizing the overall Wilks' lambda statistics, and each step added to the predictive power of the function by decreasing the lambda value. The eigenvalue is also called the *characteristic root* of each discriminant function, reflects the importance of the dimensions which classify cases of the dependent variable. It reflects 100% of variance explained in the dependent variable, thus the relative discriminating power of the discriminant function. The canonical correlation which is the multiple correlations between predictors and the discriminant function was high. The Wilks lambda value is the proportion of total variability in data that was not explained by the function and tests the significance of the discriminant function as a whole. Therefore the proportion of variability explained by the function was (1- 0.337) i.e. 0.663. Wilks's lambda is a measure of the difference between groups of the centroid of means on the independent variables. The smaller the lambda, the greater the differences. In Stepwise Wilks' lambda, the more important a trait in classifying the grouping variable, the higher its stepwise Wilks' lambda. Thus it tests the significance of each discriminant function and specifically the significance of the eigenvalue for the function. Lambda of 0.447 means that the class means differ greatly and the more these mean differentiated ESM classes. The structure matrix gave the correlation of each variable with the discriminant function. The coefficients are structure coefficients or discriminant loadings. The largest loading on the factor was AFE and helps in the naming of the factor. These discriminant unstandardized coefficients indicated the standard contribution and therefore the relative importance of each variable to the discriminant function controlling for all other variables in the equation. They were utilized to create the discriminant equation. Of the two selected variables, AFE made the biggest contribution to the discriminant equation. The classification table showed the data on the diagonal. This revealed a high level accuracy of classification. The cross-validated set of data is a more reliable presentation of the power of the discriminant function. This procedure categorized all cases but 1 to develop a discriminant function and then categorized the case that was left out. The categorization showed that early maturing class was classified with slightly better accuracy (93.8%) than the late maturing class (92.9%). On the overall, 93.3% of cross-validated classed cases were correctly classified indicating that the overall predictive accuracy (hit-ratio) of the DF was high (Ramayah et al., 2010; Sage, 2014). Since the data set was used to compute the DA function as well as the classification, this leads to over-estimation of the hit ratio. Therefore a validation data set should be used to estimate the hit ratio appropriately. As a result one obtains the probability of belonging to a specific class g and also the probability for second best class (Walde, 2014). The range of scores on the axes on the histogram, the class centroids and the small overlap of the graphs, revealed substantial discrimination between classes, and this suggested that the function did discriminate well as indicated on reported tables. The F test that was conducted on the discriminant scores signified that the model differentiated between the classes significantly better than chance (a model with just the constant; Garson, 2012).

Many researchers have used discriminant analysis to solve many problems. Rosario et al. (2008) applied canonical discriminant analysis to classify a combination of four broiler chicken strains based on performance, in relation to average feed intake, average live weight, feed conversion, carcass weight, breast weight and leg weight. He used the contrast between average feed intake and average live weight plus feed conversion to classify them.

Udeh (2014) utilized **canonical correlation analysis to relate age at first egg, bodyweight at first egg and weight of first egg with egg production at different periods in a strain of layer type chicken. He found that** Canonical weights and loadings from canonical correlation analysis showed that weight of first egg had the largest contribution to the variation in egg number at the three different periods compared with AFE and HWFE. He concluded therefore that HWFE could be used as a selection criterion for selecting good performance layers in terms of egg number. Ogah (2013) used canonical discriminant analysis to study morphometric traits in indigenous chicken genotypes. Eskindir et al. (2013) used the discriminant analysis to conduct phenotypic characterization of indigenous chicken population in Ethiopia. He correctly classified 86.2% of Horro ecotype and 80.4% of Jarso ecotype. Gwaza (2013) also used discriminant analysis in studying morphological traits in selected population of the Tiv ecotype chicken in Nigeria and concluded that linear body parameters and weight measurements can be used to increase consistency of individuals in a population and separation of individuals between populations. Yakubu et al. (2011) studied morphometric traits of Muscovy Ducks from two agro-ecological zones of Nigeria and concluded that there was an indication of high gene flow between ducks from the two agro-ecological zones and homogeneity of the genetic identity of the duck populations.

Johnson et al. (2000) in his preliminary findings on the use of discriminant function analysis as a microbial source tracking technique discussed that pets and geese appeared to be major contributors to the contamination of the Manokin River.

This research was similar to those of Rosario et al. (2008) and Udeh (2014) who worked on production in broilers and reproduction in layers respectively. The study has elucidated AFE and HWFE as related and important in classifying ESM in Parent stock pullet breeders, corroborated the conclusions of Udeh (2014) and therefore implied that in the absence of weight of first egg, then Age at first egg could be an alternative trait for classification for ESM in parent stock layer breeders and commercial pullet layers. Using above model to classify new and unknown cases would assist in assessing the predictive validity of the DF constructed. Mahalanobis distance is the distance between a case and the centroid for each class of the dependent variable in the DF model and this distance is measured in terms of SD from the centroid. Since a new case will have one Mahalanobis distance for each class, it could be classified as belonging to the class for which its distance is smallest or closest (Uddin et al., 2013; Sage, 2014). One could also compute the classification scores directly for the new case from the DF model and subsequently classify it to the class for which it has the highest score (Kumar, 2014).

Conclusions

Multivariate analysis based on canonical discriminant analysis has demonstrated its suitability for evaluating layer chicken strains for early sexual maturity (ESM) Class because there was a reduction from eight original traits to only two canonical variables. Age at first egg (AFE) and hen weight at first egg (HWFE) were the two important traits that discriminated between ESM classes in PS pullets studied. There was a clear distinction between the two ESM classes in pullet strains as presented by the multivariate mean performances and class centroids.

References

Gwaza, D. S.; Tor, N .E. T.; and Wamagi, T. I. 2013. Discriminant Analysis of Morphological Traits in Selected Population of the Tiv Local Chicken Ecotype in the Derived Guinea Savannah of Nigeria. IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS). Volume 3 (6): 60-64. e-ISSN: 2319-2380, p-ISSN: 2319-2372. Retrieved in December 2014 from www.iosrjournals.org

Dunnington, E. A. and Siegel, P. H. 1984. Age and body weight at sexual maturity in female white leghorn chickens. Poultry Science 64 (4): 828-830.

Eskindir, A.; Kefelegn, K.; Tadelle D. and Banerjee A. 2013. Phenotypic Characterization of Indigenous Chicken Population in Ethiopia. International Journal of Interdisciplinary and Multidisciplinary Studies. Vol 1 (1): 24-32. Retrieved in December 2014 from httt://www.ijims.com

Garson, G. D. 2012. Discriminant Function Analysis. Asheboro, NC: Statistical Associates Publishers. Retrieved in December 2014 from http://seoul-humanlab.com/web/bbs/board.php?bo_table =83&wr_id=31

Johnson, Y. J.; Zimmerman, N. and Neavear, L. M. 2000. Preliminary findings on the use of discriminant function analysis as a microbial source tracking technique. Proceedings of the 9th International Symposium on Veterinary Epidemiology and Economics, 2000. Retrieved in December 2014 from www.sciquest.org.nz

Kumar, N. 2014. Discriminant Analysis – Database Marketing (PPT presentation). Retrieved in December 2014 from www.utdallas.edu ~ nkumar/DBMS4.ppt.

Ogah, D. M. 2013. Canonical discriminant analysis of morphometric traits in indigenous chicken genotypes. Trakia Journal of Sciences. 2: 170-174. Retrieved in December 2014 from http://www.uni-sz.bg

Ramayah, T., Ahmad, N. H., Halim, H, A., Zainal, S. R. M. and Lo, M. 2010. Discriminant analysis: An illustrated example. African Journal of Business

Management Vol. 4 (9): 1654-1667. Retrieved in December 2014 from http://www.academicjournals.org/AJBM. ISSN 1993-8233.

Rosario, M. F., Silva, M. A. N., Coelho, A. A. D., Savino, V. J. M. and Dias, C. T. S. 2008. Canonical discriminant analysis applied to broiler chicken performance. Animal. 2 (3): 419 – 424. The Animal Consortium 2008 doi: 10.1017/S1751731107001012

Sage, 2014. Chapter 25 Discriminant Analysis. PDF. Retrieved in December 2014 from www.uk.sagepub.com/

SPSS® Statistics 17.0. 2007. Statistical Package for Social Sciences Software Package. SPSS Incorporated, Illinois. USA.

Udeh, I. 2014. Canonical Correlation Analysis Relating Age at First Egg, Bodyweight at First Egg and Weight of First Egg With Egg Production at Different Periods In A Strain of Layer Type Chicken. Global Journal of Animal scientific Research. [S.I.] Vol 2 (4): 310-314. Jul. 2014. ISSN 2345-4385. Retrieved on 24th December 2014 from http://www.gjasr.com/index.php/GJASR /article/ view/ 73/258.

Uddin, N., Meah, M. S. and Hossain, R. 2013. Discriminant analysis as an aid to human resource selection and human resource turnover minimization decisions. International Journal of Business and Management; Vol. 8 (17): 153 – 169. ISSN 1833-3850. E-ISSN 1833-8119.

Walde, J. 2014. Discriminant Analysis. Retrieved in December 2014 from www.uibk.acat/

Wright, D.; Rubin, C.; Schutz, K.; Kerje, S.; Kindmark, A.; Brandstrom, H.; Andersson, L.; Pizzari, T. and Jensen, P. 2012. Onset of sexual maturity in female chickens is genetically linked to loci associated with fecundity and a sexual ornament. Reproduction in domestic Animals. Zuchthygiene 47 (Suppl. 1: 2012 Jan.): 31-36.

Yakubu, A; Kaankuka, F. G. and Ugbo, S. B. 2011. Morphometric traits of Muscovy Ducks from two agro-ecological zones of Nigeria. Tropicultura. Vol. 29 (2): 121-124.