African Journal of Mathematics and Statistics Studies

ISSN: 2689-5323

Volume 4, Issue 1, 2021 (pp. 10-31)

www.abjournals.org

LINEAR DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION IN CLASSIFICATION AND PREDICTIVE MODELING: A COMPARATIVE APPROACH

Onuoha Desmond O. and Obimgba Blessing

Department of Mathematics and Statistics, Fed. Polytechnic Nekede, Owerri, Imo State.

E-mail: oonuoha@fpno.edu.ng

ABSTRACT: *The goal of this study was to compare two different methods of classification; Linear Discriminant Analysis and Multinomial Logistic Regression to make the choice between the two, depending on the characteristics of the data. Since both are appropriate for the development of linear classification models, Linear Discriminant Analysis makes more assumptions like normality and equal covariance among the explanatory variables on the underlying data, but when violated it is assumed that the Multinomial Logistic Regression is a more flexible and more robust method of analysis. In this work, some guidelines for proper choice were set up which was based on some predictive accuracy. The performance of the methods was studied by a real dataset and a simulated dataset. We started with the real dataset where all the assumptions failed, also, we performed an appropriate transformation on the real dataset and Linear Discriminant Analysis was performed on it. Next we compare with simulated data where all the assumptions of Linear Discriminant Analysis are satisfied. From the result where the assumptions were violated, Multinomial Logistic Regression performs better than Linear Discriminant Analysis, also the result from the analysis performed on the transformed data shows that the Multinomial Logistic Regression also performed better, and whenever the assumptions hold as in the case of the simulated data, Linear Discriminant Analysis slightly performs better. Hence Multinomial Logistic Regression serves as an alternative whenever the assumptions of discriminant analysis fail instead of transforming the data.*

KEYWORDS: Linear Discriminant Analysis, Multinomial Logistic Regression, Data, Classification & Predictive Modeling

INTRODUCTION

Background of Study

In the society today, where we have existence of diverse populations which are made up of variety or similar features, a situation may arise where we may be interested in knowing exactly which of these entities have actually the same characteristics or which of these entities are of different character. Hence, the problem of discrimination, classification and allocation sets in. For example, given a set of students with observed performance on an examination, we may wish to divide their standard of performance into success and failure and the points where we affect these divisions are arbitrary; (Ogum, 2002).

Linear Discriminant Analysis (LDA) makes assumptions of normality and equality of the unknown variance-covariance matrices about the underlying data, Multinomial Logistic Regression (MLR) makes no assumptions on the distribution of the data. Therefore, when the assumptions of the data are met, it is expected that the Linear Discriminant Analysis (LDA) produces accurate and good results than Multinomial Logistic Regression. This study puts its focus on the evaluation and comparison on the performance of both Linear Discriminant Analysis (LDA) and Multinomial Logistic Regression. Since the normality assumption is a popular issue studied by many researchers, however in practice, the assumptions are nearly always violated and therefore will check the performance of both methods with simulations. So setting some guidelines for proper choice between the two methods is required.

The major aim of this work is to compare the ability of Linear Discriminant Analysis (LDA) and Multinomial Logistic Regression (MLR) in classification and predictive modeling using real dataset and simulated dataset. Other objectives to observe include; to determine whether or not the group mean vectors are equal, to construct classification function using both Linear Discriminant Analysis (LDA) and Multinomial Logistic Regression (MLR), to evaluate the performance of the classification functions using APER and to determine how these two classification functions- how Linear Discriminant Analysis (LDA) and Multinomial Logistic Regression (MLR) behave when real dataset and simulated dataset are used. This study covers a sample 50 real dataset having eight (8) predictive variables on Multiple Banking, Age of Applicant, Collateral, Years of Experience in Business, Interest Charged, Age of Account before Facility, Amount Granted and Tenor in months; extracted from work done by Ugwuanyi (2014) and simulated dataset for a sample of 100.

LITERATURE REVIEW

According to Efron (1975), he studied the relative efficiency of Logistic Regression and discriminant analysis. He presented that while LR is less efficient, estimates will have greater variance than LDA when the data are multivariate normal, and it is robust to departures from normality. He found that typically, LR is between one half and two thirds as effective as normal discrimination.

Hossain et al. (2002)., compared the performance of Multinomial Logistic Regression (MLR) and Linear Discriminant Analysis (LDA) models to predict arrival time at the hospital. The goal was to determine the best statistical methods for prediction of arrival intervals for patients with acute myocardial infarction symptoms. One model for Multinomial Logistic regression and two models Linear Discriminant Analysis were developed using a training dataset. Correct classifications were 62.6% by MLR, 62.4% by LDA using proportional prior probabilities, and 48.1% using equal prior probabilities of the groups.

Kiang (2003) used a simulated datasets and compared the performance of neural networks and a decision tree method, and three statistical method Linear Discriminant Analysis (LDA), Logistic Regression Analysis and Kth nearest-neighbor (kNN) models in terms of the misclassification rates and used synthetic data to introduce imperfections such as nonlinearity, multicollinearity, unequal covariance and to understand the strengths and limitations of different classification methods and the effects of data characteristics on their performance in a controlled setting. Kiang concluded that Neural Network and Logistic Regression methods

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31) www.abjournals.org

provide the best relative performance under most scenarios and he showed that there is no single method that clearly outperforms all methods in all problem situations and the Logistic Model is superior to Discriminant Analysis in all cases, especially when the normality, linearity, and identical covariance assumptions do not hold and only the normality assumption has an on Discriminant Analysis.

Other researchers on this area includes Press and Wilson (1978), Maroco et al (2011), Ugwuanyi (2014), Antonogeorgos et al (2009), Polar et al (2004), Majed (2012), Manel (1999), Montgomery et al (1987), Brenn and Arnesen (1985).

METHODOLOGY

Data for this study is from a secondary source. It was extracted from a work done by Ugwuanyi (2014) on "evaluation of loans and advances using Discriminant Analysis". From the nature of the data, a random sample of 50 performing credits and non-performing credits was selected from five selected commercial banks by the method of simple random sampling and the second dataset was a simulated data with a sample of size 100, which Linear Discriminant Analysis and Multinomial Logistic Regression Analysis were used for data analysis.

Discriminant Analysis

Discriminant Analysis is a multivariate technique that is concerned with separating distinct sets of objects and with allocating new objects into previously defined groups (Johnson and Wichern, 2007). Discriminant Analysis is a powerful statistical tool that is concerned with the problem of classification. This problem of classification arises when an investigator makes a number of measurements on an individual and wishes to classify the individual into one of the several population groups on the basis of these measurements (Morrison, 1967). Once a Discriminant Analysis is contemplated, it may be important to check whether or not the research data satisfy the assumptions of Discriminant Analysis. Some of the assumptions to be examined include; equality of group covariance matrices and normality assumption.

Prior to discussing the above mentioned assumptions we shall first discuss the first step in the conduct of Discriminant Analysis - test of equality of mean vectors. If no significant differences are found, constructing a classification rule will probably be a waste of time (Ogum. 2002). To test the hypothesis that the mean vectors of the two groups under study are equal, the Hotelling's T^2 distribution will be used (Ogum, 2002).

Hotelling T^2 Distribution for Two Samples

According to Onyeagu (2003), Hotelling T^2 distribution is a multivariate generalization of the student's t-distribution for testing equality of group mean vectors; the hypotheses of interest are;

H_{01}: $\mu_1 = \mu_2\mu_1 = \mu_2$ The group means vectors of the research data are equal

H_{11}: $\mu_1 \neq \mu_2\mu_1 \neq \mu_2$The group means vectors of the research data are not equal (3.1)

The Hotelling's T^2 test statistic is given as,

$$F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T^2 F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T^2 \qquad (3.2)$$

Where

$$T^2 = \frac{n_1 n_2}{n_1+n_2} D^2 = \frac{n_1 n_2}{n_1+n_2} (\bar{x}_1 - \bar{x}_2)^T S^{-1} (\bar{x}_1 - \bar{x}_2)$$

$$T^2 = \frac{n_1 n_2}{n_1+n_2} D^2 = \frac{n_1 n_2}{n_1+n_2} (\bar{x}_1 - \bar{x}_2)^T S^{-1} (\bar{x}_1 - \bar{x}_2) \qquad (3.3)$$

Thus, equation (3.2) follows an F distribution with p and u degrees of freedom, where

$$\frac{1}{u} = \frac{1}{n_1-1}\left[\frac{\left(\bar{X}_1-\bar{X}_2\right)^T \boldsymbol{S}^{-1}\frac{\boldsymbol{S}_1}{n_1}\boldsymbol{S}^{-1}\left(\bar{X}_1-\bar{X}_2\right)}{T^2}\right]^2 + \frac{1}{n_2-1}\left[\frac{\left(\bar{X}_1-\bar{X}_2\right)^T \boldsymbol{S}^{-1}\frac{\boldsymbol{S}_2}{n_2}\boldsymbol{S}^{-1}\left(\bar{X}_1-\bar{X}_2\right)}{T^2}\right]^2 \qquad (3.4)$$

Fisher's Linear Discriminant Function assumes that the population covariance matrices across the groups are equal, because a pooled estimate of the common covariance matrix is used. To determine whether the covariance matrices for the groups under study are equal, the Box's M test will be used (Box, 1949). The calculation of Box's M test proceeds as follows. Suppose we have k groups measured on each of p variance, with n_i observations per group. Then, we estimate the within-group covariance as;

$$\boldsymbol{S}_i = \frac{\sum_{i=1}^{k}\left(X_i - \bar{X}\right)^T\left(X_i - \bar{X}\right)}{n_i - 1}, \quad i = 1, 2, \ldots, k \qquad (2.5)$$

The value of M is then calculated by $M = (N-K)In\,|S| - \sum_{i=1}^{k}(n_i-1)In|S_i|$ (3.6)

Where, $N = \sum_{i=1}^{k} n_i$

(3.7)

The Chi-square and F-ratio are used to test the significance of the value of M. These tests proceed as follows

$$A_1 = \frac{2p^2+3p-1}{6(p+1)(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i-1}\right) - \frac{1}{N-k}\right] \qquad (2.8)$$

$$A_2 = \frac{(p-1)(p+2)}{6(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i-1}\right)^2 - \frac{1}{(N-k)^2}\right] \qquad (2.9)$$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31) www.abjournals.org

$$\text{If} \quad A_2 - A_1^2 > 0, \text{ then}$$

$$V_1 = \frac{p(p+1)(k-1)}{2}. \quad (2.10)$$

$$V_2 = \frac{V_1 + 2}{A_2 - A_1^2} \quad (2.11)$$

$$b = \frac{V_1}{1 - A_1 - \left(\frac{V_1}{V_2}\right)} \quad (2.12)$$

We then carry out the following hypotheses:

$$H_0 : \sum_1 = \sum_2 = \sum_3 = \cdots = \sum_K$$
$$H_1 : \text{At least one of the } \sum_i, (i = 1, 2, \cdots, k) \text{ is unequal.} \quad (3.13)$$

Therefore, the F statistic when $A_2 - A_1^2 > 0,$ is then given by:

$$F = \frac{M}{b} \quad (2.14)$$

On the contrary, if $A_2 - A_1^2 < 0,$ then

$$V_2 = \frac{V_1 + 2}{A_1^2 - A_2} \quad (2.15)$$

$$b = \frac{V_2}{1 - A_1 + \left(\frac{2}{V_2}\right)} \quad (2.16)$$

Therefore, the F statistic when $A_2 - A_1^2 < 0,$ is given by

$$F = \frac{V_2 M}{V_1 (b - M)} \quad (2.17)$$

We shall reject the null hypothesis of equality of covariance matrices at a specified level of significance, α, if $F > F_{\alpha, V_1 V_2}$, otherwise do not reject.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

The Chi-square test statistic for conducting test of equality of covariance matrices when $(either \quad A_2 - A_1^2 < 0 \ or \ A_2 - A_1^2 > 0)$ is given as $\chi^2 = M(1 - A_1)$ (3.18)

Multivariate Normality

Discriminant Analysis assumes that data for the independent variables represent a sample from a multivariate normal distribution. To determine whether multivariate normality assumption is justified we shall employ the Q-Q plots from any Statistical software, precisely SPSS.

Outliers

Outliers are unusual observations that do not seem to belong to the pattern of variability produced by other observations. To check for the existence of outliers for a single variable, we make a dot plot and then look for observations that are far from the others, however, when the number of variables, observations *n* is large, dot plots are not feasible because the large number of scatter plots *p(p* - 1)/2 may prevent viewing them all. Here a large value of $d_j^2 = \left(X_j - \overline{X}\right)^T \boldsymbol{S}^{-1}\left(X_j - \overline{X}\right)$ will suggest an unusual observation, even though it cannot be seen visually. The steps for detecting outliers are outlined below;

Step 1. Make a dot plot for each variable.

Step 2. Make a scatter plot for each pair of variables.

Step 3. Calculate the standardized values $z_{jk} = \frac{\left(x_{jk} - \overline{x}\right)}{\sqrt{s_{kk}}}$ for j=1,2,…,n and each column *k* = 1,2, ... , *p*. Examine these standardized values for large or small values. In step 3, "large" must be interpreted.

Step 4. Calculate the generalized squared distances $d_j^2 = \left(X_j - \overline{X}\right)^T \boldsymbol{S}^{-1}\left(X_j - \overline{X}\right)$. Examine these distances for unusually large values. In a chi-square plot, these would be the points farthest from the origin. In step 4, "large" is measured by an appropriate percentile of the chi-square distribution with *p* degrees of freedom. If the sample size is *n* = 100, we would expect5 observations to have values of d_j^2 that exceed the upper fifth percentile of the chi-square distribution. A more extreme percentile must serve to determine observations that do not fit the pattern of the remaining data.

Theoretical Consideration

This theoretical consideration has three aspects in terms of data transformation to near normality, which are as follows:

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)

Original Scale	**Transformed Scale**
❖ Counts y	$\sqrt{y}\sqrt{y}$
❖ Proportion, $\hat{p}\hat{p}$	Logit $(\hat{p})\hat{p})$ = $\frac{1}{2}\log\left(\frac{\hat{p}}{1-\hat{p}}\right)\frac{1}{2}\log\left(\frac{\hat{p}}{1-\hat{p}}\right)$
❖ Correlations, r	Fisher's $z(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)z(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$

The Data Themselves

The theoretical transformation mentioned above may sometimes not improve the normality in question. It is more convenient to let the data suggest a transformation. A useful family of transformation for this purpose is the family of power transformation. The final choice should always be examined by a Q-Q plot or other checks to see whether the tentative normal assumption is satisfactory.

Fisher's Linear Discriminant Function

The steps for deriving Fisher's linear Discriminant Function (FLDF) are stated below: (Johnson and Wichern, 2002)

Step 1: Obtain the FLDF using

$$Y = \left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\right)^T \mathbf{S}^{-1} X = a^T X \qquad (2.28)$$

Where

$$\underset{(P\times 1)}{\overline{\mathbf{X}}_1} = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j} \qquad (2.29)$$

$$\underset{(P\times 1)}{\overline{\mathbf{X}}_2} = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j} \qquad (2.30)$$

$$\underset{(P\times P)}{\mathbf{S}_1} = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}\left(X_{1j} - \overline{\mathbf{X}}_1\right)\left(X_{1j} - \overline{\mathbf{X}}_1\right)^T \qquad (2.31)$$

$$\underset{(P\times P)}{\mathbf{S}_2} = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}\left(X_{2j} - \overline{\mathbf{X}}_2\right)\left(X_{2j} - \overline{\mathbf{X}}_2\right)^T \qquad (2.32)$$

$$\underset{(P\times P)}{\mathbf{S}} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2} \qquad (2.33)$$

Where,

$$\mathbf{S}_{(P\times P)}^{-1} = \frac{Adj.(\mathbf{S})}{Det.(\mathbf{S})} \qquad (2.34)$$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

$$\underset{(p\times 1)}{\boldsymbol{X}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \qquad (2.35)$$

Step 2: For population 1 (π_1), evaluate the discriminant function in equation by substituting the mean value of $x_1, x_2, ..., x_p$ and denote the values as $\bar{Y}_1$. That is,

$$\bar{Y}_1 = \left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\bar{X}_1 = a^T \bar{X}_1 \qquad (2.36)$$

Step 3: For population 2 (π_2), evaluate the discriminant function in equation by substituting the mean values of $x_1, x_2, ..., x_p$ and denote the values as $\bar{Y}_2$. That is,

$$\bar{Y}_2 = \left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\bar{X}_2 = a^T \bar{X}_2 \qquad (2.37)$$

Step 4: Obtain the critical value of the discriminant function as

$$Y_{Critical} = \frac{\bar{Y}_1 + \bar{Y}_2}{2} = \frac{\left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\bar{X}_1 + \left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\bar{X}_2}{2}$$
$$= \frac{\left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\left(\bar{X}_1 + \bar{X}_2\right)}{2} = \frac{a^T\left(\bar{X}_1 + \bar{X}_2\right)}{2}$$

Step 5: Obtain the Discriminant scores, y_0, for the population 1 (π_1) and population 2 (π_2) respectively by substituting the values of $x_1, x_2, ..., x_p$ for each individual into the Discriminant function in equation (3.19).

Step 6: State the classification Rules as follows:

(a) Allocate the individual x_0 to population 1 (π_1) if the Discriminant score, y_0, is at least the critical value of the Discriminant Function. That is, if

$$y_0 = \left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}X_0 \geq Y_{Critical} = \frac{1}{2}\left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\left(\bar{X}_1 + \bar{X}_2\right) \qquad (2.39)$$

(b) Allocate an individual x_0 to population 2 (π_2) if the Discriminant score y_0 is less than the critical values of the Discriminant Function. That is, if

$$y_0 = \left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}X_0 < Y_{Critical} = \frac{1}{2}\left(\bar{X}_1 - \bar{X}_2\right)^T \boldsymbol{S}^{-1}\left(\bar{X}_1 + \bar{X}_2\right) \qquad (2.40)$$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

The objective of building a classification rule is to correctly classify as many future units as possible. A good discriminant function should result in a few misclassifications. In other words, the probabilities of misclassification should be small. To judge how well the discriminant function performs in assigning an observation to the correct population, we shall use the Apparent Error Rate (APER) (Johnson and Wichern, 2002).

The APER is defined as the fraction of observations in the training sample that are misclassified by the sample discriminant function. A sample estimate of the error rate can be obtained by trying out the discriminant function on the same data set that has been used to compute the discriminant function. This is called Substitution or Resubstitution Method (Render, 2002). The result can be displayed in a classification table called *Confusion Matrix*, which shows the actual versus predicted group membership. For n_1 from population 1 (π_1) and n_2 from population 2 (π_2), the confusion matrix has the form shown in table 3.1.

Table 3.1: Layout of Confusion Matrix for Discriminant Analysis

Actual Membership	Predicted Membership		Number of Observations
	To Population1 (π_1)	To Population2 (π_2)	
From Population1 (π_1)	n_{1C}	$n_{1m} = n_1 - n_{1C}$	n_1
From Population2 (π_2)	$n_{2m} = n_2 - n_{2C}$	n_{2C}	n_2

Where,

n_{1C} = number of π_1 items correctly classified as π_1 items

n_{1m} = number of π_1 items misclassified as π_2 items

n_{2C} = number of π_2 items correctly classified as π_2 items

n_{2m} = number of π_2 items misclassified as π_1 items

The apparent error rate is then given as

$$APER = \frac{n_{1m} + n_{2m}}{n_1 + n_2} \qquad (2.47)$$

which is recognized as the proportion of items in the training set that are misclassified.

The APER is intuitively appealing and easy to calculate. Unfortunately, it tends to underestimate the actual error rate (AER), and the problem does not disappear unless the sample sizes n_1 and n_2 are very large. Essentially, this optimistic estimate occurs because the data used to build the classification function are also used to evaluate it. We can also compute error rates which are better than the apparent error rate, and do not require distributional assumptions. One procedure is to split the total sample into a training sample and a validation sample. The training sample is then used to construct the classification function, while the validation sample is used to evaluate the performance of the classification function. In this case the error rate is determined by finding the proportion of misclassified observation in the validation sample. Although this method overcomes the bias problem in APER by not using the same data to both build and judge the classification function, it suffers from two main defects:

(i) It requires large samples.

(ii) The function evaluated is not the function of interest. Ultimately, almost all of the data must be used to construct the classification function. If not, valuable information may be lost.

A second approach that seems to work well is called Lachenbruch's "holdout" procedure. (Johnson and Wichern, 2007). The steps for this procedure are;

Step 1. Start with the π_1 group of observations. Omit one observation from this group, and develop a

classification function based on the remaining n_1-1, n_2 observations.

Step 2. Classify the "holdout" observation, using the function constructed in Step 1.

Step 3. Repeat Steps 1 and 2 until all of the π_1 observations are classified. Let $n_{1m}^{(H)}$ be the number of

holdout *(H)* observations misclassified in this group.

Step 4. Repeat Steps 1 through 3 for the π_1 observations. Let $n_{2m}^{(H)}$ be the number of holdout observations

misclassified in this group.

Thus, estimates of P(2/1)and P(1/2) of the conditional misclassification probabilities are given by

$$\hat{P}(2/1) = \frac{n_{1m}^{(H)}}{n_1} \tag{2.48}$$

$$\hat{P}(1/2) = \frac{n_{2m}^{(H)}}{n_2} \tag{2.49}$$

and the total proportion misclassified, $\left(n_{1m}^{(H)} + n_{2m}^{(H)}\right)/(n_1 + n_2)$, is, for moderate samples, a nearly unbiased estimate of the expected actual error rate, *E(AER).*

$$\hat{E}(AER) = \frac{n_{1m}^{(H)} + n_{2m}^{(H)}}{n_1 + n_2} \tag{2.50}$$

Multinomial Logistic Regression

Regression Analysis has become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

This response variable is always quantitative in nature. A situation where the response variable becomes dichotomous, that is can only take two values say 0 and 1, regression analysis becomes insignificant hence Multinomial Logistic Regression comes into play. Even though the response may be a two outcomes qualitative variable, we can always code the two cases say 0 and 1 for instance. If we are analyzing groups; we can say group 1 = 0 and group 2 = 1, then the probability of 1 is a parameter of interest. It represents the proportion in the population which is coded 1. The mean of the distribution of 0 and 1 is p, and variance p(1-p) since:

Mean = 0 x (1-p) + 1 x p = p (3.36)

Variance = $0^2 x\ (1-p) + 1^2 x\ (p-p^2) = p(1-p)$
$0^2 x\ (1-p) + 1^2 x\ (p-p^2) = p(1-p)$ (3.52)

The probability model of 1 is given by

P = E(Y/X) = $B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_KB_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K$

The Logit Model

Instead of modeling the probability of Y directly with a linear model, we first consider the odd ratios.

Odds = $\frac{p}{1-p}\frac{p}{1-p}$
(3.39)

This is the probability of 1 to the probability of 0. This is used when we want to predict the probability that a case will be classified into one as opposed to the other of the two categories of the independent variables. If we know one probability, we know the other. From the linear model in 3.53, X can be categorical or continuous, but Y (p) is always categorical. The value Y must lie between 0 and 1. To solve this problem we replace the probability that Y=1 with the odds that Y=1. The odds that Y = 1, written odd (Y=1) is the ratio of the probability that Y = 1 to the probability that Y $\neq\neq$ 1. The odds that y = 1 is given as:

Odds = $\frac{P(Y=1)}{1-P(Y=1)}\frac{P(Y=1)}{1-P(Y=1)}$ (3.40)

In Logistic regression for a binary variable, we model the natural log of the odds ratios which is called Logit (Y); thus:

Logit (Y) = In(odds) = $In\left(\frac{P(Y=1)}{1-P(Y=1)}\right)In\left(\frac{P(Y=1)}{1-P(Y=1)}\right)$

The equation for the relationship between dependent variables and independent variables then

becomes: Logit (Y) = $B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_KB_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)

This is the Logistic Regression model which is Linear in its predictor's variables. Because it is easier for most people to think in terms of probability, we can convert from the Logit or log odds to the probability of Y, by exponentiating:

Odds(Y=1) = $e^{logit(y)} e^{logit(y)}$

Odds(Y=1) = $e^{In(odd(Y=1)} = e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K} e^{In(odd(Y=1)} = e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}$

We can then convert the odds back to the probability that (Y=1) by the formula

$$P(Y=1) = \frac{Odds\,(Y=1)}{1+Odds\,(Y=1)} \frac{Odds\,(Y=1)}{1+Odds\,(Y=1)} \qquad (3.44)$$

$$\text{This gives } P(Y) = \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} \qquad (3.45)$$

This equation has the desired property that no matter what values we substitute for the B's and the X's; P will always be a number between 0 and 1.

Logistic Classification Rule

Assign X to population 1 if the estimated discriminant scores is greater than 1

$$(Y) = B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K > 1 \qquad (3.46)$$

Assign X to population 2 if the estimated discriminant scores is less than o

$$(Y) = B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K < 0 \qquad (3.47)$$

Alternatively;

Assign X to population 1 if the probability function is equal to1

$$P(Y) = \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} = 1 \qquad (3.48)$$

Assign X to population 2 if the probability function is less than 1

$$P(Y) = \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} < 1 \qquad (3.49)$$

DATA PRESENTATION, VALIDATION, ANALYSIS AND INTERPRETATIONS

Data Presentation

Data for this study are presented on Table 1, Table 2, and Table 3 of Appendix 1 of this work.

Validation of Assumptions of the Data

Before testing the assumptions, we shall first test for equality of mean vectors. The test is presented hereunder.

Test of Equality of Mean Vectors Using Hotellings' T^2 –Distribution for Real Dataset

Therefore, we shall test the following hypotheses;

$$H_{01}: \mu_1 = \mu_2 (The\ \ group\ mean\ vectors\ are\ equal)$$

$H_{01}: \mu_1 = \mu_2 (The\ \ group\ mean\ vectors\ are\ equal)$

$$H_{11}: \mu_1 \neq \mu_2 (Thegroup\ mean\ vectors\ are\ not\ equal)$$

$H_{11}: \mu_1 \neq \mu_2 (Thegroup\ mean\ vectors\ are\ not\ equal)$

The Hotelling's T^2 test statistic is given as

$$F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2 F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

Where

$$T^2 = \frac{n_1 n_2}{n_1+n_2}D^2 = \frac{n_1 n_2}{n_1+n_2}(\bar{x}_1 - \bar{x}_2)^T S^{-1}(\bar{x}_1 - \bar{x}_2)$$

$D^2 = 1.80500802$

$$T^2 = \frac{(50)(50)}{50+50}(1.80500802) T^2 = \frac{(50)(50)}{50+50}(1.80500802) \quad => \quad T^2 = 45.1252005$$

$T^2 = 45.1252005$

$$F = \frac{50+50-8-1}{8(50+50-2)}(45.1252005) F = \frac{50+50-8-1}{8(50+50-2)}(45.1252005) \; => \quad F = \frac{91}{784}(45.1252005)$$

$F = 5.2377464 F = 5.2377464$

Conclusion: Since, $F = 5.2377 > F_{\alpha,p,n_1+n_2-p-1} = F_{0.05\ ,2,91} \equiv 2.0435$

$F = 5.2377 > F_{\alpha,p,n_1+n_2-p-1} = F_{0.05\ ,2,91} \equiv 2.0435$, we therefore reject the null hypothesis, H_0 ,and conclude that there is a significant difference between the mean vectors of the performing credits and non-performing credits at $\alpha = 0.05$.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

Test of Equality of Mean Vectors Using Hotellings' T^2 –Distribution for Simulated Dataset

$$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Where

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^T S^{-1} (\bar{x}_1 - \bar{x}_2)$$

$D^2 = 4.170330265$

$$T^2 = \frac{(100)(100)}{100+100} (4.170330265) T^2 = \frac{(100)(100)}{100+100} (4.170330265) \quad => \quad T^2 = 208.5165133$$

$$T^2 = 208.5165133$$

$$F = \frac{100+100-2-1}{2(100+100-2)} (208.5165133) F = \frac{100+100-2-1}{2(100+100-2)} (208.5165133) \quad => \quad F = 103.7316998$$

$$F = 103.7316998$$

Conclusion: Since, $F = 103.732 > F_{\alpha,p,n_1+n_2-p-1} = F_{0.05,2,197} \equiv 3.07$

$F = 103.732 > F_{\alpha,p,n_1+n_2-p-1} = F_{0.05,2,197} \equiv 3.07$, we therefore reject the null hypothesis, H_0,and conclude that there is a significant difference between the mean vectors of the simulated data at $\alpha = 0.05$.

Test of Equality of Covariance Matrices Using Box M test for Real Dataset

Applying equations (3.6) – (3.17) on data of table 4.1 of Appendix 1, we obtain the following results.

We then carry out the following hypotheses:

$$H_0 : \Sigma_1 = \Sigma_2 \; (The\ covariance\ matrices\ are\ equal\ )$$
$$H_1 : \Sigma_1 \neq \Sigma_2 \; (The\ covariance\ matrices\ are\ unequal\ )$$

The test statistic for conducting Box' M test of equality of covariance matrices is the Chi-square test statistic, given by $\chi^2 = (1 - A_1)M = (1 - 0.022108843)(225.984171) = 220.9879224$

Conclusion: Since**,** $\chi^2 = 220.988 > \chi^2_{\alpha\ ,\ V_1} = 23.27$ **,**we reject the null hypothesis, H_0, and conclude that the two covariance matrices are not equal.

Test of Equality of Covariance Matrices Using Box M test for Simulated Dataset

Applying equations (3.6) – (3.17) on data of table 4.3 of Appendix 1, we obtain the following results

We then carry out the following hypotheses:

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

$H_0 : \Sigma_1 = \Sigma_2$ (*The covariance matrices are equal*)
$H_1 : \Sigma_1 \neq \Sigma_2$ (*The covariance matrices are unequal*)

The test statistic for conducting Box' M test of equality of covariance matrices is the Chi-square test statistic, given by $\chi^2 = (1 - A_1)M = (1 - 0.95436751)(7.5222411) = 0.343258591$

Conclusion: Since, $\chi^2 = 0.343 < \chi^2_{\alpha, V_1} = 0.352$, we do not reject the null hypothesis, H_0, and conclude that the two covariance matrices are equal.

Multivariate Normality Test for Real Dataset, Simulated Dataset and Transformed Dataset

With the aid of Statistical Package SPSS, we carry out the Q-Q plots for normality test and the results for these are presented on Figure 1 and Figure 2 and figure 3 of Appendix 2

Estimations for Discriminant Analysis on the Real Dataset

Applying equation 3.16 on data of Table 1 of Appendix 1, we obtain the Linear Discriminant function on the real dataset:

$Y^I =$
$-0.108444128x_1 + 0.045190766x_2 + 0.00008853834x_3 + 0.161799961x_4 + 0.092391747x_5 + 0.014216837x_6 - 0.000039861836x_7 - 0.31074202x_8$

$Y^I =$
$-0.108444128x_1 + 0.045190766x_2 + 0.00008853834x_3 + 0.161799961x_4 + 0.092391747x_5 + 0.014216837x_6 - 0.000039861836x_7 - 0.31074202x_8$
(4.1)

Where:

$Y^I = Discriminant\ function$

$X_1 = Multiple\ Banking$

$X_2 = Age\ of\ Applicant$

$X_3 = Collaterial$

$X_4 = Years\ of\ Experience\ in\ Business$

$X_5 = Interest\ Change$

$X_6 = Age\ of\ Account\ before\ Facility$

$X_7 = Amount\ Granted$

$X_8 = Tenure\ (Months)$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)

Applying equation (3.27) and (3.28) on data of Table 1 of Appendix 1, we obtain the Linear Discriminant mean of population I and population II respectively.

$\bar{Y}_1 = 4.895822162\bar{Y}_1 = 4.895822162$

And $\bar{Y}_2 = 2.780060419\bar{Y}_2 = 2.780060419$ (4.3)

Applying equation (3.29) on data of Table 1 of Appendix 1, we obtain the critical value as

$$Y_{Critical} = \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{4.895822162 + 2.780060419}{2} = 3.837941291 \quad (4.5)$$

Applying equation (4.1) on the data of Table 1 appendix 1, we obtain the discriminant scores in the Table 4 presented in Appendix 1:

Applying the Linear Discriminant classification rule as given in equation (3.30) and (3.31) on Table 4 of Appendix 1, we obtain the Confusion Matrix given in the table 4.1 below:

Table 4.1 Confusion Matrix of the Real Dataset

		Predicted Membership	
		π_1: *Performing Credit*	π_2 *Non* − *Performing Credit*
Actual Membership	π_1: *Performing Credit*	34	7
	π_2 *Non* − *Performing Credit*	16	43
	TOTAL	**50**	**50**

Applying equation (3.35) on dataset of Table 1, we obtain the error rate as follows:

$$\text{APER} = \frac{16+7}{50+50} = \frac{23}{100} = 0.23\frac{16+7}{50+50} = \frac{23}{100} = 0.23 \quad (4.7)$$

Hence the overall probability of Misclassification is

$$\text{APER} * 100 = 0.23*100 = 23\% \quad (4.8)$$

Estimations for Discriminant Analysis on the Transformed Dataset

With the help of statistical packages, precisely MINITAB 17, discriminant analysis was applied on the transformed dataset of table 2 of appendix 1, summary results is given below:

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)
www.abjournals.org

Table 4.2 Confusion Matrix of the Transformed Dataset

		Predicted Membership	
		π_1: *Performing Credit*	π_2 *Non* − *Performing Credit*
Actual Membership	π_1: *Performing Credit*	39	9
	π_2 *Non* − *Performing Credit*	11	41
	TOTAL	50	50

$$\text{APER} = \frac{11+9}{50+50} = \frac{20}{100} = 0.20\frac{11+9}{50+50} = \frac{20}{100} = 0.20 \qquad (4.9)$$

Hence the overall probability of Misclassification is

APER * 100 = 0.20*100 = 20% (4.10)

Estimations for Discriminant Analysis on the Simulated Data

Applying equation (3.16) on data of Table 3 of Appendix 1, we obtain the Linear Discriminant function as: $Y^1 = 0.063828892X_1 - 0.016909396X_2$

$$Y^1 = 0.063828892X_1 - 0.016909396X_2 \qquad (4.11)$$

Applying equation (3.30) and (3.31) on dataset of Table 3 of Appendix 1, we obtain the Linear Discriminant mean of population I and population II respectively;

$$\bar{Y}_1 = 11.50196914$$

$$\bar{Y}_2 = 7.331638898\bar{Y}_2 = 7.331638898 \qquad (4.12)$$

Applying equation (3.35) on dataset of Table 3 of Appendix 1, we obtain the critical value as

$$Y_c = \frac{\bar{Y}_1+\bar{Y}_2}{2} = \frac{11.50196914+7.331638898}{2} = \frac{18.83360804}{2} = 9.416804019$$

$$Y_c = \frac{\bar{Y}_1+\bar{Y}_2}{2} = \frac{11.50196914+7.331638898}{2} = \frac{18.83360804}{2} = 9.416804019 \qquad (4.13)$$

Applying equation (4.11) on the data of Table 1 of Appendix 1, we obtain the discriminant scores which are presented on the dataset of table 5 of Appendix 1.

Applying the Linear Discriminant classification rule as given in equation (3.30) and (3.31) on Table 5 of Appendix 1, we obtain the Confusion Matrix given in the table 4.3 below:

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)

Table 4.3 Confusion Matrix of the Simulated Dataset

		Predicted Membership	
Actual Membership		π_1	π_2
	π_1:	96	20
	π_2	4	80
	TOTAL	100	100

Applying equation (3.35) on data of Table 4.8, we obtain the error rate as follows:

$$\text{APER} = \frac{4+20}{100+100} = \frac{24}{200} = 0.12\frac{4+20}{100+100} = \frac{24}{200} = 0.12 \quad (4.14)$$

Hence, the overall probability of Misclassification is

APER * 100 = 0.12*100 = 12% (4.15)

Estimations of Multinomial Logistic Regression on the Real Life Data

With the aid of statistical package, MINITAB 17 software, we obtain the estimate of Logistic Regression Model using the dataset on Table 1 of Appendix 1, as follows:

$\boldsymbol{Y^{II}} = 2.46 - 0.118X_1 - 0.1653X_2 + 0.000241X_3 + 0.728X_4 + 0.020X_5 + 0.0815X_6 - 0.00026X_7 - 0.458X_8$
$\boldsymbol{Y^{II}} = 2.46 - 0.118X_1 - 0.1653X_2 + 0.000241X_3 + 0.728X_4 + 0.020X_5 + 0.0815X_6 - 0.00026X_7 - 0.458X_8$

(4.16)

Applying equation (4.16) to the data of Table 1of Appendix 1, we obtain the Discriminant scores and probability scores as presented in Table 6 of Appendix 1:

Applying the Logistic classification rule as given in equation (3.48) and (3.49) on Table 6 of appendix 1, we obtain the Confusion Matrix given in the table below:

Table 4.4 Confusion Matrix for the Real Dataset

		Predicted Membership	
Actual Membership		$\pi_1: Performing\ Credit$	$\pi_2\ Non-Performing\ Credit$
	$\pi_1: Performing\ Credit$	41	7
	$\pi_2\ Non-Performing\ Credit$	9	43
	TOTAL	50	50

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 10-31)

APER = $\frac{9+7}{50+50} = \frac{16}{100} = 0.16\frac{9+7}{50+50} = \frac{16}{100} = 0.16$
(4.17)

Hence the overall probability of Misclassification is APER * 100 = 0.16*100 = 16%
(4.18)

Estimations for Multinomial Logistic Regression on the Simulated Data

With the aid of statistical package, MINITAB 17 software, we obtain the estimate of Logistic Regression Model using the dataset on Table 3 of Appendix 1, as:

$Y^I = 11.91 - 0.0759X_1 + 0.0089X_2Y^I = 11.91 - 0.0759X_1 + 0.0089X_2$
(4.19)

Applying equation (4.19) to the data of Table 3 of Appendix 1, we obtain the Discriminant scores and probability scores which is also displayed in Table 6 of Appendix 1:

Applying the Logistic classification rule as given in equation (3.48) and (3.49) on Table 6 of Appendix 1, we obtain the Confusion Matrix given in the table below:

Table 4.5 Confusion Matrix Simulated Dataset

		Predicted Membership	
Actual Membership		π_1:	π_2
	π_1:	91	17
	π_2	9	83
	TOTAL	100	100

APER = $\frac{9+17}{100+100} = \frac{26}{200} = 0.13\frac{9+17}{100+100} = \frac{26}{200} = 0.13$
(4.20)

Hence probability of Mis-classification is

APER * 100 = 0.13*100 = 13% (4.21)

FINDINGS

- The Hotelling's T^2 distribution indicated that the mean vectors of the performing credits and those of Non-performing credits are significantly different, since the mean vectors are different there is need for discriminant analysis.
- The real dataset violated the discriminant assumptions, which are; normal and that of equality of covariance matrices using Q-Q plot and Box's M test respectively.
- The simulated dataset is multivariate normal and showed equality of covariance using Q-Q plot and Box's M test respectively.

- The critical value of the discriminant was found to be approximately $Y_{critical} = 3.837941291$

- The classification rule for allocating customers into performing credits and non-performing credits for Linear Discriminant Analysis were found to be:

  (a) Allocate customer s (x_0) to performing credit group (π_1) if $\hat{y}_0 = \hat{a}x_0 \geq \hat{y}_{critical} = 3.84\hat{y}_0 = \hat{a}x_0 \geq \hat{y}_{critical} = 3.84$.

  (b) Allocate customers (x_0) to Non-performing group (π_2) if $\hat{y}_0 = \hat{a}x_0 \leq \hat{y}_{critical} = 3.84\hat{y}_0 = \hat{a}x_0 \leq \hat{y}_{critical} = 3.84$

- The classification rule for allocating customers into performing credits and non-performing credits for Multinomial Logistic Regression is as given: Assign X to population 1 if the estimated discriminant scores is greater than 1; (Y) = $B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_KB_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K > 1$ Assign X to population 2 if the estimated discriminant scores is less than 0

  (Y) = $B_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_KB_0 + B_1X_1 + B_2X_2 + \cdots + B_KX_K < 0$

  Alternatively;

  Assign X to population 1 if the probability function is equal to1

  $$P(Y) = \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}\frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} = 1$$

  Assign X to population 2 if the probability function is less than 1

  $$P(Y) = \frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}\frac{e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}}{1+e^{B_0+B_1X_1+B_2X_2+\cdots B_KX_K}} < 1$$

- The Fisher Linear Discriminant Function indicated that about 23% were misclassified, while that of Multinomial Logistic Regression gave 16% based on the real dataset. The simulated data gave the result of misclassification as 12% and 13% for Linear Discriminant Analysis and Multinomial Logistic Regression, respectively. Also the transformed dataset gave 20% of misclassification.

CONCLUSION

In this experimental study we have compared two different methods of classification: Linear Discriminant Analysis and Multinomial Logistic Regression in terms of the classification accuracy and model performance. The performance evaluation was carried out on a real dataset and also performed a simulated study to examine the group and data characteristics that may affect the performance of LDA and MLR. As a conclusion this study was very helpful for us to make the choice between the two methods easier and to understand how the two models behave under different data and group characteristics.

REFERENCES

Antonogeorgos G, Demosthenes B, Kostas N, Anastasia T (2009). Logistic Regression and Linear Discriminant Analyses in Evaluating Factors Associated with Asthma Prevalence: Divergence and Similarity of the Two Statistical Methods. International Journal of Pediatrics; 2009.

Audu I. and Usman A. (2013). Application of Group-Dependent transformation of Fisher's Discriminant Analysis in a University of Technology. *International Journal of latest Research in Science and Technology, Vol. 2, Issue2, pp 40-45.*

Box G. E. P. (1949). A General Distribution theory for a class of likelihood.*Biometrika,Volume36, pp 317–346.*

Brenn T, Arnesen E. (1985). Selecting Risk Factors. A Comparison of Discriminant Analysis, Logistic Regression and Cox's Regression Model using data from the Troms heart study. Statistics in Medicine. 1985; 44): 413 – 423.

Dinesh R.J. Pai (2009). Determining The Efficacy of Mathematical Programming Approaches for Multi-Group Classification. Pg.15. Rutgers University Electronic Theses and Dissertations.

Efron, B. (1975). The Efficiency of Logistic Regression Compared to Normal Discrimination Function Analysis. Journals of the American Statistical Association, December 1975, Volume 70, Number 352, Theory and Methods Section.

Eke, C.N and Opabisi, A. K (2019): Logistics Regression Analysis on The Effect of Inflation Rate and Exchange Rate Interaction on Positive Consumer Confidence in Nigeria. *1st International Conference of Industrial and Applied Sciences. Vol. 1, 249 – 257.*

Harrel, F.E. (1997): Translating Probability Models into Clinical Decisions. Lecture Notes.

Harrel, F.E. and Lee, K.L (1985): A Comparison of the Discrimination of Discriminant Analysis and Logistic Regression under Multivariate Normality in P.K. Sen (Ed): *Biostatistics: Statistics in Biomedical, Public health and Environmental Sciences.* North-Holland: Elsevier Science Publishers, 333 – 343.

Harrison, R.J., and Rubinfeld, D.L (1978), @Hedonic Prices and Demand of Clean Air, Journals of Environmental Economics and Management, Vol. 5, 81 – 102

Hossain M, Wright s, Petersen L, (2002). A comparing performance of Multinomial Logistic Regression and Discriminant Analysis for Monitoring Access to care for acute Myocardial Infarction. Journal of Clinical Epidemiology, Volume: 55, Issue: 4, Pages: 400 – 406.

Johnson Richard A. Wichern Dean W. (1999). Applied Multivariate Statistical Analysis. 4th Edition. New Jersey, Prentice Hall. Pg. 650 – 683.

Johnson, R.A. and Wichern D.W. (2002): *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Kaka, E, (2003): An application of Discriminant Analysis in University Admission: A case study of the Faculty of Engineering and Faculty of Science, Nnamdi Azikiwe University. Awka. Unpublished B.Sc. Project. Department of Statistics, NAU.

Kiang, M. (2003). A Comparative Assessment of Classification methods, Decision Support Systems. Pg. 35, 441 – 454.

Manel S. (1999), Comparing Discriminant Analysis, Neural Networks and Logistic Regression for Predicting Species distributions: A case study with a Himalayan River bird: Ecological Modelling 120, pg. 337 – 347.

Montgomery M.E, White M.E, Martin S.W (1987). A comparison of discriminant analysis and logistic regression for the prediction of Coliform Mastitis in Dairy Cows. Canadian Journal of Veterinary Research. 1987;51(4):495–498.

Morrison D.F. (1967). *Multivariate Statistical Methods, McGraw-Hill Book Company*, New York.

Ogum G.E.O. (2002). *Introduction to Methods of Multivariate Analysis*, Afri Towers Publisher Ltd, Aba- Nigeria.

Ogum G.E.O. (1976). *An Application of Discriminant Analysis in University Admission: A Case Study of the University of Nsukka Medical School, 1975/76.* Unpublished B.Sc. Project, Department of Statistics,UNN.

Peng Chao-Ying, Lee KukLida , Ingersoll Gary M. (2002) . An Introduction To Logistic Regression Analysis And Reporting . The Journal Of Educational Research: Journal of Educational Research 96(1): 3-13.

Peng, C. Y., Manz, B. D., & Keck, J. (2001). Modeling categorical variables by logistic regression. American Journal of Health Behavior, 25(3), 278–284.

Pohar M., Blas M., and Turk S. (2004). Comparison of logistic regression and linear discriminant analysis: a simulation study, Metodološki Zvezki, vol. 1, no. 1, pp. 143–161, 2004.

Press, S. J. & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association, 73, 699-705. regression in management research. Journal of Management. 22(2) 339-358.

Uwuanyi C.B, (2014): Application of Linear Discriminant Analysis in Evaluation of Loans and Advances. A case study of five Commercial Banks in Nigeria. Unpublished HND Project, Department of Statistics, Federal Polytechnic Nekede, Owerri Imo State. Pg. 21, 44 – 46.

Young B. (1993) .Quadratic Versus Linear Discriminant Analysis: *A paper presented at the annual meeting of the Mid-South Educational Research Association 22nd, New Orleans, LA, Nov. 11-12, 1993.*