Journal of Natural Sciences Research
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.4, No.8, 2014

Fisher's Linear Discriminant Classifier And Rank Transformation Approach To Discriminant Analysis

C.E. Onwukwe, and E.N. Ogbonna

Department of Mathematics/Statistics & Computer Science

University of Calabar, P.M.B. 1115, Cross River State, Nigeria.

E-mail: mailchrisonwukwe@gmail.com +2348037134718

Department of Statistics

Abia State Polytechnic, P.M.B.7166, Aba, Abia State, Nigeria.

Email: Ericogbonna438@yahoo.com +2348036707597

ABSTRACT

Discriminant analysis is a multivariate statistical technique used primarily for obtaining a linear function of p variables which maximizes the distance between centroids or midpoints of multivariate distributions of k groups. Linear discriminant analysis was performed using the fisher's technique which was also derived. Test for differences in the means for the two groups and their variance covariance matrices were discussed. A major shortcoming of the fisher's linear discriminant analysis is that if normality assumption does not hold, the optimal property is lost. This paper compared Fisher's linear discriminant analysis and the rank transformation approach. This was illustrated by performing discriminant analysis on the data and discriminant analysis on the ranks. If the population is not normal, the effectiveness of this method is enhanced by using the ranks of the original data rather than the data themselves. The results obtained indicate that the two methods perform equally well but the rank transformation is a better alternative to the Fisher's discriminant technique for distributions of small samples and non-normal data.

Keywords Fisher's Linear Discriminant Analysis, Rank transformation, Classification, Apparent Error Rate.

INTRODUCTION

Fisher's linear discriminant analysis is a conventional multivariate technique for dimension reduction and classification. Fisher's discriminant analysis is concerned with the problem of classifying an object of unknown origin into one or more distinct groups or population on the basis of observations made on it. (Hawkins, 1982)The goals of discriminant analysis are to construct a set of discriminants that may be used to describe or characterized group separation based upon a reduced set of variables for the purpose of analyzing the contributions of the original variable to the separation and to evaluate the degree of separation. (Neil ,2002)

Usually, when confronted with a set of objects that comes from two or more populations, we may wish to classify a new object into one of the populations with known values of the variables. For example, a university vice chancellor may be interested to know if an applicant for admission for a degree programme is qualified to be admitted or not. Should a prospective student be admitted or not based on the result of an aptitude test? A researcher may want to predict the success or failure of a new product based on data obtained relating to the item. A medical doctor may wish to conclude whether a pregnant woman will deliver through normal or cesarean -section by certain measurements such as height, weight, blood pressure and cervical size before delivery. An automobile engineer may classify an auto mobile engine into grade I, grade II, or grade III on the basis of measurements of its output, shape, size and shape. Nutritionist may classify food items into carbohydrate, protein, minerals, fat and oil based on measurements observed about the food composition. These examples illustrate range of problems that can be solved using of discriminant analysis.

Discriminant analysis was first applied by the originator R.A Fisher, an English scientist who developed how species of birds could be classified (Fisher, 1936). He considered group separation when there are only two classes, k=2. Fisher originated the idea to find a linear combination of the predictors, $z = a_1x_1 + a_2x_2 + ...,+a_px_{p,}$ that shows the largest difference in the group means relative to the within group variance. Other works that considered optimal transformation by minimizing the within and between class separation are Guo etal 2007 and Ye, J (2005).

Fisher's linear discriminant analysis is a linear combination of observed or measured variables that best describe the separations between known groups of observations. Its basic idea is to classify or predict problems where the dependent variable appear in quantitative form (Rencher, 2002)

Classification analysis is a multivariate technique associated with the development of rules for allocating or assigning observations to one or more groups. Classification rules require knowledge of the parametric structure of the groups. The goal of classification is to create rules for assigning observation to groups that minimize the total probability of misclassification. Linear discriminant functions are used to develop classification rules and the goals of the two procedures tend to overlap. Classification is a statistical method used to build predictive models to separate and classify new data points. It is a learning function that classifies a data item into one of several predefined classes (McLachlan,1992).

Johnson and Wichern (2007) tried to find "discriminants "whose numerical values are such that the populations are separated as much as possible. The emphasis is on deriving a rule that can be used to optimally assign new objects to the labeled classes.

This work intends to develop a robust classification function, perform discriminant classification analysis and classify individuals into groups and determine the apparent error rate by applying the Fisher's method and the rank transformation approach. Conover and Iman,(1980), compared rank transformation method with Linear discriminant function (LDF) and Quadratic discriminant function (QDF) using simulation studies. Gessaman and Gessaman (1972), compared the probabilities of misclassification for several types of discrimination methods.

Huberty and Stephen (2006) stated the rank transformation procedure as follows;

let $X_{(l)}$ denote the l^{th} ordered observation on a given variable in the original data set;

let X_{N+1} denote the observation for a new unit, and let $R(X_{(N+1)}$ denote its rank which may be assigned as follows;

1. If $X_{N+1} < X_{(1)}$, then $R(X_{N+1}) = R(X_{(l)} = 1$
2. if $X_{N+1} > X_{(N)}$, then $R(X_{N+1}) = R(X_{(N)} = N$
3. If $X_{N+1} = X_{(l)}$, then $R(X_{N+1}) = R(X_{(l)}$.
4. If $X_{(l)} < X_{N+1} < X_{(l+1)}$, $l = 1,2,.........., N-1$, then

$$R(X_{N+1}) = R(X_{(l)} + \left[R(X_{(l+1)}) - R(X_{(l)}\right]\frac{X_{N+1} - X_{(l)}}{X_{(l+1)} - X_{(l)}}$$

METHODOLOGY

Fisher proposed the transformation of multivariate observations to univariate observations such that they could be maximally separated. The mean difference of these observations determines the separation. Fisher classification rule maximizes the variation between samples variability.

Let the transformation of multivariate variables

$$X^{'} = \left[X_1, X_2, ..., X_p\right] \quad (1)$$

to a univariate variable Y be a linear function of the X variables. Let the data matrix be represented as

$$\underset{n_1xp}{X_1} = \begin{bmatrix} x^{'}_{11} \\ x^{'}_{22} \\ . \\ . \\ x^{'}_{1n_1} \end{bmatrix} \qquad \underset{n_1xp}{X_2} = \begin{bmatrix} x^{'}_{21} \\ x^{'}_{22} \\ . \\ . \\ x^{'}_{2n_2} \end{bmatrix} \qquad \text{data set } = \begin{bmatrix} 1 \\ 1 \\ . \\ 2 \\ 2 \end{bmatrix} \text{ population}$$

Which are transformed to scalars

$$Y_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ . \\ y_{n_1} \end{bmatrix} \qquad Y_2 = \begin{bmatrix} y_{21} \\ y_{22} \\ . \\ y_{n_2} \end{bmatrix}$$

The sample mean vectors and covariance matrices are given as

$$\underset{(px1)}{\bar{x}_1} = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j}, \qquad \underset{(pxp)}{S_1^{\ 2}} = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(x_{1j} - x_1)(x_{1j} - \bar{x}_1)^{'} \quad (2)$$

$$\underset{(px1)}{\bar{x}_2} = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}, \qquad \underset{(pxp)}{S_2^{\ 2}} = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(x_{2j} - x_2)(x_{2j} - \bar{x}_2)^{'} \quad (3)$$

The pooled sample variance covariance matrices which is an unbiased estimate of Σ is

$$S_p^{\ 2} = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_1^{\ 2} + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_2^{\ 2} \quad (4)$$

$$S_p^{\ 2} = \frac{(n_1 - 1)S_1^{\ 2} + (n_2 - 1)S_2^{\ 2}}{n_1 + n_2 - 2} \quad (5)$$

We find the vector 'a' that maximally separate the two groups.

The separation is measured by $\frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^{\ 2}}$ (6)

The linear combination $\hat{y} = a' x = (\vec{x}_1 - \bar{x}_2)' S_p^{\ -1} x$ maximizes the ratio

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{s^2{}_y} = \frac{(\hat{a}'\bar{x}_1 - \hat{a}' x_2)^2}{\hat{a}' S_p \hat{a}'} = \frac{(\hat{a}' d)^2}{\hat{a}' S_p \hat{a}} \quad (7)$$

Over all possible coefficient vectors $\hat{a}'$, where d $= (\bar{x}_1 - \bar{x}_2)$. The maximum of the ratio above is

$$D^2 = (\bar{x}_1 - \bar{x}_2)' S_p^{\ -1} (\bar{x}_1 - \bar{x}_2) \quad (8)$$

which is known as the Mahalanobi's squared distance.

The vector $\hat{a}$ that maximizes the standardize difference and the square distance is the ratio

$$\frac{a' (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)' a'}{a' S_p a} \quad (9)$$

and the maximization occurs when

$$a = (\bar{x}_1 - \bar{x}_2)' S_p^{\ -1} \quad (10)$$

The optimum direction of $\hat{a}' = (x_1 - x_2) S_p^{\ -1}$ is parallel to the line joining $\bar{x}_1 - \bar{x}_2$ because D^2 is equivalent to the standardized distance between $\bar{x}_1$ and $\bar{x}_2$. Hence we can state that

$$D^2 = (\bar{x}_1 - \bar{x}_2)' S_p^{\ -1} (\bar{x}_1 - \bar{x}_2) = a' (\bar{x}_1 - \bar{x}_2), \quad (11)$$

and any other direction other than $a' = (\bar{x}_1 - \bar{x}_2) S_p^{\ -1}$ would yield a smaller difference between $a' x_1$ and $a' x_2$.

Therefore the linear combination and the linear discriminant function is

$$\hat{y} = a'x = (\bar{x}_1 - \bar{x}_2)S_p^{-1}x$$

(12)

Fisher's discriminant function and classification rule

A simple procedure for classifying or allocating new observations x_o to any of the two groups would be

i To calculate $Y = (\bar{x}_1 - \bar{x}_2)' S^{-1}x$ (13)

ii To determine the midpoint 'm' which is given as

$$m = \frac{(\bar{y}_1 + \bar{y}_2)}{2} \quad (14)$$

$$m = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S_p^{-1}(\bar{x}_1 + \bar{x}_2) \quad (15)$$

Iii We proceed and apply the allocation rule based on Fisher's discriminant function.

Allocate x_o to π_1 if $y_o = (\bar{x}_1 - \bar{x}_2)S_p^{-1}x_o \geq \hat{m}$ (16)

Allocate x_o to π_2 if $y_o = (\bar{x}_1 - \bar{x}_2)S_p^{-1}x_o < \hat{m}$

Where $m = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S_p^{-1}(\bar{x}_1 + \bar{x}_2)$ (17)

This rule is known as Fisher's linear discriminant function.

$$m = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S_p^{-1}(\bar{x}_1 + \bar{x}_2) = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) \quad (18)$$

$$\bar{y}_1 = (\bar{x}_1 - \bar{x}_2)' S_p^{-1}\bar{x}_1 = \hat{a}'\bar{x}_1$$

$$\bar{y}_2 = (\bar{x}_1 - \bar{x}_2)' S_p^{-1}\bar{x}_2 = \hat{a}'\bar{x}_2$$

.

$$Y_c = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) \quad (19)$$

The classification rule based on the cut off value is

Assign an observation x_o to π_1 if $Y \geq Y_c$

Assign an observation x_o to π_2 if $Y < Y_c$

This is called the linear discriminant analysis and this technique adhere strictly to equal variance covariance matrix for two normal population (Fisher,1936,).

Rank Transformation.

Rank transformation procedure is one in which the usual parametric procedure is applied to the ranks of the data instead of the data themselves. Here, we simply rank each variable in all the observations from the smallest, with rank 1, to the largest with rank N, separately for each of the p component variables. The discriminant function procedure is then applied to the data vectors of ranks with each of their elements replaced by their corresponding ranks. This procedure is repeated across all predictors and the usual parametric procedures are followed to obtain the discriminant function and apparent error rate. This procedure is not only simple and easy, but the approach is a development to solving new nonparametric problems where parametric procedure exists.

Let X_{ij} be the jth individual observation from π_1, $j = 1,2,......., N; i = 1,2,........, k$.

X_{ijm} denotes the p components of X_{ij} where $m = 1,2,......p$. The mth components of all X_{ij} are ranked from the smallest, with rank 1, to the largest, with rank $N = N_1 + N_2 + ... + N_k$. Each component is ranked separately for $m = 1$ to $m = p$. Then the sample means $R(\bar{x}_i)$ and sample covariance matrices $R(S_i)$ are computed on the ranks of the observations from each population separately. The new observation x_0 to be classified is compared, component by component, with all original observations and each component of x_0 is replaced by a rank, which is a number obtained by linear interpolation between the two adjacent ranks. Average ranks are used when there is a tie.

In our investigation, we considered Fisher's linear discriminant analysis and Rank Transformation Approach using two groups of observations.

Multivariate test of significance.

Test of equality for group mean vectors

Hotellings: T^2 Test Criteria

Hypothesis

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Test statistic

$$T_c^{\,2} = \frac{n_1 n_2}{n_1 + n_2}\left(\bar{X}_1 - \bar{X}_2\right)' S_{pooled}^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

Anderson (1984) (20)

$$T_t^{\,2} = \frac{(n_1 + n_2 - 2)}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}(\alpha) \tag{21}$$

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

where $F_{p, n_1 + n_1 - p - 1}(\alpha)$ is the value of F with degree of freedom of numerator $n_1 + n_2 - p - 1.$

Decision Rule

Reject H_0 at α level of significance if

$$F > F_{p, n_1 + n_2 - p - 1}(\alpha)$$

Wlks' lamda is a direct measure of the proportion of variance in the combination of dependent variables that is unaccounted for by the independent variables. Wilks' (1932) suggested the wilks' Lambda criteria which are given as.

$$\Lambda^* = \frac{|W|}{|B + W|} = \frac{\sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(x_{i\,j} - x_{i.}\right)^2}{\sum n_i (x_i. - \bar{x}..)^2 + \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i.\right)^2} \qquad (22)$$

where

W = 'within' sum of squares and products matrix

B = 'Between' sum of squares and products matrix

B + W = Total sum of squares and products matrix

Critical value

$$F = \left(\frac{\sum n_i - p - 1}{p} \right) \left(\frac{1 - \wedge^*}{\wedge^*} \right)$$

. Reject H_0 if Λ^* > F approximation of Wilks' lamda at α level of significance.

Test of equality of covariance matrices

Box's M test statistic is usually applied to test for equality of covariance matrices for two groups. Box (1949) presented the Box M test statistic which is given as

$$M = (N - g) Log / S / - \sum_{i=1}^{g} v_i Log / S_i / \qquad (23)$$

Where $S = \frac{\sum_{i=1}^{g} v_i S_i}{n - g}$

This approximation is appropriate provided $n_i > 20$ and p and g are less that 6. The chi-square approximation is

$$\chi_c^{\ 2} = (1-c)M$$

Where $$C = \frac{2p^2 + 3P - 1}{6(p+1)(g-1)}\left[\sum_{i=1}^{g}\frac{1}{v_i} - \frac{1}{N-g}\right] \qquad (24)$$

H_o is rejected if $\chi_c^{\ 2} > \chi^2{}_{\alpha(v)}$

where $$v = \frac{p(p+1)(g-1)}{2}$$

Classification rule

In executing the classification rules, let's define $f_1(x)$ and $f_2(x)$ as the probability density functions associated with the random vector Y for populations π_1 and π_2. Let the prior probabilities be p_1 and p_2, with $p_1 + p_2 = 1$. Let $c_1 = C(2|1)$ and $c_2 = C(1|2)$ represent the misclassification of assigning an individual from π_2 to π_1 and from π_1 to π_2 respectively. If $f_1(x)$ and $f_2(x)$ are known, the total probability of misclassification (TPM) is equal to p_1 times the probability of assigning an individual to π_2 given that it is from π_1, $p(2|1)$ plus p_2 times the probability that an individual is classified into π_1 given that it is from π_2, $p(1|2)$. Thus,

TPM = $p_1 P(2|1) + p_2 P(1|2)$ Johnson and Wichern (2002) (25)

The optimal error rate (OER) is the error rate that maximizes the TPM. If cost is taking into account, the average or expected cost of misclassification is given by

(ECM) = $p_1 P(2|1)C(2|1) + p_2 P(1|2)C(1|2)$ (26)

The classification rule is to make the ECM as small as possible although cost of misclassification is not usually known in practice. (Neil, 2002). The classification performance of discriminant procedures is determined by the "error rates" or the misclassification probabilities.

The optimum error rate (OER) is the minimum (smallest value) total probability of misclassification over all classification rules. This is given as

$$OER = \min[p_1 \int_{R_2} f_1(x)dx + p_2 \int_{R_1} f_2(x)dx] \quad (27)$$

The actual error rate (AER) is the sample performance of the total probability of misclassification of the simulated data. This is given as

$$AER = TPM = p_1 \int_{\hat{R}_2} f_1(x)dx + p_2 \int_{\hat{R}_1} f_2(x)dx$$

(28)

Where $\hat{R}_1$ and $\hat{R}_2$ are the classification regions of n_1 and. n_2

The Apparent error rate (APER) was evaluated using the confusion matrix presented below. It indicates the number of correct and incorrect classified individuals in the data set.

Confusion matrix table 1.

Predicted membership

	Actual	π_1	π_2	
Membership	π_1	n_1c	n_2m	n_1
	π_2	n_2m	n_2c	n_2

Where

n_1c is the number of individuals from π_1 correctly classified as π_1

n_2c is the number of individuals from π_2 correctly classified as π_2

n_1m is the number of individuals from π_1 misclassified as π_2

n_2m is the number of individuals from π_2 misclassified as π_1

The apparent error rate (APER) is given as

$$\text{APER} = \frac{n_1m + n_2m}{n_1 + n_2}$$

$$\hat{p}(2|1) = \frac{n_{1m}}{n_1} \qquad \hat{p}(1|2) = \frac{n_{2m}}{n_2} \qquad \hat{p}_1 = \frac{n_1}{n_1 + n_2} \qquad \hat{p}_2 = \frac{n_2}{n_1 + n_2}$$

The confusion matrix and APER is to justify how good or bad the rule is. The APER is an estimate of the probability that a classification procedure based on a given data will misclassify a future observation.

Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.4, No.8, 2014

DATA

Two groups of Car owners and Non- car owners were considered for best sales prospects by an automobile manufacturing company who were interested in classifying families on the basis of Income and lot size. The sample taken were $n_1 = n_2 = 12$ car owners and Non car owners. The figures are given below:

Car Owners		Non Car Owners	
Income (In $)	Lot size (in ft)	Income (in $)	Lot Size (in ft)
1. 90.0	18.4	105.0	19.6
2. 115.5	16.8	82.8	20.8
3. 94.8	21.6	94.8	17.2
4. 91.5	20.8	73.2	20.4
5. 117.0	23.6	114.0	17.6
6. 140.1	19.2	79.2	17.6
7. 138.0	17.6	89.4	16.0
8. 112.8	22.4	96.0	18.4
9. 99.0	20.0	77.4	16.4
10. 123.0	20.8	63.0	18.8
11. 81.0	22.0	81.0	14.0
12. 111.0	20.0	93.0	14.8

Source: Applied Multivariate Statistical Analysis by Johnson and Wichern 2007.

DATA ANALYSIS

Summary of the data: The result of the original data for Fisher's technique is presented in T**able 2 -7** while that of the ranked data for Rank transformation is shown in **Table 8 -14.**

Table 2: Tests of Equality of Group Means

	Wilks' Lambda	F	Df1	Df2	Sig.
X1	.076	267.705	1	22	.000
X2	.072	284.685	1	22	.000

Table 3: Box's Test of Equality of Covariance Matrices

Log Determinants

Group	Rank	Log Determinants
1	2	11.161
2	2	2.830
Pooled within groups	2	9.808

Table 4: Test Results

Box's M	61.872
F Approx.	18.605
Df1	3
Df2	87120.000
Sig.	.000

Table 5: Eigen values

Function	Eigenvalue	% of Variance	Cummulative %	Canonical Correlation
1	27.271	100.0	100.0	.982

Table 6: Wilk's Lambda

Test of Function	Wilk's Lambda	Chi-square	df	Sig.
1	.035	70.178	2	.000

Table 7: Classification Function Coefficients

	1	2
Rx1	.669	.125
Rx2	.922	.185
Constant	-77.616	-3.588

Fisher's Linear Discriminant Function

Table 7: Classification Results

Predicted group Membership

Group		1	2	Total
Original Count .	1	12	0	12
	2	0	12	12
%	1	100.0	0	100.0
	2	0	100.0	100.0
Count	2	12	0	12
Crossvalidated	2	0	12	12
%	1	100.0	0	100.0
	2	0	100.0	100.0

100% of original grouped correctly classified

100% of cross validated grouped correctly classified

Table 8: Tests of Equality of Grouped Means

	Wilk's Lambda	F	Df1	Df2	Sig.
Rx1	.245	77.046	1	25	.000
Rx2	.245	76.933	1	25	.000

Table 9: Box's Test of Equality of Covariance Matrices

Log Determinants

Groups	Rank	Log Determinant
1	2	5.120
2	2	5.255
Pooled within groups	2	5.209

Table 10: Test Results

Box's M	.326
F Approx.	.100
Df1	3
Df2	132508.071
Sig	.960

•

Table 11 Eigenvalues

Function	Eigenvalue	% of variance	Cumulative%	Canonical Correlation
1	5.982	100.0	100.0	.926

Table 12 Wilk's Lambda

Test of Function(s)	Wilk's Lambda	Chi-square	df	Sig.
1	.143	46.641	2	.000

Table 13: Classification function Coefficients

	1	2
Rx1	1.329	.421
Rx2	1.327	.430
Constant	-25.262	-3.277

Fisher's linear discrminant functions

Table 14: Classification Results

Groups	1	2	Total
Original count 1	12	0	12
2	0	15	15
% 1	100.0	0	100.0
Cross validated 1	12	0	12
2	0	12	15
% Count 1	100.0	0	100.0
2	0	100.0	100.0

100.0% of groups cases correctly classified

100.0% of cases validated cases correctly classified

RESULTS AND DISCUSSION

Discriminant analysis was carried out on two procedures-Fisher's linear discriminant function and the rank transformation approach. It was done to predict sales prospect by an automobile manufacturing company who was interested in classifying families on the basis of Car owners and Non-car owners. The predictor variables are income and lot size. **Table 2** and **table 8** showed significant mean differences of ($p<.000$) in both income and lot size which meant that the data is suitable for discrimination. In **table 3**, the log determinants of the Fisher's Discriminant (11.161, 2.830, 9.808) were not similar and the Box's M test result of **table 4** showed that the assumption of equality of covariance matrices was violated at a significance value of (P <.000). However, in **table 9**, the log determinants (5.120,5.225, 5.209) were similar which is an indication of equality of covariance matrices. This was confirmed by the Box's M test results of **table 10** which showed a non-significance value of (p < .960). **Table 5** revealed that there is a significant association between groups and the predictors which explained for 96.4% of between groups variations. **Table 11** also revealed significant association which explained for 85.7% of between group variation. The discriminant function for both approach is highly significant as indicated in **Table 6** and **table 12** which showed (p <.000). The two tables also showed variation within groups which was unexplained to be 3.6% and 14.3% respectively. The cross-validated classification in table 7 and table 13 showed that the hit ratio for both techniques is 100% and their probabilities of misclassification are therefore zero.

CONCLUSION

The results of the experiment conducted suggested a comparable classification procedure to Fisher's linear discriminant function. The data used was employed to establish a preferable alternative approach to the Fisher's linear discriminant function. The Rank transformation approach is compared to the Fisher's linear discriminant function and the Rank procedure gave the same results as the Fisher's technique. The hit ratio and probabilities of misclassification for both methods are the same. However, the Rank transformation approach is more ideal to adopt when the sample size is small and the data is non-normal. In distributions where normality assumptions is violated, the Rank transformation becomes a better alternative. This was established in this study when Box M test for equal covariance matrices was violated as displayed in table 4.

This showed that the rank transformation method is more robust and efficient than the Fisher's technique and therefore is recommended as a better technique for small samples of non-normal data.

Acknowledgments This work was actually motivated by regular seminar presentation of the department of Mathematics/Statistics and Computer Science, University of Calabar. The 2009 Education Trust Fund of the Nigerian government via Abia State Polytechnic,Aba, lecturers who contributed by their candid suggestion and the invaluable silent reviewers for their comments.

REFERENCES

Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis, second edition, John Wiley. and Sons Inc., New York.

Box,G.E.P., (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36: 317- 346

Conover, W.J. and Iman, R.I. (1980). The Rank Transformation as Method of Discrimination with Some Examples". Communication in statistics- Theory and Methods, A9, 465 – 485.

Fisher, R., (1936). The use of multiple measurements in taxonomic problems, Ann. Eugenics 7, 179 – 188.

Gessaman,M.P.& Gessaman,P.H. (1972). A comparison of some Multivariate Discrimination Procedures. Journal of American Statistical Association, 67, 468 – 477.

Guo, Y.,Hastie,T., & Tibshirani, R (2007). Regularized inear discriminant analysis and its application in microarrays. Biostatistics, 8(1) 86-100.

Hardle, W. & Simar, L.(2007). Applied Multivariate Statistical Analysis. Springer, second edition, New York.

Hawkins, D. M.(1982). Topics in Applied Multivariate Analysis. Cambridge university press, first edition new York.

Huberty, C.J.& Stephen, O.(2006). Applied Manova and Discriminant Analysis. 2nd edition. John W

Johnson, R. A. & Wichern, D, W. (2002). Applied Multivariate Statistical analysis (4th edition) Englewood Cliffs, NJ: Prentice Hall.

Johnson R. A. & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis 6th edition, Pearson Hall, upper Saddle River, New jersey 07458.

McLachlan, G. J. (1992). Discriminant Analysis and Statistical Pattern recognition. New York: Wiley-Interscience.

Neil, H.T. (2002). Applied Multivariate Analysis. Springer-verlag New York.

Rencher, A.C.(2002). Methods of Multvariate Analysis. 2nd ed., John Wiley and sons, inc. New York.

Wilk's, S.S. (1932). Certain generalization in the Analysis of Variance. Biometrica, 24, 471-494

Ye, J. (2005). Characterization of a family of algorithyms for generalized discriminant analysis on undersampled sampled problems. Journal of machine learning research, 6, 483-502