Fjrs.futa.edu.ng
FUTA Journal of Research in Sciences
ISSN: 2315 – 8239 (Print); E-ISSN: 2489 - 0413

FUTA Journal of Research in Sciences, Vol. 13 (1) 2017:113-121

APPLICATION OF DISCRIMINANT ANALYSIS IN MODELLING STUDENTS' RESULTS

* A.H. Bello, O.B. Aladeniyi and E. Osafu

Department of Statistics, School of Sciences, Federal University of Technology, Akure.

Corresponding Author's e-mail:bellab_2011@yahoo.com]

ABSTRACT

The research is aimed at applying discriminant analysis in developing a model for the prediction of the final year graduating class of degree and also to determine the influence of hundred and two hundred levels student's results on the final graduating class of degree in the Department of Mathematical Sciences, Federal University of Technology, Akure, Ondo State, Nigeria. Data were collected by taking a random sample of 150 academic records of alumni students and categorized into three groups as historical data in the department of Mathematical Sciences. Fisher's discriminant analysis were used to analyze the collected data that resulted in two eigenvalues ($\lambda_1 = 8.0072,\ \lambda_2 = 0.0198$) and corresponding eigenvectors. The first discriminant function associated with λ_1 coefficients (β) revealed positive contribution for the variables: first semester result, hundred level ($\beta = 0.3438$), two hundred level (β = 0.612), second semester result, hundred level (β = 0.417) and two hundred level (β = 0.578). The second function which is not significant revealed a negative contribution for the variables. This revealed that these variables highly contribute to the final class of graduation. The models obtained were able to classify different groups in the historical data up to 97.33% (hit ratio). The function was a valid tool for classifying students of unknown group.

Key words: Multivariate normality test, Discriminant Analysis, Wilks' Lambda test, Classification matrix, Press Q Statistic.

INTRODUCTION

Discriminant Analysis

Discriminant Analysis is computationally equivalent to regression analysis. It involves deriving linear combinations of two or more independent variables that will discriminate best between prior defined groups, subject to the decision rule of maximizing the between-group variance, relative to within-group variance. The two techniques commonly used in discriminant analysis are descriptive discriminant analysis (DDA) and predictive discriminant analysis (PDA). Stevens (1996) described the difference between descriptive discriminant analysis and predictive discriminant analysis. Descriptive discriminant analysis tends to focus more on the separation among different groups, while Predictive discriminant analysis focuses more on the classification of group members into their respective groups. Different techniques are applied in both analysis (Huberty and Barton (1989). Research has shown that predictive discriminant analysis performs quite well with categorical data (Gilbert, 1968; Moore, 1973) than regression analysis. Discrimination procedures based on normal populations predominate in statistical practice because of their simplicity and reasonably high efficiency across a wide variety of population models (Johnson and Wichem; 1992).

The number of dependent variable groups (categories) can be two or more, but these groups must be mutually exclusive and exhaustive. Hair *et al* (2007) considered a case where three or more categories are created and the possibility that arises when examining only the extreme groups in a two-group discriminant analysis. This procedure is called the polar-extremes approach. This involves comparing only the extreme two groups and excluding the middle group from the discriminant analysis. Independent variables are usually selected in two ways: either from previous research or from intuition – selecting variables for which no previous research or theory exists but that might logically be related to predicting the groups for the dependent variable. Discriminant analysis is quite sensitive to the ratio of sample size to the number of predictor variables. Hair *et al* (2007) also suggested a ratio of 20 observations for each predictor variable, although this will often be unachievable. At a minimum, though the smallest group size must exceed the number of independent variables. Many times the sample is divided into two subsamples, one used for estimation of the discriminant function (the analysis sample) and another for validation purposes (the holdout sample). This method of validating the function is referred to as the split-sample or cross-validation approach. No definite guidelines have been established for dividing the sample into analysis and holdout groups. The most popular procedure is to divide the total group so that one-half of the respondents are placed in the analysis sample and the other half are placed in the holdout sample. Some researchers prefer a 60-40 or a 75-25 split however if the sample size isn't large enough to split in this way (if $n < 100$) then one compromise would be to develop the function on the entire sample and then use the function to classify the same group used to develop the function. This gives an inflated idea of the predictive accuracy of the function. The challenge of designing an education model of any kind in higher education has been of great interest to many researchers over the years. Usoro (2006) carried out a study to classify students into various departments on National diploma, based on their cumulative results obtained from year Foundation programme otherwise known as Pre-National Diploma (Pre-ND) in Polytechnic system. Charles and June (1970) carried out a study to determine if a differentiation or separation among students graduating, withdrawn or failing could be identified. Erimafa *et al* (2009) applied discriminant analysis in predicting the class of degree of students. Julianti *et al* (2012) performed a research to ascertain which course determines students' final graduating result. Adebayo and Jolayemi (1998) applied the T statistics to investigate how predictable final year result would be, using the first year result or grade point average (GPA) of some selected university graduates. Thomas and Paschal (2013) performed a comparison of the performance of students in pre-degree and University matriculation examination (UME) classes in a University system. Gary *et al* (2004) predicted MBA no-shows and graduation success with discriminant analysis, the paper uses discriminate analysis to examine five years of MBA admission records in order to separate no-shows from the successful program graduates. In this study our major task is to identify students who might be at risk of not graduating or graduating with different classes of degree. Concerning this research the groups under study are three namely: Group one, Group two and Group three. The Group one were a group of Students that graduated with either 'First Class' or 'Second Class Upper Division'. The Group two was the Group of Students that graduated with 'Second Class Lower Division' or 'Third Classes. The Group Three' were the Group of Students not graduating or graduated with 'Pass'. Identification of those courses that are Grade Point Average (GPA) booster by student might tend to change the class of degree of a student. This is a course that the understanding of its concept has a booster effect on sectional Grade Point Average (GPA). This student identification task, performed by the discriminant analysis seems more appropriate than commonly used educational measures such as correlation and regression weights because the variable being predicted is categorical. Research has shown that predictive discriminant analysis perform quite well with categorical data. Fisher linear discriminant analysis (otherwise known as discriminant analysis), will discriminate between groups better than any other linear function (Fisher 1936). And also violation of the assumptions underlying regression modeling can attract serious repercussions.

The objective of this research is to use students' Grade Point Average (GPA) (First and Second Year GPA) to design a discriminant analysis model for predicting the final classes of degree of students in Mathematical Science Department, Federal University of Technology, Akure, Ondo State, Nigeria.

METHODOLOGY

The data for this study were from academic records of First year and Second year (independent variables) of alumni students and their corresponding final classes of degree (categorical variable) in the Department of Mathematical Sciences, Federal University of Technology, Akure, Ondo State from 2005 to 2007 academic session .

The historical data were collected for three groups (categories), namely Group 1, Group 2, and Group3. The collected data are grade point average (GPA) of each student (both First and Second Semester first year and second year). The number of students n_i in group i is the sample size for that group.

X_1 Represent First Year Grade Point Average (GPA) for First Semester.

X_2 Represent First Year Grade Point Average (GPA) for Second Semester.

X_3 Represent Second Year Grade Point Average (GPA) for First semester.

X_4 Represent hundred level Grade Point Average (GPA) for First semester.

$\bar{X}_1, \bar{X}_2 and\, \bar{X}_3$ Represent the mean of Group 1, Group 2 and Group 3 respectively and $\bar{X}$ be the grand mean of the entire Group.

$\bar{X}_{ij}$ Represent the mean of Group i and Variable j

Total samples taken from each group is $n_i = 50$

Fisher discriminant analysis

Fisher discriminant analysis make use of the between (**B**) and within (**W**) sum of squares in accomplish its aim.

$$\boldsymbol{B} = \sum_{i=1}^{g} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Also an estimate of $\sum$ is based on the sample within group matrix

$$\boldsymbol{W} = ((n_i - 1)S_i) = \sum_{i=1}^{g} = \sum_{i=1}^{g}\sum_{j=i}^{n_i}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \dots\dots (2)$$

Let $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p\, be$ an eigenvalue (characteristic root) of $\boldsymbol{W^{-1}B}$ *that is*

$$|\boldsymbol{W^{-1}B} - \boldsymbol{\lambda I}| = 0 \dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Where $\boldsymbol{I}$ **is** identity matrix? Equation (3) is called the characteristic equation.

Let $V_1, V_2, V_3, \dots V_p$ *be the normalize eigen vector*(characteristic vector) that satisfies $\boldsymbol{W^{-1}BV} = \boldsymbol{\lambda V}$.

The first discriminant function is obtained by forming a linear combination of $V_1'X$ that is $Z_1 = X_1V_{11} + X_2V_{12} + X_3V_{13} + \cdots + X_pV_{1p}$

The second variate/function is also formed by $V_2'X$ which is equal to

$$Z_2 = X_1V_{21} + X_2V_{22} + X_3V_{23} + \cdots + X_pV_{2p}$$

$$Z_p = X_1V_{p1} + X_2V_{p2} + X_3V_{p3} + \cdots + X_pV_{pp}$$

Where P is the number of independent variables measured in each group/population.

The number of positive eigenvalue is $\mathbf{Minimum}(g-1, p)$, where **g** is the number of groups and **p** is the number of independent variables.

DATA ANALYSIS

Test for normality

Using the Mahalanobis test of multivariate normality, first is to obtain the Mahalanobis distance (MD) for each students.

Total sample for group1, group2 and group 3 are equal. That is $n_1 = n_2 = n_3 = 50$

$$\bar{X}_i = \begin{bmatrix} \bar{X}_{i1} \\ \bar{X}_{i2} \\ \bar{X}_{i3} \\ \bar{X}_{i4} \end{bmatrix}$$

$$\bar{X}_1 = \begin{bmatrix} 3.6754 \\ 3.6670 \\ 3.7352 \\ 3.7408 \end{bmatrix} \bar{X}_2 = \begin{bmatrix} 2.7832 \\ 2.7124 \\ 2.8518 \\ 2.7254 \end{bmatrix} \bar{X}_3 = \begin{bmatrix} 1.7300 \\ 1.8068 \\ 1.7472 \\ 1.7590 \end{bmatrix}$$

$$\bar{X} = \begin{bmatrix} 2.7295 \\ 2.7287 \\ 2.7781 \\ 2.7417 \end{bmatrix}$$

$$S_1 = \begin{pmatrix} 0.27658 & 0.082994 & 0.085833 & 0.075453 \\ 0.082994 & 0.236103 & 0.056136 & 0.106423 \\ 0.085833 & 0.056136 & 0.241838 & 0.076349 \\ 0.075453 & 0.106423 & 0.076349 & 0.209452 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 0.204451 & 0.043898 & 0.037212 & 0.027346 \\ 0.043898 & 0.209786 & 0.026353 & -0.00228 \\ 0.037212 & 0.026353 & 0.171909 & -0.03607 \\ 0.027346 & -0.00228 & -0.03607 & 0.203776 \end{pmatrix}$$

$$S_3 = \begin{pmatrix} 0.17758 & 0.063655 & 0.021459 & 0.032084 \\ 0.063655 & 0.150953 & 0.036652 & 0.006213 \\ 0.021459 & 0.036652 & 0.172339 & 0.002785 \\ 0.032084 & 0.006213 & 0.002785 & 0.212572 \end{pmatrix}$$

Where S1, S2 and S3 are variance covariance matrix of group1, group2 and group3 respectively

$$MD_{ij} = (X_{ij} - \bar{X}_i)' S_i^{-1} (X_{ij} - \bar{X}_i), j = 1, 2, \ldots, n \text{ and } i = 1, 2, 3$$

Where X_{ij} is the jth element (student) on the ith group.

Since each of the data are univariate normally distributed and none of the students' squared Mahalanobis distance is $>$ the $\chi^2_{4\left(1-\frac{0.05}{2}\right)} = 11.1, \alpha = 0.05$ we can say that data set is multivariate normally distributed.

Sum of squares

The total sum of square = the between sum of square + the within sum of square

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\bar{X}_{ij} - \bar{X})(\bar{X}_{ij} - \bar{X})' = \sum_{i=1}^{3} n_i(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' + \sum_{i=1}^{g}\sum_{i=1}^{n_i}(\bar{X}_{ij} - \bar{X}_i)(\bar{X}_{ij} - \bar{X}_i)$$

$$\boldsymbol{T} = \boldsymbol{B} + \boldsymbol{W}$$

$$\boldsymbol{B} = \sum_{i=1}^{3} n_i(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'$$

$$\bar{X}_1 - \bar{X} = \begin{pmatrix} 0.9459 \\ 0.9383 \\ 0.9571 \\ 0.9991 \end{pmatrix}, \bar{X}_2 - \bar{X} = \begin{pmatrix} 0.0537 \\ -0.0163 \\ 0.0737 \\ -0.0163 \end{pmatrix}, \bar{X}_3 - \bar{X} = \begin{pmatrix} -0.9995 \\ -0.9219 \\ -1.0309 \\ -0.9827 \end{pmatrix}$$

$$\boldsymbol{B} = \left[50 * \begin{pmatrix} 0.9459 \\ 0.9383 \\ 0.9571 \\ 0.9991 \end{pmatrix} (0.9459 \quad 0.9383 \quad 0.9571 \quad 0.9991) \right]$$

$$+ \left[50 * \begin{pmatrix} 0.0537 \\ -0.0163 \\ 0.0737 \\ -0.0163 \end{pmatrix} (0.0537 \quad -0.0163 \quad 0.0737 \quad -0.0163) \right]$$

$$+$$

$$\left[50 * \begin{pmatrix} -0.9995 \\ -0.9219 \\ -1.0309 \\ -0.9827 \end{pmatrix} \begin{pmatrix} -0.9995 & -0.9219 & -1.0309 & -0.9827 \end{pmatrix}\right]$$

$$\boldsymbol{B} = \begin{pmatrix} 94.8305375 & 90.4050855 & 96.9831565 & 96.3191015 \\ 90.4050855 & 86.5286095 & 92.3616165 & 92.1836175 \\ 96.9831565 & 92.3616165 & 99.2113455 & 98.4051365 \\ 96.3191015 & 92.1836175 & 98.4051365 & 98.2082895 \end{pmatrix}$$

$$W = \sum_{l=1}^{g}\sum_{i=1}^{n_l} (\bar{X}_{lj} - \bar{X}_l)(\bar{X}_{lj} - \bar{X}_l)' = (n_1 - 1)S_1 + (n_2 - 1)S_2 + (n_3 - 1)S_3$$

$$\boldsymbol{W} = (n_1 - 1)S_1 + (n_2 - 1)S_2 + (n_3 - 1)S_3$$

$$\boldsymbol{W} = \begin{pmatrix} 32.271939 & 9.336803 & 7.080696 & 6.609267 \\ 9.336803 & 29.245258 & 5.837909 & 5.407444 \\ 7.080696 & 5.837909 & 28.718214 & 2.110136 \\ 6.609267 & 5.407444 & 2.110136 & 30.6642 \end{pmatrix}$$

Multivariate analysis of variance shows that there was significant difference among the three groups. This difference will improve the classification ability of discriminant analysis.

Discriminant analysis

The MANOVA test shows that the groups were significantly different from each other. Therefore discriminant analysis could be used to separate the groups. Let $^{W-1}$ be the inverse of within sums of square

$$\boldsymbol{W}^{-1}\boldsymbol{B} = \begin{pmatrix} 1.4111490878799 & 1.3407196224688 & 1.4444593725611 & 1.4284977086265 \\ 1.702747047829 & 1.6355320137235 & 1.7379787602421 & 1.7423294940517 \\ 2.5092928074449 & 2.3866974428439 & 2.5677871902303 & 2.5429146512623 \\ 2.3639939084197 & 2.2645997473697 & 2.414604516878 & 2.4125697659033 \end{pmatrix}$$

$$let\ \boldsymbol{R} = \boldsymbol{W}^{-1}\boldsymbol{B}$$

The number of eigenvalue/vector to be obtained is Min (p, g-1) = Min (4,2) = 2.$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq |\lambda_4|$.

The largest eigenvalue is $\leq Max(\, r_i = (\sum_{j=1}^{4} r_{ij})) = 11.43.\, where\ r_i\, is\ the\ ith\ row\ of\ R$

Using the **Scaled Power Method** to obtain the eigenvector and the corresponding eigenvalue of **R**. we have below:

The corresponding eigenvector and eigenvalue are:

$$\boldsymbol{V_1} = \begin{pmatrix} 0.3438098252159 \\ 0.4167459831371 \\ 0.6116408499452 \\ 0.5779386302216 \end{pmatrix}, \lambda_1 = 8.0072778724361$$

$$\boldsymbol{V_2} = \begin{pmatrix} -0.5322616122519 \\ 0.2132822184365 \\ -0.8192654936787 \\ 0.0035103118415 \end{pmatrix}, \lambda_2 = 0.0197601743569$$

Therefore the discriminant functions $Z_1\, and\ Z_2$ are given as:

$Z_1 = \boldsymbol{V}_1'\boldsymbol{X}$

$$[0.3438098252159 \quad 0.4167459831371 \quad 0.6116408499452 \quad 0.5779386302216]\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$\boldsymbol{Z_1} = 0.3438098252159X_1 + 0.4167459831371X_2 + 0.6116408499452X_3 + 0.5779386302216X_4$$

$Z_2 = \boldsymbol{V}_2'\boldsymbol{X}$

$$[0.5322616122519 \quad -0.2132822184365 \quad 0.8192654936787 \quad 0.5921080114859]\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$\boldsymbol{Z_2} = -0.5322616122519X_1 + 0.2132822184365X_2 - 0.8192654936787X_3 + 0.0035103118415X_4$$

Test of significance

Testing the significance of the model is important to the interpretation of the experiment. The hypothesis is to be tested on both models, to ascertain the predictive ability of each model.

Wilks' Lambda test

Using the Wilks' lambda to test for the significance of the models.

$$\boldsymbol{W} = \frac{\boldsymbol{1}}{\boldsymbol{1+\lambda}}$$

The test statistic is:

$Q = -\left[(n-1) - \frac{p+g}{2}\right] InW$ Follows χ^2 (chi square) distribution with $p(g-1)$ degrees of freedom.

For the first model, $\lambda_1 = 9.6568$

Null hypothesis $H_o: \lambda_1 = 0$

Alternative hypothesis $H_1: \lambda_1 \neq 0$

$$W_1 = \frac{1}{1+\lambda_1} = \frac{1}{1+8.0072778724361} = 0.111021333$$

$$Q_1 = -\left[(150-1) - \frac{4+3}{2}\right] In(0.111021333) = -145.5(In(0.111021333)) = 319.8137874$$

$$\boldsymbol{\chi^2}_{4(3-1)} \boldsymbol{at\ \alpha = 0.05 = 15.51}$$

Since Q_1 is greater than $\chi^2{}_8$, we have the statistical reason to reject H_0 *and* conclude that the model prediction is more than chance. This model alone can be used for prediction of future occurrence or classification. But the second model is not significant; this means that the second model alone cannot predict future classes of degree of students.

Functions at group centroid

This is the mean of the discriminant scores of both functions in a group.

With the centroid classification can be done for all students whose results served as the historical data for this research.

Table 1: Group Centroid

FUNCTIONS

GROUP	$\bar{Z}_1$	$\bar{Z}_2$
1	$\bar{G}_{11} = 7.2384$	$\bar{G}_{12} = -4.22116$
2	$\bar{G}_{21} = 5.406665$	$\bar{G}_{22} = -3.2297$
3	$\bar{G}_{31} = 3.573967$	$\bar{G}_{32} = -3.575192$

$\bar{G}_{ij}$ *is the mean of ith group jth discriminant scores.*

We classify by first determining the difference between each discriminant scores and the group centroids. A student is predicted to be in the ith group if the distance between its discriminant

score and the ith group centroid is the minimum among other groups.
A student having a discriminant score Z, is predicted to be in group k if:

$$G_k = [|Z - \bar{G}_{k1}|], \quad for \quad k = 1,2,3$$

Where k is the group and G_k is the minimum. Therefore student is predicted to belong to group k. On applying this principle called the **distance method**, we were able to classify our historical data.

Table 2: Classification matrix

			Predicted Group Member			Total
			1	**2**	**3**	
		1	**48**	2	0	50
Original Group	Count	**2**	1	**49**	0	50
		3	0	1	**49**	50
		1	**96**	4	0	100
	%	**2**	2	**98**	0	100
		3	0	2	**98**	100

The number of students correctly classified in the historical data is the diagonal elements on the classification matrix above.
Therefore the **Hit** ratio $H = \frac{48+49+49}{150} = 0.973333333333$
The percentage correctly classified is $H\% = 97.33\ percent$

Press Q statistics

This simple measure compares the number of correct classifications with the total sample size and the number of groups. The calculated value is compared with a critical value from the Chi-Square distribution with 1 degree of freedom.

$$Press's\ Q\ statistic\ (P) = \frac{(N - nk)^2}{N(k-1)}$$

H_o: The result exceed classification by accuracy by chance
Vs.
H_1: The result does not exceed classification by chance at α = 0.05 level of significance.

$N = (50 + 50 + 50) = 150, n = (48 + 49 + 49) = 146, k = 3$

$$P = \left(\frac{[150 - (146 * 3)]^2}{150 * (3-1)}\right) = \frac{82944}{300} = 276.48$$

Since P is greater than $\chi^2_{\alpha,1} = 3.84$ we have the statistical reason to reject H_0 and conclude that the classification result is more than just chance.

RESULTS AND DISCUSSION

The Mahalanobis distance showed that the data does not contain any outlier and it was multivariate normally distributed. Multivariate Analysis of Variance showed that there was a significant difference between each group at 5% level of significance and that there was homogeneity within each group and each group was sufficiently different from one another.
Two eigenvalue $(\lambda_1 = 8.0072, \lambda_2 = 0.0197\ )$ and corresponding eigenvector were obtained using the scaled power method. The vector was used to form two Fisher's discriminating functions;

$$Z_1 = 0.343809825215 9X_1 + 0.4167459831371X_2 + 0.611640849945 2X_3 + 0.5779386302216X_4$$

$$Z_2 = -0.5322616122519X_1 + 0.2132822184365X_2 - 0.819265493678 7X_3 + 0.0035103118415X_4$$

Wilk's lambda tests were used to test the significance of these models, which rendered $\boldsymbol{Z_2}$ function not significant. The significant function $\boldsymbol{Z_1}$was used to classify the historical data that gave a hit ratio of 97.33%.
The models obtained above have a predictive ability of about 97.3% at 95% confidence level.
The result of this analysis showed that, with the first year and second year grade point averages (GPA) of a student from the Department of Mathematical Sciences can be used in determining graduating class of degree of that student. The model also inform the student the distance of his/her discriminant score from a desire class of that student, which inform the student the effort needed to either remain in a particular class or to belong in other classes of degree. This model can

be used by the student advisers to counsel students of this department. This study also showed how statistical models, with every assumptions satisfied can be used for predictions with a minimal risk. The result obtained also confirms the conclusion made by Erimafa et al (2009) about the high predictive ability of discriminant analysis when all assumptions are satisfied.

CONCLUSION

This study examined the influence of the 100 and 200 level results on the final class of degree of student. Results of discriminant analysis based on the criteria of values of unstandardized canonical coefficient (eigenvectors), correlation matrix, eigenvalues, values of Wilk's Lambda and classification routine identified that these variables highly contribute up to $\left(\frac{8.007}{8.027}\right) * 100 = 99.7\%$ to the final class of degree. Each of the coefficient of the discriminant function contributed positively to the discriminating of each group corresponding to the highest eigenvalue. The two hundred (both first and second semester) contributed more on the final class of degree. The model was able to predict up to 97.33% of the entire historical group.

RECOMMENDATION

This study recommended that discriminating model should be developed for every department throughout the University. A model developed for a particular department varies from models from other departments, therefore cannot be too accurate for prediction in other departments. Variations in the performance of student might be greatly affected by the variables (courses) involved. Educational research showed that some psychological factors, such as learning style, self-efficacy, motivation and interest, and teaching and learning environment, also play a role in students' learning and thus affect students' achievement. Therefore, future studies should include psychological variables in the models so as to increase their prediction accuracy.

REFERENCE

Adebayo, S.B and Jolayemi, E.T (1998). *Effect of Rare Outcome on the Measure of Agreement Index. J. Nig. Stat. Assoc. 13:1-10.*

Charles, B.K, June E. H (1970). *Predicting Graduation Withdrawal and Failure in College by Multiple Discriminant Analysis. J. Educational Measure. 7(2): 91-95.*

Gary, E. C and Tom, C (2004).*Predicting MBA No-Shows and Graduation Success using Discriminant Analysis. IAER: August 2004, Vol. 10, No. 3*

Erimafa, J. T, Iduseri, A. and Edokpa, I. W(2009).*Application of Discriminant Analysis to Predict the Class of Degree for Graduating Students in a University System: inter journal of physical Sciences, vol.4,Pp. 16-21,2008.*

Gilbert, E. S (1968). *On Discrimination using Qualitative Variable. J. Amer. Stat. Soc. pp. 399 – 404.*

Huberty, C. J., and Barton, R. M (1989). *An Introduction to Discriminant Analysis Measurement and Evaluation in Counseling and Development, 22: 158 – 168.*

Johnson, R. A., and Wichern, D. W (2007). *Applied Multivariate Statistical Analysis, Sixth Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey. 07458.*

Julianti, Kasih, Sani and Susanto (2012). *Predicting Students' Final Results through Discriminant Analysis, World Transactions on Engineering and Technology Education (WIETE) Vol.10, No.2, 2012.*

Moore, D. H (1973). *Evaluation of five*

Discrimination procedures for Binary Variables.
J. Amer. Stat. Soc. 13: 399 – 404.

Stevens, J (1996). *Applied Multivariate Statistics for the Social Sciences.In: Mahwah NJ, (3rd Ed). Lawrence Erlbaum Associate, Inc.*

Thomas, A. U and Paschal, B. N (*2013****).*** *A Comparison of the Performance of Students in Pre-Degree and University Matriculation Examination (UME) Classes in a University System: A Discriminant Analysis Approach, International Journal of Advancements in Research and Technology, vol. 2,September.*

Usoro, A. E (2006). *Discriminant Analysis and its Application to theClassification of Students On the Basis of their Academic Performances. Journal of Research in Physical Science. 2(3): 53 – 55.*