On Proper Classification and Placement of Students in Nigerian University Systems Using Discriminant Analysis

G. O. Nwafor[1,*], C. E Onwukwe[2]

[1]Department of Mathematics and Statistics, Cross River University of Technology Calabar, Nigeria
[2]Department of Mathematics, Statistics and Computer Science, University of Calabar, Cross River State, Nigeria
*Corresponding author: gonwafor@gmail.com

Abstract This paper attempts to provide more realistic and reliable way of placing Nigerian students seeking admission into Nigerian University system using Discriminant Analysis and also providing quantitative analysis of a Discriminant Analysis approach to prediction of student's admission scores into a university system. The paper utilizes secondary data sourced from the Admission and Public Relation Department of Cross Rivers University of Technology Calabar, the conditions for predictive and classification discriminant analysis were obtained, and the empirical result yield a discriminant linear classification function results on the various faculties of interest on the method of admission system obtainable in Nigeria university system. The study reveals university mandatory examination (UME) and Aptitude Test score of students in their various faculties. The linear function established a hit ratio of 83% of which successfully predicted the student admission scores. The study had apparent error rate of 17% which explains the probabilities of misclassification.

Cite This Article: G. O. Nwafor, and C. E Onwukwe, "On Proper Classification and Placement of Students in Nigerian University Systems Using Discriminant Analysis." *American Journal of Applied Mathematics and Statistics*, vol. 2, no. 6 (2014): 394-397. doi: 10.12691/ajams-2-6-6.

1. Introduction

Discriminant Analysis is used in situation where the clusters are known in advance. The aim of discriminate analysis is to classify an observation, or several observations, into these known groups. The admission system adopted in major university based on the education intervention of the kind in higher education has been of great and area for many researcher or educations, over the years. [1] carried out a study on classification of students into various Departments on the basis of their cumulative results for one year foundation programme otherwise known as pre-national diploma (PRE-ND) in polytechnics system. [2] carried out a study in application of discriminant analysis to predict the classes of degree for graduating students in a university system. The analyses yield a linear discriminant function which successfully classified or predicted the graduating student's class of degrees. Discriminant analysis builds a predictive model for group membership the model is composed of a discriminant function (or, for more than two groups, a set of linear classification function) based on linear combination of the predictive variables that proved the best discriminant among groups. [3] carried out a study to determine of a differentiation or separation among students graduating, withdrawing or failing could be

identified. [4] applied the t-statistic [5] to investigate how predictable the final years result would be using the first year result or grade point average (GPA) of some selected university graduates. In the past 35 years, research in academic prediction has centered on graduation, withdrawal, failure and selection of the students on the basis of either their collegiate success or cumulative results of remedial or pre- nd, and literature to date suggests no loss of interest.

In the study, our major task is to empirically yield a discriminant linear classification function on the various faculties of interest on the method of admission system obtainable in Nigeria University, the UME and aptitude test score. The first four groups are students from the faculty of science, education, engineering and environmental science with their UME and Aptitude test scores, respectively. This prediction of student's admission scores into a university system, formed by the discriminant analysis seems more appropriate than commonly used educational measure such as regression, weights etc. Research has show that predictive discriminant analysis perform quite well categorical date [6] also violations of the assumption underlining regression modeling can have serious repercussions [7].

2. Materials and Methodology

The data type and sources of this study was secondary data from the Admission and Public Relation unit of Cross River University of Technology Calabar. Other augmenting sources of this study include published articles and journals, working papers, textbooks and relevant internet resources. In the first state of data collection, this system of scores was established, the UME SCORE AND APPTITUDE TEST SCORES from the admission unit which includes various department and their respective faculties, but the interest of the researcher was on the faculty level comprising the entire departments. A simple random sampling technique was employed to obtain the desired faculties of four (4) the faculty of science, Engineering, Education and Environmental science. A systematic sampling technique was quantitatively applied to obtain the desired sample size of 150 per faculty in the two systems of admission scores.

2.1. Data Analysis

The method is to set up a procedure based on the students/applicants admission score of the two systems in their respective faculties by the discriminate analysis function and placement to the program by score by the classification function for the groups /faculties. The fisher linear discriminant function [8] will be adopted in the study since it will discriminate between the groups better than any other linear function [9].

The analysis yields variance-covariance matrices for the groups are given as

$$S_{ij} = \sum y_i y_j - \frac{\sum y_i \sum y_j}{n} \qquad (1)$$

$$S_{ii} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \qquad (2)$$

$$S_{jj} = \sum y_j^2 - \frac{(\sum y_j)^2}{n} \qquad (3)$$

$$S_1 = \begin{bmatrix} 267.156 & 22.766 \\ 22.766 & 132.131 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 351.270 & -6.096 \\ -6.096 & 185.542 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 338.330 & 13.887 \\ 13.887 & 94.483 \end{bmatrix}$$

$$S_4 = \begin{bmatrix} 202.888 & -6.064 \\ -6.064 & 116.660 \end{bmatrix}$$

And the pooled covariance matrices given as

$$s_{pooled} = \frac{\begin{bmatrix} (n_1 - 1)S_1 + (n_2 - 1)S_2 \\ + (n_3 - 1)S_3 + (n_4 - 1)S_4 \end{bmatrix}}{n_1 + n_2 + n_3 + n_4 - 4} \qquad (4)$$

$$s_{pooled} = \begin{bmatrix} 289.911 & 6.123 \\ 6.123 & 132.204 \end{bmatrix}$$

$$s_{pooled}^{-1} = \begin{bmatrix} 0.003453 & -0.00016 \\ -0.00016 & 0.007571 \end{bmatrix}$$

The Table 1 indicates the descriptive properties of the study groups.

Table 1. Descriptive Statistics

GROUP		Mean	Std. Deviation
Science	UME	203.23	16.345
	APPT	56.17	11.495
Education	UME	201.54	18.742
	APPT	51.04	13.621
Engineering	UME	202.21	18.394
	APPT	63.03	9.720
Environmental	UME	198.28	14.244
	APPT	61.23	10.801
Total	UME	201.32	17.085
	APPT	57.87	12.386

2.1.1. TEST OF SIGNIFICANT

The hypothesis and test statistics is:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_0 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

With the F-transformation of Hotelling's T^2 given

$$F = \frac{n_1 + n_2 + n_3 + n_4 - P - 1}{(n_1 + n_2 + n_3 + n_4 - 4)P} * T^2 \qquad (5)$$

Were $T^2 = \frac{n_1 n_2 n_3 n_4}{n_1 + n_2 + n_3 + n_4}(D^2 = mahalanobisdistance.)$

H_0 is rejected if $F_{Cal} > F_{p, n_1 + n_2 + n_3 + n_4 - P - 1; 1 - \propto}$

This indicates that at 5% level of significant; reject the hypothesis of equality of group means. This implies that there exist significant differences among the group means.

2.1.2. DISCRIMINANT LINEAR CLASSIFICATION FUNCTION

Obtain a linear classification function, which is denoted by

$$l_i(y) = \overline{y}_i^1 s_p^{-1} - \frac{1}{2} y_i^1 s_p^{-1} \overline{y}_i \forall i = 1, 2, \ldots, k \qquad (6)$$

Is maximum (largest).Assigning y to a group using the linear classification function and prior probabilities i.e., the distance function which $p_i f(y/G_i)$ is maximum with this rule, the probability of misclassification is minimized and the rule becomes

$$L_i^1(y) = \ln P_i + \overline{y}_i^1 s_i^1 y - \frac{1}{2} y_i^1 s_p^{-1} \overline{y}_i \forall i = 1, 2, \ldots, k \qquad (7)$$

[10] Certain facts exist; if $P_1 = P_2 = \ldots = P_k$ then equation (2) which optimizes the classification rate for normal distribution reduces to equation (1) which was based on the heuristic approach of minimizing the distance of y to $\overline{y}_i$.

APER –The apparent error rate is easily obtained and routinely provided by most classification software programs. It is an estimate of the probability that our classification functions based on the sample will misclassify a future observation.

$$APER = \frac{n_{14} + n_{23} + n_{32} + n_{41}}{n_{1} + n_{2} + n_{3} + n_{4}} \qquad (8)$$

Similarly,

Apparent correction classification rate

$$APCR = 1 - \left(\frac{n_{14} + n_{23} + n_{32} + n_{41}}{n_{1+} + n_{2+} + n_{3+} + n_4}\right). \tag{9}$$

The error and correction classification rate judge the ability of classification procedures to predict group member. The usually use the probability of misclassification which is known as the Aper and its complement Apcr.

3. Result

The predictive analytical software (PASW) was employed on the data which yields the following findings and results.

Table 2. Linear Classification Function Coefficients

Classification Function Coefficients				
	GROUP			
	Science	Education	Engineering	Environmental
UME	.693	.688	.688	.675
APPT	.393	.354	.445	.432
(Constant)	-82.807	-79.725	-84.978	-81.508
Fisher's linear discriminant functions				

This is the distribution of observation into the gender groups used as a starting point in the analysis. The default prior distribution is the allocation into groups, as seen in the analysis. SPSS allows users to specify different prior with the priors subcommand. The discriminant analysis and classification builds a predictive model for group membership and classification of linear function .The linear classification function coefficients provide the following model for placement of students in their respective programs.

$Science(score) = 0.693UME + 0.393APPT - 82.807$

$Education(score) = 0.688UME + 0.354APPT - 79.725$

$Engineering(score) = 0.688UME + 0.445APPT - 84.978$

$Environmental(score) = 0.675UME + 0.432APPT - 81.508$

Based on the linear combination of the prediction variables that provide the best classification /placement among group of students. The functions are generated from a sample of cases for which group membership is known, the functions can then be applied to new cases that have measurements for the predictor variable but have unknown group membership. This study provides a classification model for placement of student in Cross River University of Technology Calabar.

Table 3. Confusion matrix for actual and predicted groups

	GROUP	Predicted Group Membership				Total
		1	2	3	4	
Original Count	Science	38	55	31	26	150
	Education	12	82	25	31	150
	Engineering	21	31	50	48	150
	Environmental	20	37	42	51	150
%	Science	25.3	36.7	20.7	17.3	100
	Education	8.0	54.7	16.7	20.7	100
	Engineering	14.0	20.7	33.3	32.0	100
	Environmental	13.3	24.7	28.0	34.0	100

Probability of misclassification 0.17. Hit ratio = 1-0.17= 83%

The study on the discriminant analysis to predict student scores was achieved. The apparent correction rates i.e. the hit ratio yield 83% for the study and probability of misclassification (17%). The Table 3 the confusion matrices which is the measure of predictive ability indicates that discriminant analysis can be used to predict students scores. The linear classification function for the faculties was used to obtain the students aggregate scores for their respective faculties. This quantitative measure of discriminant analysis should be adopted by the study groups to obtain the actual scores for the respective candidates, and having standard cutoff scores by the faculty/university management. The study also reveals the descriptive statistics for the groups of the sampled data. The Table 2 shows the linear discriminant classification function for the study.

4. Discussion

The concern for the predictive ability of the linear discriminant function has obscured and even confused the fact that two sets of techniques based on the purposed of analysis exist , i.e., predictive discriminant analysis (PDA) and descriptive discriminant analysis (DDA).The distinction between PDA and DDA is that predictive discriminant analysis focus on classifying subjects into one of several groups (or to predicate group membership), whereas in descriptive discriminant analysis focus on revealing major differences among the groups. [11] aptly stated the purposes of the two analyses are different; the techniques in the two analyses are different. There is, perhaps, some feasibility of the "mixing of DDA for purpose of corroboration of results. Generally research questions are of the descriptive type or of the predictive type; only seldom would both types of questions be addressed in a given research situation. The two types of discriminant analyses i.e., PDA and DDA have different histories of development. According to [12] "discriminant analysis for the first three or four decades focused on the prediction of group membership", PDA, whereas DDA usage did not appear until the 1960s and "its use has been very limited to apply research settings over the past two decades". Hence, PDA is appropriate when the researcher is interested in assigning units (individuals) to groups based on composites scores on several predictor variables. The accuracy of such prediction can be assessed by examining "hit rates" as against chance; for example, the most basic question answered by PDA is "given the individual scores on several predictor variables, which group represents their true membership group?" Again, the focus of PDA is predication and the accuracy of hit rates. As [11] noted with respect to PDA, "one is basically interested in determining a classification rule and assessing its accuracy".

References

[1] Usoro A. E "Discriminant analysis and its application to the classification students on the basis of their academic performances" *J. Res. Phy. Sci.* 2(3):53-55. 2006.

[2] Erimafa J.T, Iduseri A. and Edokpa I. W "Application of discriminant analysis to predict the class of degree for graduating students in a university" *International journal of physical science* vol.4 (1), pp.016-021. 2009.

[3] Charles B K and June E H "predicting graduation withdrawal and failure in college by multiple discriminant analysis" *J. Edu. Measure* 7(2):91-95.1970.

[4] Adebayo SB and Jolayemi ET "Effect of rare outcome on the measure of agreement index" *j. Nig. Stat. assoc.* 13:1:10. 1990.

[5] Adebayo SB and Jolayemi E T "on the effect of rare outcome on some agreements/Concordance indices" Nig. J pure and appl. Sci. 13:718-723. 1998.

[6] Gilbert E S "on discriminant using qualitative variable" *J. Amr Stat Soc.* Pp 399-4d4. 1968.

[7] Cook RD and Weisbery S "Residuals and influence in regression:" London; chapman and hall.1982.

[8] Fisher R A "the use of multiple measurements in taxonomic problems Ann. Eug" 7:179-188. 1936.

[9] Anderson T W "an introduction to multivariate statistical analysis" New York john wiley. 1958.

[10] Alvin C. Rencher "Methods of multivariate Analysis." John Wiley & Son. Inc. Publication. 2002.

[11] Huberty C J, Barton R M "an introduction to discriminant analysis Measurement and evaluations in counseling and development" 22: 158-168. 1989.

[12] Hubert C J "Applied discriminant analysis" New York Wiley and sons. 1994.