International Journal of Applied Science and Mathematics
Volume 3, Issue 5, ISSN (Online): 2394-2894

An Application of Discriminant Analysis to the Classification of Students on the Basis of their Academic Performances

Joseph O. Onu and Onyedikachi O. John

***Abstract* – In this paper, a discriminant analysis was carried out and the rule for discrimination was obtained. The analysis yielded a linear discriminant model used as a function with which some students were re-classified based on their academic performances. The rule discriminated 4 students from Statistics department to Computer Science department, and 5 students from Computer Science department to Statistics department based on their scores in four departmental courses they did together.**

***Keywords* – Classification, Confusion Matrix, Discimination, Linear Discriminant Function.**

I. INTRODUCTION

Discriminant analysis 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 several categories or population groups on the basis of these measurements, Onyeagu [7]. In routine banking or commercial finance, an officer or analyst may wish to classify loan applicants as low or high credit risks on the basis of the elements of certain accounting statements, Morrison [6]. [Lanchenbruch [5] viewed the problem of discriminant analysis as that of assigning an unknown observation to a group with a low error rate. The function or functions used for the assignment may be identical to those used in the multivariate analysis of variance. Johnson and Wichern [4] defined discriminant analysis and classification as multivariate techniques concerned with separating distinct sets of objects (or observations), and with allocating new objects (observations) to previously defined groups.

Usoro [9] viewed discriminant analysis as a multivariate statistical method which seeks to differentiate between two groups with respect to certain characteristics. According to him, the objective of discrimanant analysis is to have a critical study in naturally or unnaturally occurring groups, and to distinguish between two or more predefined groups. He further stated that in most cases observations or individuals under careful examination may possess some properties which restrict their fundamental membership to two predefined groups of populations of which one group will have standard and similar properties amongst its members distinct from the other. In spite of the common similarities that exist in group membership, there may be some differentials based on certain characteristics which could make some members of a particular group or population different from other members.

Jerome [3] observed that the formal purpose of classification or discriminant analysis is to assign objects to one of several, k, groups or classes based on a set of measurements $X = (x_1, x_2, ..., x_p)$ obtained from each object or observation. He further observed classification techniques are also used informally to study the separability of labeled groups of observations in the measurement space. He further stated that in the formal setting, an object is assumed to be a member of one and only one class, and an error is incurred if it is assigned to a different one.

Onyeagu [7] gave an example of discriminant analysis problem as follows: prospecting students applying for admission to college are given a battery of tests; the vector of scores is a set of measurements X. The prospective student may be a member of one population group consisting of those students who will successfully complete college training or rather have potentialities for successfully completing training, or he may be a member of the other group; those who will not complete the college course successfully. The problem is to classify student for admission on the basis of his scores on the entrance examination.

Usoro [9] applied discriminant analysis in classifying students on the basis of their academic performance. He used the cumulative results of PRE-ND students of Accountancy and Business Administration department based on the five courses they offered for 2004/2005 academic session. Based on their scores, 78 students were discriminated from Business Administration to Accountancy, and 37 students from Accountancy to Business Administration.

Erimafa *et al* [2] applied discriminant analysis to identify students who might be “At risk” (AR) and “Not At Risk” (NAR). The first group, are students who are in danger of graduating with a poor class of degree, and the second group are those that will graduate with better class of degree within their first two years of study. His analysis successfully classified or predicted 87.5 percent of the graduating students’ class of degree.

Thomas and Paschal [8] applied discriminant analysis to compare the performance of students who gained admission into the university system through pre-degree programme and those who passed through the University Matriculation Examination, (UME). It was observed that there is no difference in the performance of UME and pre-degree students on the average at 5% level of significance.

The objective of this research work is to apply discriminant analysis in reclassifying the National Diploma students of Abia State polytechnic, Aba, either from Statistics department to computer science department, or from computer science department to statistics department based on their academic

performances in the departmental courses they did together in their first year.

II. Methodology

A. Data Source

The data for this work are the results of National Diploma I students in Statistics and Computer Science departments, Abia State Polytechnic, Aba. The results involve four courses namely STA 111 (Descriptive Statistics), STA 112 (Elementary Probability Theory), COM 101 (Introduction to Computer) and COM 123 (Computer Packages 1) which were offered by the concerned students in 2014/2015 academic session.

B. Method of Analysis

We made use of mean vector from both departments and their variance-covariance matrices. The description of vectors as used in the analysis is as follows: X_1 represents the scores made by students in STA 111, X_2 represents the scores made by students in STA 112, X_3 represents the scores made by students in COM 101, and X_4 represents the scores made by students in COM 123.

The sum and difference between the mean vectors are given by

$$\bar{X}_s - \bar{X}_c = \begin{bmatrix} \bar{X}_{1s} - \bar{X}_{1c} \\ \bar{X}_{2s} - \bar{X}_{2c} \\ \bar{X}_{3s} - \bar{X}_{3c} \\ \bar{X}_{4s} - \bar{X}_{4c} \end{bmatrix} \text{ and } \bar{X}_s + \bar{X}_c = \begin{bmatrix} \bar{X}_{1s} + \bar{X}_{1c} \\ \bar{X}_{2s} + \bar{X}_{2c} \\ \bar{X}_{3s} + \bar{X}_{3c} \\ \bar{X}_{4s} + \bar{X}_{4c} \end{bmatrix} \quad (1)$$

The variance-covariance matrix for a particular department can be represented as follows:

$$S = \begin{bmatrix} var(X_1) & cov(X_1X_2) & cov(X_1X_3) & cov(X_1X_4) \\ & var(X_2) & cov(X_2X_3) & cov(X_2X_4) \\ & & var(X_3) & cov(X_3X_4) \\ & & & var(X_4) \end{bmatrix} \quad (2)$$

The variance-covariance matrices of the vectors for Statistics and Computer Science departments are denoted by S_s and S_c respectively, and their pooled covariance matrix becomes

$$S_{pooled} = \frac{S_s + S_c}{n_s + n_c - 2} \quad (3)$$

where n_s and n_c are the number of sample results selected in each of the Statistics and Computer Science departments. Here, $S_s = S_c = 60$; which refers to the first 60 students that registered from each department.

The linear discriminant function

$$W = X' S_{pooled}^{-1}(\bar{X}_s - \bar{X}_c) - \frac{1}{2}(\bar{X}_s + \bar{X}_c)' S_{pooled}^{-1}(\bar{X}_s - \bar{X}_c) \quad (4)$$

called Wald-Anderson statistics, Wald [10], Anderson [1], is applied with the following classification rule:

Assign a student to statistics if $W \geq 0$

Assign a student to computer if $W < 0$

III. Results and Discussion

The column vectors for the sum and difference are as follows:

$$\bar{X}_s - \bar{X}_c = \begin{bmatrix} -11.07 \\ 10.97 \\ 9.30 \\ -11.89 \end{bmatrix} \text{ and } \bar{X}_s + \bar{X}_c = \begin{bmatrix} 111.53 \\ 106.43 \\ 115.00 \\ 127.75 \end{bmatrix}$$

The pooled covariance matrix and its inverse are

$$S = \begin{bmatrix} 77.6392 & 5.6186 & 55.8647 & 7.7848 \\ & 66.8595 & 13.4481 & 16.5256 \\ & & 166.3669 & 27.6933 \\ & & & 136.6098 \end{bmatrix}$$

$$S_{pooled}^{-1} = \begin{bmatrix} 0.016996 & -0.00033521 & -0.0057185 & 0.00023127 \\ & 0.015577 & -0.00086528 & -0.0016898 \\ & & 0.0082062 & -0.001233 \\ & & & 0.007761 \end{bmatrix}$$

The linear discriminant model is given by

$$W = -0.2478X_1 + 0.1866X_2 + 0.1448X_3 - 0.1248X_4 + 3.3553 \quad (5)$$

The analysis shows that 56 students were correctly classified and retained in Statistics department, while 4 students were discriminated from Computer Science department to Statistics department as seen in table I; confusion matrix. Furthermore, 55 students were correctly classified and retained in Computer Science department, while 5 students were misclassified and then discriminated from Statistics department to Computer Science department based on their performances.

Table I: Confusion matrix

Department	Statistics	Computer Science
Statistics	56	4
Computer Science	5	55

IV. Conclusion

It is a well known fact that every unit of a population possesses some characteristics similar to or different from other members of the group in which its membership is established. Hence the application of discriminant analysis in this research work is motivated by the need to re-classify some individuals who by one reason or the other belonged to the wrong group(s). The need for discrimination arises if there are some individuals in a group whose characteristics are significantly different from the general characteristics of members forming the greater proportion of number in the group. In this work, the discriminant analysis carried out has provided an understanding of discrimination process involving two departments in a polytechnic as students' scores from four departmental courses offered together were used to discover their actual performances and then re-classified some either from Statistics department to Computer Science department or vice versa based on their academic performances in the selected courses.

References

[1] T. W. Anderson, "Classification by Multivariate Analysis," Psychometrika, vol. 16, 1951, pp. 31-50.

[2] J. T. Erimafa, A. Idusari, and I. W. Edokpa, "Application of Discriminant Analysis to Predict the Class of Degree of Graduating Students in the University System," International Journal of Physical Science, vol. 4, 2008, pp. 16-21.

[3] H. F. Jerome, "Regularized Discriminant Analysis," Journal of the American Statistical Association, vol. 84, no. 405, Theory and Methods, 1989.
[4] R. A. Johnson and D. W. Wichern, "Applied Multivariate Statistical Analysis 3rd edition," Englewood Cliffs, New Jersey: Prentice-Hall, 1992, pp 493.
[5] P. A. Lachenbruch, "Discriminant Analysis," New York: Hafner Press, 1975.
[6] D. F. Morrison, "Multivariate Statistical Methods 3rd edition," New York, McGraw-Hill Inc. 1990, pp. 269.
[7] S. I. Onyeagu, "A First Course in Multivariate Analysis," Mega Concept, 2003.
[8] A. U. Thomas and B. N. Pascal, "A Comparison of the Performance of Students in Pre-degree and University Matriculation Examination Classes in a University System: A Discriminant Analysis Approach," International Journal of Adv. in Research and Technology, vol. 2, Issue 9, 2013.
[9] A. E. Usoro, "Discriminant Analysis and its Application to the Classification of Students on the Basis of their Academic Performances," Journal of Research in Physical Sciences, vol. 2, no. 3, 2006.
[10] A. Wald, "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups," Annals of Mathematical Statistics, vol. 15, 1944, pp. 145-162

AUTHORS' PROFILES

Joseph Omocho Onu, (with email id, josephonu35@yahoo.com) was born in Aba, Abia State , Nigeria, on 28th March, 1978. He obtained his Bachelor of Science degree in Statistics from Michael Okpara University of Agriculture, Umudike, 2010, and Master of Science degree in Statistics from Michal Okpara University of Agriculture, Umudike, Nigeria, 2016. He is currently doing his Ph.D in Statistics Michael Okpara University of Agriculture, Nigeria. He is a LECTURER with Abia State Polytechnic, Aba, Nigeria, in the department of statistics.
He is a member of the Nigerian Statistical Association.

Onyedikachi Ogba John, (with email id, johnkady@yahoo.com) was born in Aba, Nigeria, on 2nd March, 1972. He has a Bachelor of Science degree in mathematics/statistics from University of Calabar, Nigeria, 1997, and a Master of Science degree in statistics from University of Port Harcourt, Nigeria, 2009. He is a Ph.D (Statistics) candidate in Michael Okpara University of Agriculture, Umudike, Nigeria.
He is currently a LECTURER in Rhema University Aba, Nigeria, and author of the journals, "Quantile Regression as a Robust Alternative to Ordinary Least Squares," Port Harcourt, Nigeria, Scientia Africana, vol. 8 (No. 2), December, 2009 pp 61-65, "Robustness of Quantile Regression to Outliers,"American Journal of Applied Mathematics and Statistics, vol. 3, no. 2, 2015, "Stock Market Price and its Determinants: A Case Study of Nigerian Banks," Journal of the International Association of Advanced Technology and Science, vol. 1, Issue 12, 2015, pp. 25-31.
Mr. John is a member of Nigerian Statistical Association.