

Modelling Academic Risks of Students in a Polytechnic System With the Use of Discriminant Analysis

Fagoyinbo, I. S.^{[a],*}, Akinbo, R. Y.^[a], Ajibode, I. A.^[a]

^[a]Department of Mathematics and Statistics Federal Polytechnic, Ilaro, Ogun State, Nigeria

* Corresponding author

Address: Department of Mathematics and Statistics Federal Polytechnic, Ilaro, Ogun State, Nigeria. E-mail: ilesanmiajibode@yahoo.com

Received: May 1, 2013 / Accepted: October 12, 2013/ Published: October 31, 2013

Abstract: This research work “Modelling Academic risks of students in a Polytechnic System With the Use of Discriminant Analysis”: A Case Study of Federal Polytechnic Ilaro, Ogun State, identified students at academic risks i.e. those who are in danger of failing, repeating on probation or being withdrawn due to the level of their academic performance. Several methods exist for student’s identification for academic risks; these include the Bayesian approach, Von Mises (Minimax), Multiple Regression Analysis, etc. For this research work, the method adopted was the discriminant analysis which assist in classifying students into classes of grades i.e. Distinction, upper credit, Lower Credit, Pass and others who are in the risk group, the method was adopted due to its simplicity and its systemic classification of the phenomenon under study.

Key words: Discriminant; Classification; Risk; Regression; Measurement; Discriminant function

Fagoyinbo, I. S., Akinbo, R. Y., & Ajibode, I. A. (2013). Modelling Academic Risks of Students in a Polytechnic System with the Use of Discriminant Analysis. *Progress in Applied Mathematics*, 6(2), 59-68. Available from <http://www.cscanada.net/index.php/pam/article/view/j.pam.1925252820130602.1738> DOI:10.3968/j.pam.1925252820130602.1738

1. INTRODUCTION

Discriminant Analysis is a statistical technique which allows the researcher to study the difference between two or more groups of objects with respect to several variables simultaneously. In designing an institutional intervention of any kind, one needs to accomplish at least three tasks:

- I. Determination of the factors that are relevant to the successive performance of the task at hand.
- II. Evaluate the impact of any new programme on student's performance.
- III. Implementation of the new programme and comparing it with the old programme.

In social sciences, there are wide variety of situations in which this technique can be useful, for example, a researcher team that has been commissioned to study the outcomes of terrorist take-over involving hostages. In particular:

- (a) They want to know what elements of the situation would predict the safe release of hostages even though the terrorist demands have not been met.
- (b) How these variables might be combined into a mathematical equation to predict the most likely outcome and accuracy of the derived equation.

The first task mentioned above in the first paragraph means a careful delineation of the problem. Given the large number of possible variables, it is not surprising that individuals have different ideas based on different assumptions about what causes subpar performance or performance enhancement laying these different ideas "out on the table" may provide the group as a whole an opportunity to attack the problem in the best way possible. This will show how discriminant analysis can be used to help determine what variables have a relationship with performance, and how such relationship can be used to help shape intervention. The second major task is identifying the appropriate individuals with which to use the intervention may typically mean identifying students who might be termed "at risk". These are students who are in danger of failing a class, not understanding a certain concept etc. However, some intervention might be targeted at different levels of performance e.g. "the gifted student" might be chosen for some supplemented instruction. It is in this student identification task that discriminant analysis will be seen to be most advantageous traditional approaches.

Finally, it is always critical to evaluate the intervention. This step has many purposes. The most obvious is that it lets the research team and finding agent know whether or not a given intervention worked. In addition, evaluation can be used to shape the research program and guide it towards a more effective intervention. We will attempt to show how one can use discriminant analysis to assist in this difficult task.

The main purpose of discriminant analysis is to identify the appropriate individuals with which to use the intervention, may typically mean identifying

students who might be academic risks i.e. students who are in danger of failing a class, not understanding a certain concept etc. Therefore the study identifies students at academic risk (AR) and not at academic risk (NAR). The first group are the students who are in danger of graduating with poor class of degrees; PCD (i.e. Pass and Fail) and the second group are those that will graduate with better class of degrees B,C,D; (i.e. Distinction, Upper Credit and Lower Credit) within the two years of study; whether at Ordinary National Diploma (OND) or at the Higher National Diploma (HND). In this study, discriminant function analysis would be used to predict their grades on successive completion of their courses based on their grade point average (GPA) and undergraduate grade points.

2. LITERATURE REVIEW

Researchers have used discriminant analysis in a wide variety of settings: It was first developed by Fisher (1930), who was seeking to solve problems in physical anthropology and Biology. In the social sciences, some of the first application dealt with psychological and educational testing. Political scientists have found discriminant analysis to be useful in studying citizen and court cases and it has also being used in educational interventions. The technique is especially useful, however, in analysis experimental data when assignment to a “treatment” group is presumed to affect scores on several criterion variables.

Ronald Fisher developed discriminant analysis for use with categorical data. It is based on assumptions very similar in nature to multiple regressions, except that it is designed for categorical criterion. While not specifically intended for use with categorical predictors. Discriminant analysis forms linear combination of the predictors which are used to classify cases into the various “group” of the criterion one may conceptualize discriminant analysis in terms of evaluating the centroid of a group of cases. In the present context the student cases are placed in grade. The mean value of a discriminating variable (e.g. SAT or a preceding course grade), or predictors, for the students in a particular group is evaluated. The bigger the difference between the mean values of the predictors related to various groups, the more discriminating is that variable. Discriminant analysis simultaneously analyses all of these mean differences and determines which predictors are most discriminating (based on backward probabilities).

Also, Huberty and Barton (1989) aptly stated that the purpose of the two analysis are different, the two types of discriminant analysis that is; predictive discriminant analysis performs quite well with categorical data.

In this study, discriminant analysis is used to identify student at risk in the department of Mathematics/Statistics of the Federal Polytechnic, Ilaro. This paper presents a discussion of the collinearity problem in regression and discriminant analysis. It described the reason for the prediction ability and classification ability of the classical methods. The discussion is based on the formular for prediction error; special emphasis is put on differences and similarities between regression and classification.

Multivariate regression and discriminant analysis are among the most used and useful technique in modern applied statistics. These methods are used in a number of different areas and application ranging from chemical spectroscopy to machine and social sciences. One of the main problems when applying some of the classical techniques is the collinearity among the variables used in the model. Such collinearity problem can be sometimes lead to a serious problem when the methods are applied, Weisberg (1985). A number of different methods can be used for diagnosing multi-collinearity. This range from simple method based on principal component to more specialized techniques for regularization. The most frequently used method for collinearity and regression classifications resemble each other strongly and are based on similar principles. Often the collinearity problem is described in terms of instability of the small eigen values and the effect that this may have in the empirical inverse. Covariance matrix which is involved both in regression and classification. This information is relevant for the regression coefficient and relevant classification criteria.

Linear discriminant analysis is perhaps the most widely used method for classification because of its simplicity and optimal properties. In the classical discriminant problem as proposed by Fisher (1936), Anderson and Bahadur (1962) studied procedures for classifying two multivariate normal distributions with unequal covariance matrices. They showed how to construct a discriminant function that minimizes probability of misclassification given the other and how to obtain a minimax discriminant procedure is non-linear. The best linear discriminant for this unequal covariance matrix content was found by Clunies-Ross and Riffenberg (1960) and later in (1962) by Anderson , Bohadur and Chenoff (1972, 1973) suggested same measures that indicate how well one can discriminate between two multivariate normal populations with unequal covariance matrices using linear discriminant functions. He then proceeded using such criteria to compare the performance of linear discriminant functions based on balanced and unbalanced designs.

Linear discriminant analysis is known to be optimal for two multivariate normal groups with equal covariance matrices. An important result on non-parametric estimation of linear classification was suggested by Greer (1979, 1984); he considered the algorithm designed to produce a hyper-plane that minimizes classification rules in a completely non-parametric manner for a large set of loss function. The Fisher's linear discriminant analysis problem, it minimizes the expected loss in the case of known prior probability and it is an admissible rule when prior probabilities are not known. Although Fisher's linear discriminant function has been used in so many practical situations, its statistical properties under non optimal conditions have not received much attention until recently.

3. METHODOLOGY

The data for this study was obtained from students' record of Department of Mathematics and Statistics, Federal Polytechnic, Ilaro. There were 61 students

in all. The method adopted in the analysis is the discriminant function. It is a method of finding linear combination of variables which best separates two or more classes. Discriminant analysis is not a classification algorithm although it makes use of class labels. However, discriminant analysis result is mostly used as a part of a linear classifier. The other alternate used in making a dimension reduction before using nonlinear classification algorithms. Discriminant analysis can be used in the same circumstances as multiple regressions. Given a list of potential predictors, one can determine which are most effective in predicting performances. It provides a discriminant function which includes only those variables that should be used in predicting performances. Probably the biggest advantage of discriminant function over regression is that it measures the predictive ability in terms of correct classification. This is possible since the unit of analysis is categorical. It predicts category membership; given the true grouping of criterion. One can determine how many predictions produced by the equations are right.

Moreover, discriminant function analysis is a multivariate analysis of variance, the independent variables are the groups and the dependent variables are the predictors.

The linear discriminant analysis function is given by:

$$Z_D = U_1X_1 + U_2X_2 \quad ..1$$

$$\text{Where, } U = W^{-1}d \quad ..2$$

$$\text{And } W = (N_1 - 1)S_1^2 + (N_2 - 1)S_2^2 \quad ..3$$

$$d = \begin{bmatrix} \bar{X}_1 G_1 - \bar{X}_2 G_2 \\ \bar{X}_2 G_1 - \bar{X}_1 G_2 \end{bmatrix} \quad ..4$$

The basis of the analysis is built on the following foundation, suppose that our population consists of two groups' i.e. π_1 and π_2 . We observe a $p \times 1$ vector X and must assign the individual whose measurement are given by X to π_1 or π_2 . We need a rule to assign X to π_1 or π_2 . If the parameter of these distributions of X in π_1 or π_2 are known. We may use this knowledge in the construction of an assignment rule. If not, we use samples of size N_1 from π_1 and N_2 from π_2 to estimate the parameters. We need a criterion of goodness of classification Fisher (1936) suggested using a linear combination of observation and using the coefficient so that the ratio of the difference of the means of the linear combination in the two groups to its variance is maximized in the two groups to its variance is maximized. In the Fisher's approach, let the linear combination be denoted by:

$$Y = \lambda'X \quad ..5$$

The mean of Y is denoted by $\lambda'\mu_1$ in π_1 and $\lambda'\mu_2$ in π_2 and its variance is given by $\lambda'\epsilon\lambda$ in both populations if we assume that the covariance matrices $\epsilon_1 = \epsilon_2 =$

ε , we wish to choose N to maximize:

$$Q = \frac{(\lambda' \mu_1 - \lambda' \mu_2)^2}{\lambda' \varepsilon \lambda} \quad ..6$$

Differentiating Q with respect to λ' , we have:

$$\frac{dQ}{d\lambda} = \frac{2(\mu_1 - \mu_2)(\lambda' \mu_1 - \lambda' \mu_2)\lambda' \varepsilon \lambda - 2\varepsilon \lambda (\lambda' \mu_1 - \lambda' \mu_2)^2}{(\lambda' \varepsilon \lambda)^2} \quad ..7$$

Equating $\frac{dQ}{d\lambda}$ to zero, it simplifies to:

$$\mu_1 - \mu_2 = \frac{\varepsilon \lambda (\lambda' \mu_1 - \lambda' \mu_2)}{\lambda' \varepsilon \lambda} \quad ..8$$

Since λ is used to separate the two population then it may be multiplied by any constant. Thus λ proportional to $\varepsilon^{-1}(\mu_1 - \mu_2)$ i.e. $\lambda \mu \varepsilon^{-1}(\mu_1 - \mu_2)$. If the parameters are known, then the linear discriminant function is given by $Y = (\mu_1 - \mu_2)\varepsilon^{-1}X$, but if the parameters are unknown, then the usual practice is to estimate them by:

$$Y = (\mu_1 - \mu_2)S^{-1}X \quad ..9$$

Generally, p-measurement made on sample of sizes N_1 and N_2 from population π_1 and π_2 . The mean vector of measurement is given by $X_1 = (x_1, x_2, \dots, x_p)$ is $1 \times p$ matrix. The mean vector of measurement in π_2 is given by $X_2 = (x_1, x_2, \dots, x_p)$. The covariance between the measurement in π_1 is given by $p \times p$ matrix.

$$S_1 = \frac{1}{N_1 - 1} \sum_{i=1}^n (X_{i1} - X_1)(X_{i1} - X_1) \quad ..10$$

The covariance between the measurement in π_2 is given by $p \times p$ matrix.

$$S_2 = \frac{1}{N_2 - 1} \sum_{i=1}^n (X_{i2} - X_2)(X_{i2} - X_2) \quad ..11$$

In general, the covariance between measurement in π_i is given by:

$$S_i = \frac{1}{N_i-1} \sum_{j=1}^n (X_{ji} - X_i)(X_{ji} - X_i) \text{ where } j= 1,2,3, \dots, n; i = 1, 2. \quad ..12$$

The pooled variance is given by p x p matrix:

$$S^2 = \frac{(N_1-1)S_1^2 + (N_2-1)S_2^2}{N_1+N_2-2} \quad ..13$$

The inverse of $S_{p \times p}$ is given by $S_{p \times p}^{-1}$.

The cross product term of the measurement in Group 1 is given below:

$$\sum_{i=1}^{51} X_{1i} X_{2i} = 6646.64$$

The covariance matrix for Group 1 is given by:

$$S_1 = \begin{bmatrix} 0.2651 & -28.3272 \\ -28.3272 & 175.3977 \end{bmatrix}$$

The cross product of measurement in Group II is given below:

$$\sum_{i=1}^{51} X_{1i} X_{2i} = 3868.19$$

The covariance matrix for the measurement in Group II is given by:

$$S_2 = \begin{bmatrix} 0.0703 & 0.4200 \\ 0.4200 & 77.4571 \end{bmatrix}$$

The linear discriminant function is given by:

$$Z_D = U_1X_1 + U_2X_2 \text{ from equation (1)}$$

The pooled sum of squares and cross product matrix is computed as:

$$Z_D = U_1X_1 + U_2X_2$$

$$U = W^{-1}d$$

$$W = (N_1 - 1) S_1 + (N_2 - 1) S_2$$

$$W = \begin{bmatrix} 14.2392 & -1410.48 \\ -1410.48 & 9854.2844 \end{bmatrix}$$

$$W^{-1} = \begin{bmatrix} -0.00533 & -0.00076 \\ -0.00076 & -0.000008 \end{bmatrix}$$

$$d = \begin{bmatrix} -0.5472 \\ -20.4471 \end{bmatrix}$$

$$U = \begin{bmatrix} -0.0185 \\ -0.0006 \end{bmatrix}$$

$$Z_D = -0.0185X_1 - 0.0006X_2$$

Classification rule:

We define the cut-off as: $C = \frac{(\overline{Z_{1D}} + \overline{Z_{2D}})}{2}$

We compute $\overline{Z_{1D}}$ and $\overline{Z_{2D}}$ which denote the function at group centroids, substituting the CGPA and STA 222/STA 423 for each of the two groups.

The discriminant scores of the students in Group I are obtained as follows:

$$Z_D = -0.0185X_1 - 0.0006X_2$$

$$Z_{D1} = -0.0185(2.46) - 0.0006(54) = -0.0779$$

$$Z_{D2} = -0.0185(2.47) - 0.0006(66) = -0.0853$$

This is computed for all the members of the group up to the last member as:

$$Z_{D51} = -0.0185(2.48) - 0.0006(40) = -0.0699$$

The discriminant scores of the students in Group II are calculated as:

$$Z_D = -0.0185X_1 - 0.0006X_2$$

$$Z_{D1} = -0.0185(3.29) - 0.0006(65) = -0.0999$$

$$Z_{D2} = -0.0185(3.32) - 0.0006(80) = -0.1094$$

The last member of the group is $Z_{D51} = 0.0699$

$$\overline{Z_{1D}} = -0.08580196, \text{ similarly } \overline{Z_{2D}} = -1.08093333$$

The cut off mark is defined as: $C = \frac{(\overline{Z_{1D}} + \overline{Z_{2D}})}{2}$

$$C = -0.0969$$

The discriminant procedure is as follows:

Assign an individual to Group 1 if and only if $C > -0.0969$, also to second Group if and only if $C \leq -0.0969$.

Table 1: Probability of misclassification into the two Groups

| Group | Predicted at Risk | Predicted not at Risk | Total |
|-----------------|-------------------|-----------------------|-------|
| Original | 1 39 | 12 | 51 |
| | 2 1 | 14 | 15 |
| % | 1 76.5 | 23.5 | 100 |
| | 2 6.67 | 93.3 | 100 |

The total probability of misclassification is denoted by:

$$P = \frac{q_1}{N_1} + \frac{q_2}{N_2} \quad \dots 14$$

q_1 represents number of students wrongly classified into Group I and q_2 represents number of students wrongly classified into Group II. N_1 and N_2 is the sample size for Groups I and II respectively.

Overall percentage of correct classifications is 80.30%

Overall percentage of correct classifications is 19.70%

4. CONCLUSION

At the end of the data analysis, it was discovered that 39 out of 51 students at risk were predicted at risk. Also, 14 out of 15 not at risk were predicted not at risk. This gives overall correct classification as 80.3%.

5. RECOMMENDATIONS

Discriminant analysis should be the preferred method of operation in educational interventions regardless of the other benefits provided. It is more effective because of the weights added to variables under consideration.

Moreover, Discriminant analysis can serve as a better basis for comparison than regression analysis for situations where control groups are not feasible.

6. REFERENCES

- [1] Anderson, T. W. (1958). *An introduction to multivariate analysis*, John Wiley and Sons Inc.
- [2] Anderson, T.W. & Bahadur, R.R. (1962). Classification into two multivariate normal distributions with different covariance matrices. *Annals of Mathematical Statistics*. V. 33, N. 2, pp. 420 – 431.
- [3] Chernoff, H. (1972). The selection of effective attribute for deciding between hypothesis using linear Discriminant function in frontiers of Pattern Recognition, In: *Frontiers of Pattern Recognition* (pp. 55 – 60). New York Academic Press: Ed. S. Watanabe.
- [4] Clunnies-Ross, C.W. & Riffenberg, R.H. (1960). Geometry and linear discrimination. *Biometrics*, V. 47, pp. 185 – 189.
- [5] Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, V. 7, pp. 179 – 188.
- [6] Greer, R. L. (1979). *Consistent Non-parametric estimation of best linear classification rule/solving in consistent systems of linear inequalities*. (Technical report No. 129). Stanford: Department of Statistics, Stanford University.
- [7] Greer, R. L. (1984). *Trees and Hills: Methodology for maximizing functions of systems of Linear Relations*. Amsterdam: North Holland.

- [8] Huberty, C.J. & Barton, R.M. (1989). An Introduction to Discriminant and Evaluation in Counselling and Development. Vol. 22, pp. 158 – 168.
- [9] Weisberg, S. (1985). Applied Linear Regression. J. Wiley and Sons, NY.