

Efficiency of Discriminant Analysis in Identifying Performance of Students at Risk in Nigerian Private Universities.

Abayomi A. Akomolafe^{1*} and G.N. Amahia²

¹Petroleum of Mathematical Sciences, Ondo State University of Science and Technology, PMB 353, Okitipupa, Ondo State, Nigeria.

²Department of Statistics, University of Ibadan, Nigeria.

E-mail: akomolafe01@yahoo.com*

ABSTRACT

Discriminant analysis is a statistical technique designed to investigate the differences between two or more groups of people with respect to several underlying variables. This technique is more appropriate than commonly used educational measures because the variable being predicted is categorical. Moreover, this approach result is more useful in evaluating educational program.

We have used discriminant analysis to predict students' performance in Nigerian Universities in their science-based mathematical courses in three southwest private universities in Nigeria. In these courses there is a high failure rate (greater than 30%, earn a grade of 'F') which results in a great cost to the universities and to students as success in the course is a prerequisite for all science Students. A technique that could identify the factors that are predictive of course performance and identify students who are at risk would be of great benefits to the design, implementation, and evaluation of educational program.

We will present a case study with over 1,600 science students where discriminant analysis was used successfully to determine the predictor of the student performance. We looked at over 15 possible predictor variables (NECO, WAEC, etc.) and determine which variable would be most useful in identifying students at risk. Using information available from students' records, we were able to successfully, predict 50% of students who eventually failed in the class.

We will also discuss how discriminant analysis can be used as an evaluation technique. We show that this approach provides results that are both more interpretable and statistically sound than traditional measures.

(Keywords: discriminant analysis, student performance, predictor variable, traditional measure, mathematics education)

INTRODUCTION

In designing educational program of any kind, one needs to accomplish at least three tasks: (1) determine the factors that are relevant to successful performance of the task at hand, (2) identify those individuals most likely to benefit from the program, and (3) evaluate the impact of any new program on students' performance. This paper discusses the use of discriminant analysis as a useful technique in addressing these critical issues.

The first issue mentioned above means a thorough diagnosis of the existing problem. Given a large number of possible variables, it is not surprising that individuals have different ideas based on different assumptions about what causes sub-par performance or performance enhancement. This paper will show how discriminant function can be used to help determine what variables have a relationship with performance, and how such relationship can be used to help shape interventions.

The second major issue, identifying appropriate student with which to use the intervention, may typically means identifying students who might be termed "at risk". Those are the students who are in danger of failing a class, not understanding a certain concept, etc. However, some programs might be targeted at different levels of performance (for example, the "gifted students" might be chosen for some supplemental instruction).

It is in this student-identification task that discriminant analysis will be seen to be most

advantageous over traditional approaches. We will show how we used the analytical technique to pinpoint students “at risk” in a class where previous attempt as such classification had been less ideal.

Finally, it is always critical to evaluate the program. This step has many purposes. The most obvious is that it lets both the government and research team know whether or not a given program is having a positive direct impact on students’ academic achievement.

METHODOLOGY

Several statistical approaches are available to help answer the above question the traditional approaches to data analysis in this arena has been to use simple means for assessing the cumulative effects of several variables when allowed to work in combination. Most researchers turn to some form of multiple regression (2) this technique determine the linear relationship between a set of predictions and a criterion in terms of the model.

$$Y = a_1x_1 + a_2x_2 + \dots + a_kx_k + b.$$

Here, Y is a criterion such as course grade

x_i are predictors such as MAT grade

a_i 's are scores or weight associated with each predictors

and b is a constant.

This technique provides a measure of association in terms of the amount of variant accounted for each variable along with an estimate of their combined effect on the criterion.

While multiple regression has long been used in evaluating educational programs, its purpose and model's equation are based on the assumptions that predictors and criterion are continuous in nature. Most variables used in educational interventions are not continuous course grades, minority status, sex, value, etc., are clearly discrete and MAT are better characterized, statistically as polychromatic ones since MAT's range only from 001 to 100. On the criterion side, most variables in educational program are likewise categorical (i.e., grade received).

Research has shown that violations of the assumptions underlying regression modelling can have serious repercussions (3). A more appropriate technique under the current circumstance is known as discriminant analysis.

Ronald Fisher (2004) 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, research has shown (4, 5) that it performs quite well using such data discriminant analysis forms linear combinations of the prediction which are used to classify cases into various groups of the criterion.

One may conceptualize discriminant analysis in terms of evaluating the centroid of a group or cases. In the present context, the student cases are placed in criterion “groups” depending on the final course grade. The mean value of a discriminating variable (e.g., MAT or a preceding grade or predictors) for a student in a particular group is evaluated. The bigger the difference between the mean values of the predictors related to the various groups, the more discriminating is that variable.

Discriminant analysis simultaneously analyzes all of these mean difference and determine, (based on backward probability). The key notion here is the breaking of the criterion group into separate identifiable units. This allows for discriminant analysis to better accounts for discontinuous relationship among the variable.

Discriminant analysis is very simplistic. In reality, discriminant analysis based on solving an equation such as:

$$l^1y = k^1x + B$$

While l^1 and k^1 represents matrices consisting of all possible linear combinations of responses on the predictors, respectively. This results in linear combinations that are solved simultaneously. They are also subject to several classification rules. The mathematics involved quickly became quite complicated. Most statistical packages include a discriminant analysis procedure. For more detailed analysis of the technique see Hubert (5).

Using the technique, one can determine predictors that are most effective in predicting performance. Unlike regression, significance tests are provided for each possible variable and the equation as a whole. Probably, the biggest advantage of discriminant function over regression is that its measure of prediction ability is in terms of the percent of correct classification. This is possible since the unit of classification is categorical, it predicts category membership given the time grouping of the criteria, and one can determine how many predictions produced by the equation are right. This quantity (% correct) seems more interpretable than % of variance accounted for in regression. It also seems more translatable to educators and administrators not familiar with statistics and concept of variance partitioning.

ANALYSIS AND RESULT

In order to demonstrate the methodology, we now present a case study from Joseph Ayo Babalola University, Bowen University, and Lead City University show that a high percentage of science students fail to make satisfactory grades in the required MAT 111 and CHM 111 courses. At these Universities up to 30% of all the students achieve an unsatisfactory grade in these courses. Since science students receiving such grades are required to repeat the course, this problem represents a great cost to both the students and then universities.

Our intention was to analyze the situation and determine what parameters in the students' past record might represent criteria for identifying those who are likely to be at risk of obtaining unsatisfactory grade. We set out to accomplish this task with the help of discriminant analysis using historical data (1,622 students all together registered and sat for exam in the aforementioned

courses). The first step was to determine in what ways students who performed satisfactorily differed from those who did not. Using data from student records we looked at many possible predictor variables including the following: ethnic group, gender, major, university NECO result, WAEC result, overall GPA, overall CGPA and then performance in continuous assessment within the testing period.

Univariate F-tests shared that all academic performance variables had relationships with performance variables in MAT 111 and CHM 111. However, discriminant analysis found that only three of these variables made significant independent contribution. These are the students overall GPA grade in the MAT 111 course.

The next question is how successful these three variables are at predicting student performance. Discriminant analysis, title regression, produces a discriminant function equation classified students as being "at risk" (predicted grade of E) or not at risk (predicted grade of A, B or C). As can be seen in Table 1, 319 of the 661 students predicted to fail did so. This represents a hit rate of 48%. Also, of 961 students predicted not be at risk, some 828 or 86% achieved a satisfactory grade of C or better. Note that the greatest misclassification occur for "C" students.

It is interesting that discriminant analysis is highly successful in identifying these students who will perform successfully and only 50% successful in identifying students who will perform poorly. A review of the students' records shows that the better students have erratic records. The data sample is inevitably skewed and cannot represent "normal" distribution as the students with the very weakest records would have left the system at an earlier stage and are not part of the analyzed group.

Table 1: Predicted and Actual Grades for MAT 111 over the Years.

Actual Grade	Predicted At Risk (E)	Predicted Not at Risk (A,B,C)	Total
E	319	133	452
C	203	291	494
B	114	328	442
A	125	209	234
Total	661	961	1622

We used the above in several ways. First, these three discriminating parameters serve as a pointer for possible reasons underlying poor performance in CHM 111. They help confirm the groups previously suspicious that student inability to understand, retain, and apply basic concepts. It is easy to see how such an analysis would be useful in designing an educational program. Without the discriminant function to direct the research program, one might have reached a vastly different conclusion regarding the etiologic of poor performance.

The next step in using discriminant analysis was to discover if we could use the discriminant function derived from the analysis to identify students at risk in future session. To do this, the initial discriminant function with three significant variables, was applied to a new group of students entering the course. The results are presented in Table 2. Success in identifying students not at risk was 84%, essentially the same as the historical sample, success in identifying students at risk was 50%, again similar to the historical sample.

The use of this technique was more successful than previous attempts at classification. Use of the discriminant function in this manner allowed us to target our future program on the students who they were designed to help (i.e., those at risk)

The third use of discriminant function is in evaluation of educational program. In CHM 111, we have seen a consistent relationship between

the three predictor variables and performance. The consistency might be useful as a basis of comparison. In most educational program, it is difficult to offer a beneficial program to one group and set aside a true control group. The possible repercussions of such actions should be obvious. If a consistent baseline has been established via discriminant function, it is feasible to substitute this for a control group.

CONCLUSION

Discriminant analysis should be the preferred method of operation in educational program regardless of the other benefits provided. As we have seen, it provides other benefits in addition to being statistically correct procedure.

Data are reduced more efficiently, and non-predictive variables are eliminated earlier the analysis process students at risk are more reliably identified in more easily identifiable terms. The technique also allows for more detailed analysis of errors or prediction than does regression, and does so with much more meaningful measure of effect (% correct prediction, or as confidence interval).

Finally, discriminant analysis can serve as a better basis for comparison than regression analysis for situations where control group are not feasible.

Table 2: Predicted and Actual Grades for CHM 111 over the Years.

Historical Actual Grade	Sample Predicted at Risk	Predicted Not at Risk	Total
F	108	52	160
C	78	90	168
B	26	116	142
A	05	59	64
Total	217	317	534

REFERENCES

1. Akomolafe, A.A., O.M. Olayiwola and O.A. Bello. 2008. "Using Multivariate Analysis to Identify Students at Risk in Nigerian Universities". *Journal of Educational Research and Development*. 3(1) 32-37.
2. Cook, R.D. and S. Weisberg. 1982. *Residual and Influence in Regression*. Chapman and Hall: London, UK.
3. Gilbert, E.S. 1988. "On Discriminant using Qualitative Variable". *Journal of American Statistical Society*. 3(1):18-28.
4. Hedges, L.V. 1998. "Themeta- Analysis of Test Validity Studies: Some New Approaches". In: H. Wanner and H.J. Brawn (eds). *Test Validity*. Erlbaum: London, UK. 191-212.
5. Huberty, C.J. 1994. "Applied Discriminant Analysis". John Wiley and Sons: New York, NY.
6. Lewis-Beck, M.S. 1980. *Applied Regression: An Introduction*. Sage Publication: Boston, MA.
7. Moore, D.H. 1993. "Evaluation of Five Discriminant Procedures for Binary Variables". *Journal of the American Statistical Society*. 399-404.

SUGGESTED CITATION

Akomolafe, A.A. and G.N. Amahia. 2015. "Efficiency of Discriminant Analysis in Identifying Performance of Students at Risk in Nigerian Private Universities". *Pacific Journal of Science and Technology*. 16(1):201-205.

 [Pacific Journal of Science and Technology](http://www.akamaiuniversity.us/PJST.htm)