Continental J. Applied Sciences
Hamidu A. Chamalwa *et al.* (2019) 14 (1): 8 – 24 DOI: 10.5281/zenodo.2612919

Research Article

Multivariate Modeling of Student Placement in Nigeria Higher Institutions in University of Maiduguri: An Application of Fisher's Linear Discriminant Analysis

Hamidu Aliyu Chamalwa[1] Harun Rann Bakari[2] and Emmanuel Sambo Uba[3]
[1,2]Department of Mathematical Sciences, University of Maiduguri
[3]Department of Mathematics and Statistics, Ramat Polytechnic Maiduguri

Abstract

Students' placement into courses of study in most (if not all) tertiary institutions in Nigeria had long been a problematic exercise. Most students get placed into courses of study they are ill suited to and then for which they are ill prepared. This research work titled "Multivariate Modeling of Student Placement in Nigeria Higher Institutions in University of Maiduguri: An Application of Fisher's Linear Discriminant Analysis". Focused on the Statistical Analysis of classifying students into various department based on their scores on Unified Tertiary Matriculation Examination (UTME). The study only covered the Department of Chemistry and Department of Biological science, Faculty of Science, University of Maiduguri. Discriminant Analysis technique was used and results obtained indicated that many students were wrongly admitted in to both Departments (Department of Chemistry and Department of Biological Science). 45% of students were wrongly admitted into Chemistry Department and 41% were also wrongly admitted into Biological Science Department based on the Apparent Error Rate (APER) calculated. Based on the result; it is however recommended that proper statistical measures (statistical application) should be introduced into the admission system of various higher institutions so as to minimized the problems highlighted.

Keywords: Analysis, Discriminant, Linear, placement, Students

Received: 30 December 2018 Accepted: 02 February 2019

ISSN: 1597 – 9928 Science and Education Development Inst., Nigeria

Introduction

It is very obvious that most university graduates do not practice what they spent years to study in the university: lack of passion for the course they studied as they just endure through their years of study or because they graduate with poor grades and cannot find a placement for jobs; due to misplacement of students in their year one of studies.

Also, the number of students that graduate with a Cumulative Grade Point Average (CGPA) of 3.50-5.00 is decreasing annually. Most students just want to graduate and leave the university not considering the class of degree anymore. Some degree students spend more years in the school than stipulated by their course of study. Some students put in several efforts into their studies and yet, do not get the desired result.

All these problems mentioned above are matters arising due to misclassification of students into the appropriate department in the year one of their university studies. One cost of misclassification is a measure of the undesirability of starting a student through university when he/she will not be able to finish and the other is a measure of the undesirability of refusing to admit a student who can complete his course. There the objective of this study is to classify students into mutually exclusive courses on the basis of some apriori information such as Joint Admission Matriculation Examination scores and first year (university) results.

Literature Review

Discriminant analysis is a technique for classifying a set of observation in to predefined classes. Discriminant analysis is concerned with problem of assigning an unknown observation to a group with low error rate.

Johnson and Wichern (1992) defined discriminant analysis and classification as multivariate techniques concerned with separating distinct set of observations (or objects) and with allocating new observations (or objects) to previously defined groups. They defined two goals namely;

Goal 1: To describe either graphically (in at most three dimensions) or algebraically the differential features of objects (or observation) from several known collections (or populations) and

Goal 2: To sort observations (or objects) in to two or more labeled classes. The emphasis is on deriving a rule that can be used to optimally assign new observations to the labeled classes. They used the term discrimination to refer to Goal 1 and used the term

ISSN: 1597 – 9928 Science and Education Development Inst., Nigeria

classification or allocation to refer to Goal 2. A more descriptive term for Goal 1 is separation.

They also explained that a function that separates may sometimes serve as an allocator or classificatory and conversely an allocation rule may suggest a discriminator procedure. Also that Goal 1 and Goal 2 frequently overlap and the distinction between separation and allocation becomes blurred.

Costanza and Afifi (1979) computationally stated that discriminant function analysis is very similar to analysis of variance (ANOVA). Cooley and Lohnes (1971`) viewed discriminant analysis technique measure for description and testing of between group differences. Lachenbruch (1975) viewed discriminant analysis as the problem of assigning an unknown observation to group with low error rate. The theoretical basis by Lachenbruch (1968) elaborated that the basic problem in discriminant analysis is to assign an unknown subject to one of two or more groups on the basis of a multivariate observation. It is important to consider the costs of the assignment, the a priori probabilities of belonging to one of the groups and the number of groups involved.

Bayesian approach to classification is the process of assigning an observation to a group with the greatest posterior probabilities. Bayesian approaches to classification problem in the case of normally distributed observations with unknown parameters were discussed by Geisser (1964, 1966, and 1967) as well as Dunmore (1966). Classifications procedure based on Baye's formula was also discussed by Birnbaum and Maxwell (1960).

Welch (1939) explicitly showed the optimality for linear discriminant function under the assumption of multivariate normal densities with common covariance matrix. He obtained his linear discriminant function by minimizing the total probability of misclassification.

Martinez (2001) also reported on the ability of discriminant analysis as a statistical predictive tool in placing students into the right course that matches students' capabilities after conducting a series of standardized tests on the students after admission. Subsequent researches focused not only in predicting student's placement through the use of performance, but also on other factors that might influence a student's performance in the school. Dyer et al. (2002) identifying a student at risk of failure of the academic programme Pyrt (1986) used discriminant analysis to illustrate

ISSN: 1597 – 9928 **Science and Education Development Inst., Nigeria**

how the technique can maximize the effectiveness and efficiency of the screening procedure for identifying intellectually gifted children.

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 one probability of misclassification given the other and how to obtain a minimax discriminant procedure is non-linear. The best linear discriminant function for this unequal covariance matrix context was found by (Clunies-ross and Riffenberg) (1960) and later in 1962 by Anderson and Bahadur. Chernoff (1972, 1973) suggested some 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.

The robustness of linear discriminant analysis and the effect of failure of assumption to hold, have been studied by Gilbert (1969), Moore (1973), Marks and Dunn (1974), Lachenbruch et al. (1973) and Lachenbruch (1975).

The effect of unequal covariance matrices on the linear discriminant analysis was studied by Gilbert (1969) for the large sample case and later by Marks and Dunn (1974) for small sample case. They both concluded that the linear function is quite satisfactory provided that the covariance matrices are not too different.

An important result of non-parametric estimation of linear classification was suggested by Greer (1979, 1984), he considered the algorithms designed to produce a hyper plane that minimizes classification rules in a completely non parametric manner for a large set of loss function.

The estimation procedure associated with a suitable loss function is consistent for any two underlying distributions, whether these are continuous, discrete or mixture of the two variables. Breiman et al. (1984) proposed a non parametric method that yielded a perfect selection of the two groups, when possible. Their method however, has no continuous system.

The main crux of multivariate analysis is to explain variation according to one or more of the variables respectively. The aim of most applied studies is concerned with variation on some characteristics. This variation can be explained according to one or more of the following:

1. Determination of the nature and degree of association between a set of criterion variables and a set of predictor variables, often called "dependent" and "independent" variables respectively.
2. Finding a function or formula by which one can estimate values of the criterion variables.
3. Placing confidence limits or intervals on the estimates via the test of statistical significant.
4. Among the several techniques available in the application of multivariate analysis, one would be singled out for application in this particular study, namely; Discriminant Analysis (DA).

Fisher (1936) Who first develop the linear discriminant function in terms of the analysis of variance, gave as an example the assigning of Irish plant to one of two species on the basis of the lengths, widths of the sepals and petals; Six measurements on a skull found in England were used to determine whether it belonged to the Bronze age or the iron age (Rao, 1952). A patient is admitted to a hospital with a diagnosis of myocardial infection. Systolic blood pressure, heart rate, stroke index, and mean arterial pressure are obtained. Is it possible to predict whether a patient will survive? Can we use these measurements to compute the probability of survival for a patient?. Indian men have been classified into three (3) castes on the basis of stature, sitting height and nasal depth and height (Rao, 1948).

An example of a problem of classification of Prospective students applying for admission into college is given a battery of test. The prospective student may be a member of one population consisting of those students who will successfully complete college training or, rather have potentialities for successfully completing training or the student may be a member of the other population, those who will not complete the college course successfully. The problem is to classify a student applying for admission on the basis of his scores on the entrance examination.

Materials and Methods

Fisher's linear discriminant analysis was used in this study to classify the students in various courses on the basis of the information of JAMB and First year results

Assignment Rule

For $I = 1, 2, \ldots N_j, \quad j = 1, 2$

Assign the ith individual to $\prod_1$

Iff $Y_i > \frac{\bar{Y}_1 + \bar{Y}_2}{2}$ else to $\prod_2$

Iff $Y_i \leq \frac{\bar{Y}_1 + \bar{Y}_2}{2}$

Bayesian Approach to Classification

A Bayesian criterion for classification is one that assigns an observation of a population with the greatest posterior probability. A Bayesian criterion for classification is to place the observation. In $\prod_1$

If $P(\prod_1 / x) > P(\prod_2 / x)$

By Bayes theorem

$$P(\prod_i / x) = \frac{P(\prod : x)}{P(x)}$$

$$= \frac{q_1 \, f_1(x)}{q_1 \, f_1(x) + q_2 \, f_2(x)}$$

Hence the observation X is assigned to $\prod_1$

If $\frac{q_1 \, f_1(x)}{q_1 \, f_1(x) + q_2 \, f_2(x)} > \frac{q_2 \, f_2(x)}{q_1 \, f_1(x) + q_2 \, f_2(x)}$

The above rule reduces to assigning the observation to $\prod_1$ if

$\frac{f_1(x)}{f_2(x)} > \frac{q_2}{q_1}$

And to $\prod_2$ otherwise

ISSN: 1597 – 9928 Science and Education Development Inst., Nigeria

Where

$$\frac{f_1(x)}{f_2(x)} = exp\{\left[X - \frac{1}{2}\,(\mu_1 - \mu_2\,)\right]' \Sigma^{-1}(\mu_1 - \mu_2\,)\}$$

q_1 = the proportion of $\prod_1$ in the population

$q_2 = (1 - q_1)$ = the proportion of $\prod_2$ in the population.

Welch's Criterion

An alternative way to determine the discriminant function is due to Welch (1939). Let the density functions of $\prod_1$ and $\prod_2$ be denoted by $F_1(x)$ and $F_2(x)$ respectively. Let q_1 be the proportion of $\prod_2$ in the population. Suppose we assign X to $\prod_1$ if X is in some region R_1 and to $\prod_2$ if X is in some region R_2. We assume that R_1 and R_2 are mutually exclusive and their union includes the entire space R. the total probability of miscalculation is

$$T(R, f) = \; q_1 \int_R f_1(x)dx + \; q_2 \int_R f_2(x)dx$$

$$= q_1(1 - \int_R f_1(x)dx) + \; q_2 \int_R f_2(x)dx$$

$$= \; q_1 + \int_R (q_2 f_1(x) + \; q_1 f_1(x))dx$$

This quantity if minimized if R_1 is chosen such that $q_2 f_2(x) \; = \; q_1 f_1(x) < 0$ for all points in R_1.

Thus, the classification rule is:

Assign X to $\prod_1$ of

$$\frac{f_1(x)}{f_2(x)} > \frac{q_2}{q_1}$$

And to $\prod_2$ if otherwise, it is pertinent to note that this rule minimizes the total probability of misclassification.

Probabilities of Misclassification

In constructing a procedure of classification, it is desired to minimize the probability of misclassifications or more specifically, it is desired to minimize on the average the bad effects of misclassification. Suppose we have an observation from either population $\prod_1$

or population $\prod_2$ the classification of the observation depends on the vector of measurements.

$X^{'} = (X_1, X_2, \dots X_{IK})'$ on the observation. We set up a rule that if an observation is characterized by certain sets of values of$X_1, X_2, \dots X_{IK}$, we classify it as from $\prod_2$. We think of an observation as appoint in a K-dimensional space. We divide the space into two regions or groups. If the observation falls inR_2, we classify it as coming from population$\prod_2$.

In following a given classification procedure, the statistician can make two kinds of errors in classification. If the observation is actually from$\prod_1$, the statistician or researcher can classify it as coming from$\prod_2$; or if it from$\prod_2$, the statistician may classify it as from $\prod_1$. We need to know the relative undesirability of those two kinds of misclassification.

Let the probability that an observation comes from population $\prod_1$ be q_1 and from population $\prod_2$ be q_2. Let the density function of population $\prod_1$ be $f_1(x)$ and that of population $\prod_2$ be $f_2(x)$. Let the regions of classification from $\prod_1$ be R_1 and from $\prod_2$ beR_2. Then the probability of correctly classifying as observation that is actually drawn from $\prod_1$ is $\int_{R_1} f(x)dx$ where $dx = dx_1,\ dx_2, \dots dx_K$ and the probability of misclassifying an observation from $\prod_1$ is $P_1 = \int_{R_2} f(x)dx$

Similarly the probability of correctly classifying an observation from $\prod_2$ is $\int f_2(x)dx$ and the probability of misclassifying such an observation is$P_2 = \int_{R_2} f(x)dx$.

The probability of misclassification can be computed for the discriminant function. Two cases have been considered. (A) When the population parameter are known.

(B) When the population parameter are not known but estimated from samples drawn from the two populations.

Confusion Matrix

Population		$\prod_1$	$\prod_2$
	$\prod_1$	Correct classification	P_1
	$\prod_2$	P_1	Correct classification

Apparent Error Rates (APER)

One of the objectives of evaluating discriminant function is to determine its performance in the classification of future observations. When the (APER) is

$$T(R,f) = q_1\int_{R_2} f_1(x)dx + q_2\int_{R_1} f_2(x)dx$$

If $f_1(x)$ is multivariate normal with mean (μ_1) and covariance($\sum$), we can easily calculate these rates. When the parameters are not known, a number of error rates may be defined. The function $T(R,f)$ defines the error rates (APER). The first argument is presumed distribution of the observation that will be classified.

Results and Discussions

Table 1: Group Statistics

COURSE		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
CHEMISTRY	ENGLISH	45.3700	6.33199	100	100.000
	PHYSICS	42.6500	6.87460	100	100.000
	CHEMISTRY	42.2700	7.60098	100	100.000
	BIOLOGY	44.0100	7.10882	100	100.000
BIOLOGICAL SCIENCE	ENGLISH	45.0100	7.60316	100	100.000
	PHYSICS	44.3300	7.61585	100	100.000
	CHEMISTRY	43.7600	7.52118	100	100.000
	BIOLOGY	42.7900	8.15047	100	100.000
Total	ENGLISH	45.1900	6.98123	200	200.000
	PHYSICS	43.4900	7.28528	200	200.000
	CHEMISTRY	43.0150	7.57905	200	200.000
	BIOLOGY	43.4000	7.65263	200	200.000

This table presents the means and standard deviations for the samples; values suggest that they differ to some extent.

Table 2: Covariance Matrices

COURSE		ENGLISH	PHYSICS	CHEMISTRY	BIOLOGY
CHEMISTRY	ENGLISH	40.094	-2.839	-2.515	8.946
	PHYSICS	-2.839	47.260	6.489	-6.128
	CHEMISTRY	-2.515	6.489	57.775	-1.215
	BIOLOGY	8.946	-6.128	-1.215	50.535
BIOLOGICAL SCIENCE	ENGLISH	57.808	15.391	19.447	9.366
	PHYSICS	15.391	58.001	17.474	.979
	CHEMISTRY	19.447	17.474	56.568	16.636
	BIOLOGY	9.366	.979	16.636	66.430

Table 2 shows the covariance for the pairs of the variables

Box's Test of Equality of Covariance Matrices

Table 3: Log Determinants

COURSE	Rank	Log Determinant
CHEMISTRY	4	15.451
BIOLOGICAL SCIENCE	4	16.007
Pooled within-groups	4	15.835

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Table 4: Test Results

Box's M		21.026
F	Approx.	2.057
	df1	10
	df2	187429.482
	Sig.	.024

The box- test for equality of population covariance (log Determinant and test results) suggest that difference is statistically significant at 15 level of significance F (10, 187429.482) = 2.057, p= 0.024. This shows that the equality of covariance matrices assumption required for Fisher's linear discriminant approach is valid.

Tests null hypothesis of equal population covariance matrices.

Summary of Canonical Discriminant Functions

Table 5: Eigen values

Function	Eigen value	% of Variance	Cumulative %	Canonical Correlation
1	.028[a]	100.0	100.0	.166

The Eigen value is 0.028, representing the between group and within group sum of squares. The canonical correlation is 0.166, which is simply the Person's correlation between discriminant function scores and group membership i. 0.027 (2.556%) of variance in the discriminant function score can be explained by group differences. The Wilk's Lambda provides a test for assessing the H_0, which the population vector of means of the three scores is in the two groups. This is simply proportion of the total variance in the discriminant scores not explained by difference among the groups, 97.3%. The chi-square-test confirms that the scores do not differ significantly between the courses. $X^2(4) = 5.454$, $p=0.044$

Table 6: Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.973	5.454	4	.044

Table 6a:Standardized Canonical Discriminant Function Coefficients

	Function
	1
ENGLISH	-.232
PHYSICS	.575
CHEMISTRY	.566
BIOLOGY	-.486

The Wilk's Lambda provides a test for assessing the H_0

Table 6b: Structure Matrix

	Function
	1
PHYSICS	.693
CHEMISTRY	.590
BIOLOGY	-.477
ENGLISH	-.154

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions Variables ordered by absolute size of correlation within function.

Table .6c: Canonical Discriminant Function Coefficients

	Function
	1
ENGLISH	-.033
PHYSICS	.079
CHEMISTRY	.075
BIOLOGY	-.064
(Constant)	-2.408

Unstandardized coefficients

Table 7: Functions at Group Centroids

COURSE	Function
	1
CHEMISTRY	-.167
BIOLOGICAL SCIENCE	.167

Unstandardized canonical discriminant functions evaluated at group means

Classification Statistics

Table 8: Prior Probabilities for Groups

COURSE	Prior	Cases Used in Analysis	
		Unweighted	Weighted
CHEMISTRY	.500	100	100.000
BIOLOGICAL SCIENCE	.500	100	100.000
Total	1.000	200	200.000

Table 9a: Classification Function Coefficients

	COURSE	
	CHEMISTRY	BIOLOGICAL SCIENCE
ENGLISH	.652	.641
PHYSICS	.668	.694
CHEMISTRY	.419	.444
BIOLOGY	.625	.603
(Constant)	-52.319	-53.124

Continental J. Applied Sciences

Fisher's linear discriminant functions

Table 9b: Classification Results[a,c]

		COURSE	Predicted Group Membership		Total
			CHEMISTRY	BIOLOGICAL SCIENCE	
Original	Count	CHEMISTRY	59	41	100
		BIOLOGICAL SCIENCE	45	55	100
	%	CHEMISTRY	59.0	41.0	100.0
		BIOLOGICAL SCIENCE	45.0	55.0	100.0
Cross-validated[b]	Count	CHEMISTRY	58	42	100
		BIOLOGICAL SCIENCE	52	48	100
	33%	CHEMISTRY	58.0	42.0	100.0
		BIOLOGICAL SCIENCE	52.0	48.0	100.0

a. 57.0% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 53.0% of cross-validated grouped cases correctly classified.

The Fisher's Linear Discriminant Function is simply derived by subtracting the coefficients given for the variables in each group in the **Classification Function Coefficients table 9a above** as follows

$$Y = 0.011\text{English} - 0.026\text{Physics} - 0.025\text{Chemisty} + 0.022\text{Biology} \qquad **$$

The difference between the constant coefficients provides the sample mean of the discriminant function scores.

$$\bar{Y} = 0.805$$

The coefficients defining fisher's linear discriminant function in ** above is proportional to the canonical discriminant function coefficients

The discriminant scores are centered so that they have a zero mean. These scores can be compared with the average of the group means to allocate students into courses. Here, the threshold against which students discriminant score is evaluated is

$$½(-0.167 + 0.16) = 0$$

Therefore, a student with discriminat score greater than 0 will be assigned to Biological science, otherwise, they would be assigned to Chemistry. According to the estimate 57% of the students are correctly classified into chemistry and Biological Science on the basis of the discriminant rule as shown in Classification **Results table 9b.** 41% and 45% of the students were wrongly classified into chemistry and biological science respectively (**Table 9b**)

Conclusion and Recommendations

A reference to the analysis and discussion of results it was discovered that many students were mistakenly admitted into both departments (department of Chemistry and Biological Science); 41% of students were mistakenly admitted into chemistry department and 45% of students were also mistakenly admitted into Biological Science department. This might hinder the performance of such student who has been wrongly misclassified into either of both departments. However, those students could perform to their maximum capacity if they had been admitted to their appropriate programme.

It is however recommended that institutions can utilize this dynamic and powerful procedure into the system of conducting admission in the various institutions, so as to reduce and remove bias, misclassification and minimize errors, in order to ensure optimum performance by the students.

The Fisher's linear discriminant analysis could also make a good achievement if implore to "Post University Tertiary Matriculation Examination and Pre-degree (REMEDIAL) programme" which can be effectively used to classify prospective student into various Universities, Polytechnics and or Colleges of education.
The following will be achieved if the school will adopt the use of this statistical analysis in the admission processes.

1. The students will have a better performance.
2. The students will be able to graduate with good result and will be able to find a placement for job after graduation.
3. The number of students having spill over will decrease and students will not suffer much to pass their examinations.
4. Many graduates will find it easy to practice what he/she studied.

References

Altman E.I. (1978), Financial ratios *Discriminant Analysis and the Predicting of Corporate Bankruptcy*. J. Finance, 23: page 589-609.

Anderson T.W. (1958), *An Introduction of Multivariate Analysis*. John Willey and Sons Inc. Page 207-210.

Anderson T.W. (1973), *Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions*. In Discriminant analysis and applications, T. Cacoullos edition New York Academic Press: page 17-35.

Anderson T.W. and Bahadur R.R. (1962), *Classification into two Multivariate Distributions with Different co-variance matrices*. The Annals of mathematical statistics, 33: page 420-431.

Birnbaum A. and Maxwell A.E. (1960), *Classification Procedure Based on Baye's Formula*. Applied Statistics 9: Page 152 – 169.

Breiman L., Friedman J.H., Olshen R.A., and Stone C.J (1984), *Classification and Regression Tree*. Belmont, C.A., Wadsworh.

Chernoff H. (1972), *The Selection of Effective Attribute for Deciding between Hypothesis Using Linear Discriminant Functions*, In Frontiers of Pathern Recognition, ed. S. Watanabe, New York, Academic Press: Page 55-60.

Chentir A., Guerti M. and Hirst D.J. (2008), *Classification by Discriminant Analysis of energy in view of the detection of accented syllables in standard Arabic*. J. Comput. Sci., 4: Page 668-673.

Clcnies- ross C.W. and Riffenberg R.H. (1960), *Geometry Linear and Discriminant*, Biometric, 47: Page 185-189.

Cooley W.W. and P.R. Lohnes (1971), *Multivariate Data Analysis*. John Wiley and Sons, Inc.

Costanza W.J. and Afifi A.A. (1979), Comparison of stopping rules in forward stepwise Discriminant Analysis. *Journal of American Statistical Association,* 74: Page 777-785.

Dunsmore I.R. (1966), *A Bayesian Approach to Classification*, Royal Statistical Society Journal, Ser. B, 28: Page 568-577.

Dyer J.E., L.M. Breja and P.S.H. Wittler, (2002), *Predictors of Student Retention in College of Agriculture.*

Fisher R.A. (1936), *The use of Multiple Measurements in Taxonomic Problems*, Annals of Eugenics, 7: Page 179-188.

Geisser S. (1964), *posterior odds for Multivariate Normal Classifications*, Journal of Royal Statistical Society, Ser. B, 26: Page 69-76.

Geisser S. (1966), *Predictive Discriminant in Multivariate Analysis*, Proc. International Symp. Dayton, Ohio, New York, Academic Press, Inc., Page 149-163.

Gilbert E.S. (1969), *The Effect of Unequal Variance Co-variance Matrices on Fisher's Linear Discriminant Function*. Biometrics, 25: Page 505-516.

Greer R.L. (1979), *Linear Classification Rules solving Inconsistent System of Linear Inequalities Consistent Non-parametric Estimation of Best*. Technical Report 129, Stanford University, Dept. of Statistics.

Greer R.L. (1984), Tree and Hills: *Methology for Maximizing Functions of Systems of Linear Relations*, Amsterdam: North-Holland.

Johnson R.A. and Wichern, D.W. (1992), *Applied Multivariate Statistical Analysis*. Prentice Hall: Page 246.

Lachenbruch P.A. (1968), *On the expected values of probabilities of misclassification in discriminant analysis, necessary size and a relation with the multiple correlation coefficients*. Biometrics 24: Page 823.

Lachenbruch P.A. (1975), *Discriminant Analysis*. Hafner Press New York.

Lachenbruch P.A. and Kupper (1973), *Discriminant Analysis When One Population is a Mixture of Norma*l, Biometrics Z 15: Page 191-197.

Lachenbruch P.A., Sneeringer C. and Revo L.T. (1973), *Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-normality*, Communications in Statistics, 1: Page 39-56.

Marks S., and Dunn, O.J. (1974), *Discriminant Functions When Covariance Matrices are Unequal.*

Martinez D. (2001). *Predicting Students Outcomes Using Discriminant Analysis*. Proceeding of the 39th Annual meeting of the research and planning group, May 2-4 ERIC, Lake Arrowhead, C. A: Page 1-22.

Moore D.H. (1973), *Evaluation of five Discriminant Procedures for Binary Variable*, J.A.S.A., 68: Page 399-404.

Pyryt M.C. (1986), *Using Discriminant Analysis to Identify gifted Children*, J. Educ. Gifted, 9: Page 233-238.

Rao C.R. (1948), *The Utilization of Multiple Measurements in Problem of Biological Classification*, Journal of the Royal Statistical Society, Ser. B, 10: Page 159-193.

Welch B. L. (1939), *Note on Discriminant Functions*, Biometric 31: Page 218-220.