AN EVALUATION OF THE VALIDITY OF DISCRIMINANT ANALYSIS: THE CASE OF M.B.A. PROJECTS, FACULTY OF COMMERCE, UNIVERSITY OF NAIROBI

BY

JOSEPH K. NG'ETHE

A MANAGEMENT RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER IN BUSINESS AND ADMINISTRATION, FACULTY OF COMMERCE, UNIVERSITY OF NAIROBI.

JULY 1991

DECLARATION

This Research Project is my original work and has not been presented for a degree in any other University.

Joseph K .Ng'ethe

This Management Project has been submitted for examination with my approval as University supervisor.

Michael K. Chirchir.

Date 8-10-91

Acknowledgement

I wish to express my deep and sincere gratitude to all those who contributed directly or indirectly in facilitating this study.

Special thanks to my supervisor, Mr Michael K. Chirchir without whose input this project is unlikely to have been what it is.

My profound thanks go to Danny Fernandes, who not only aroused my interest in this study, but also contributed a lot especially during the proposal presentation stage.

I also wish to thank my colleagues in the M.B.A. class for their contribution. My indebtedness to Suneeta Prabhakar is both extensive and deep for her company, support and encouragement. Special thanks to my colleague Gichuki, who provided painstaking and thorough reviews especially of the data analysis. On more than one occasion he showed me how complicated ideas could be explained with a minimum of fuss.

I wish to register my appreciation to my parents Paul Ng'ethe and Monica Mugure for the love and support they gave me and ensuring that I never lacked anything that could hinder my academic progress.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The primary objective of this study was to evaluate the validity of the discriminant method and results as used in Master of Business and Administration (M.B.A.) projects in the University of Nairobi. This was done by testing whether data used in the analysis met the underlying assumptions of discriminant analysis, testing the significance and stability of the coefficients and lastly by determining the error of classification.

Two M.B.A. projects were analysed, one written by Chirchir (1989) and the other by Nderitu (1989) For Chirchir's project both assumptions for multivariate normality and equality of group dispersion held true. It was also found out that the error of classification was quite high. It was also found that Economics, Fundamentals of Accounting and Quantitative Methods were significant.

In the case of Nderitu's project, the variables used were dummy variables and therefore the two most important assumptions did not hold true. In fact in her project she agrees that her model lacked in validity. For dummy variables it would have been more appropriate to develop a quadratic discriminant function (Dillon 1979). Therefore, given the importance of her study one could replicate it

and this time develop a quadratic discriminant function.

Both studies are quite applicable in the University of Nairobi. Chirchir's findings could be used to guide students on option selection. If Nderitu's study could be improved on then it could also be useful to the M.B.A. admission board in selection of students.

1. INTRODUCTION

1.1 Background

In all organisations managers are confronted by a set of interacting problems for which they have responsibility. Modelling is used to structure these problems and to derive suggestions for action. A model is a representation of the real situation. For a model to be meaningful it should be as close to reality as possible. In modelling a real situation, various assumptions are made to enable a tool of analysis be applied. The results from such a model will be subject to the assumptions. To apply these results it is important to find out to what extent the real situation could be represented by the model. The process by which this is done is termed as validation. Fishman and Kiviat(1968) state that "*model validation tests the agreement between the behaviour of the model and the real world system being modelled* "

Recent studies in validation of statistical models have found out that many researchers who use these tools for data analysis do not validate them. Crask and Perreault(1977) addressed their research on the validation of discriminant analysis in marketing research. They found out that only a few researchers validated their findings.

Another study by Finlay(1987) specifically looked at validation of research models. Finlay studied a total of 25 projects in 15 industrial/commercial concerns. He found very few cases where the data used was formally validated and where validation was done, it was rarely in a carefully structured manner. Finlay observes that widespread availability of computers has altered the emphasis within model building. To him this is the major factor that has led to the misuse and misapplication of tools of analysis. He argues that validation will prevent silly models being developed with the very easy to use packages that are now available.

Discriminant analysis has become an increasingly useful tool in the analysis of data used in M.B.A projects in the University of Nairobi. The increased usage of discriminant analysis has been due to a number of factors such as:

a) Introduction of microcomputers in the faculty of commerce. This has made it easier for students to analyse their

data using packages such as the STATGRAPHICS package.

b) Students are increasingly getting exposed to these statistical tools especially in the management science seminars offered in the Faculty of Commerce.

In research the choice of an appropriate tool of analysis is crucial and students need to exercise greater care and good judgement in the application of, not only discriminant analysis, but also any other tool of analysis. Thus it is important that ways of improving the choice of tools of analysis, especially for M.B.A projects are sought.

No statistical model will ever perfectly fit the data and there will be discrepancies, large or small, between the data and their model. This project addresses the issue of assessing the suitability of a proposed discriminant analysis model. This is because an important issue to someone interested in the utilisation of results from such projects that used discriminant analysis is the validity and reliability of these results seen in terms of assumptions, rightful inference and with what probabilty confidence.

1.2 Statement of the problem

In evaluating the validity of the results of discriminant analysis two important issues are addressed.

These are:

a) How well the assumptions on which discriminant analysis is based hold true.

b) The stability and significance of the coefficients.

Any researcher who uses discriminant analysis for data analysis makes several assumptions. There are at least four assumptions that are commonly made and must hold true in order to make the results of discriminant analysis valid. If these assumptions do not hold true then the validity of the findings would be grossly jeopardised as outlined in the table below hence the need for the study.

TABLE 1.1: ASSUMPTIONS

ASSUMPTION	VIOLATION EFFECT
1. Multivariate normality	bias test of significance and estimated error rates.
2 Equality of Group Dispersion.	affects significance tests for the differences in group means and the appropriate form of the classification
3 Scale of measurement Interval or stronger	the assumptions may not hold
4. Discrete and Identifiable Variables	Classification not possible

Therefore as it can be seen if any of these assumptions are violated the effect may be to render the results of the study invalid. As was noted by Kiilu(1990) many students using regression analysis have been using them indiscriminately without due regard to their assumptions and without testing for the significance of the estimates of the models they develop. This project sets out to find whether the same conclusions can be made on projects that used discriminant analysis.

1.3 Objectives of the study

The objective of the study is to evaluate the validity of the results of M.B.A projects that have used discriminant analysis.Specifically this will entail :

a) Determining whether the data used in the analysis met the underlying assumptions of discriminant analysis.

b) Testing the significance and stability of the coefficients.

c) Determining the error of classification in the two projects

1.4 Importance of the study

The study will:

1. Bring into light the need to validate models

2. Provide a framework that can be used by researchers to validate discriminant analysis.

3. Bring into light the need for researchers to examine the structure of the data before using the discriminant analysis technique.

4. The researcher will therefore be able to choose a discrimination technique and a method of evaluation which should lead to increased performance.

1.5 Overview of the Study Report

The study report is organised into five chapters. The material presented so far is the introduction which is chapter one. This is followed by a literature review in chapter two. Chapter three is the research design.

The fourth chapter is devoted entirely on data analysis and discussion of the results. Chapter five, the conclusion, gives a summary of the findings and their implications followed by a discussion on the limitations of the study and finally, directions for further research in this area are contemplated.

CHAPTER 2 LITERATURE REVIEW

2.0 An Overview

In this chapter a review of the literature on validation approaches in general is first presented. This is then followed by a review of literature on validation of discriminant analysis. Lastly a summary of the findings of the two projects under study is presented.

2.1 Approaches to Validation

The question that is usually asked when looking at validation is: Are we measuring what intended to measure? The emphasis in this question is on what is being measured. For example,a researcher may be interested in measuring the understanding of a certain concept by respondents. He includes in the test only factual items about the concept. Such a test is not valid, because while it may reliably measure the respondents factual knowledge of the concept, it does not measure their understanding of the concept. In other words, it may measure what it measures quite well, but it does not measure what the researcher intended it to measure.

Therefore, the problem of determining the validity of a particular research effort or measuring instrument lies with the basic definition of validity itself-- that is, are we really measuring what we are trying to measure? If we did know the true value of the characteristic involved, there would have been no reason to conduct research in an attempt to measure it.

Validation is looked at from different aspects. Depending on what aspect of validation one is interested on, one may want to find out how well the model is measuring what it is supposed to measure(content validity) or how well the underlying assumptions have been met(construct validity). also one may also be interested in answering the question; How valid are sample-based results with respect to the broader population of interest? This is called external and/or predictive validity.

The four main approaches to validity are summarised in table 2.1.

TABLE 2.1

TYPES OF VALIDITY

VALIDITY TYPE	WHAT IS MEASURED
Construct	Whether underlying assumptions are met by the model.
Content	How well the model meets its objectives
Predictive	predictive ability of the model
External	inference

A further discussion of these four types of validity follows:

a) *Construct Validity:*

Construct validity measures the the extent to which the underlying concepts and assumptions of a model are met.

It is more abstract and theoretical. Various methods to be discussed later have been introduced to measure this validity.

b) *content validity:* Concerns the extent to which the model "*appears to be*" measuring whatever it is intended to measure.This is usually subjective and involves judgement of experts.

c) *Predictive validity:*refers to the ability of the model to predict some future value associated with variable(s) presumed to have measure.

d) *External Validity:* This measures how well the results can be generalized to a wider population.
All the four dimensions of validity are related especially as far as the technicalities of their measurements are concerned.This will come out more clearly as the various methods are discussed.

2.2 DISCRIMINANT ANALYSIS: VALIDATION

Discriminant analysis begins with a desire to statistically distinguish between two or more groups of cases. To distinguish between the groups the researcher selects a collection of discriminating variables that measure characteristics on which the groups are expected to differ. The three major purposes of discriminant analysis are:

a) developing predictive models to classify individuals into groups.

b) profiling characteristics of groups which are most dominant in terms of discrimination.

c) identifying the major underlying dimensions (i.e discriminant functions) which differentiate among groups.

In validating discriminant analytical models researchers are mainly interested in answering the question: How valid are sample-based dicriminant analysis results with respect to the broader population of interest? This can also be extended for each of the purposes of discriminant analysis, so that in testing for validation one will be interested in answering the following questions:

a) is actual classification potential as high as sample-based estimates indicate?

b) are the true population profiles what they appear to be from the sample results?

c) are the underlying sample-based dimensions generalised in the population?

With the advent of microcomputers, discriminant analysis can be done with ease. This makes it more crucial for such results to be validated. A researcher needs to validate the results of the analysis he has made, as management would not commit its time and funds on the basis of possibly inaccurate results.

The major methods used in validating discriminant analytical model are:

a) Hold-out method

b) Monte-carlo simulations

c) Jackknife method

d) U-method

Holdout Method:

This is the most frequently suggested validation approach. Here the sample is randomly split. One of the subsamples is used to develop estimates of the discriminant coefficients, and these coefficients are applied to the observations in the other subsample for classification purposes.

The method can only be used when the data base is large. It is difficult to split an already small sample, since this will make the derived coefficients even less reliable. Furthermore, as typically applied, this approach is only useful in considering classification and does not help in determining the validity of the profiles or the underlying dimensions.

Monte-carlo simulations

This method has been suggested by Frank R.E (1965)for evaluating discriminant results. Here synthetic data are generated and discriminant functions are derived with the same degrees of freedom as the original data. This approach is useful when predictors are independent. When the predictors are not independent, a problem occurs as it is impossible when generating synthetic data, to model the covariance structure between the predictor variables

Jackknife Method

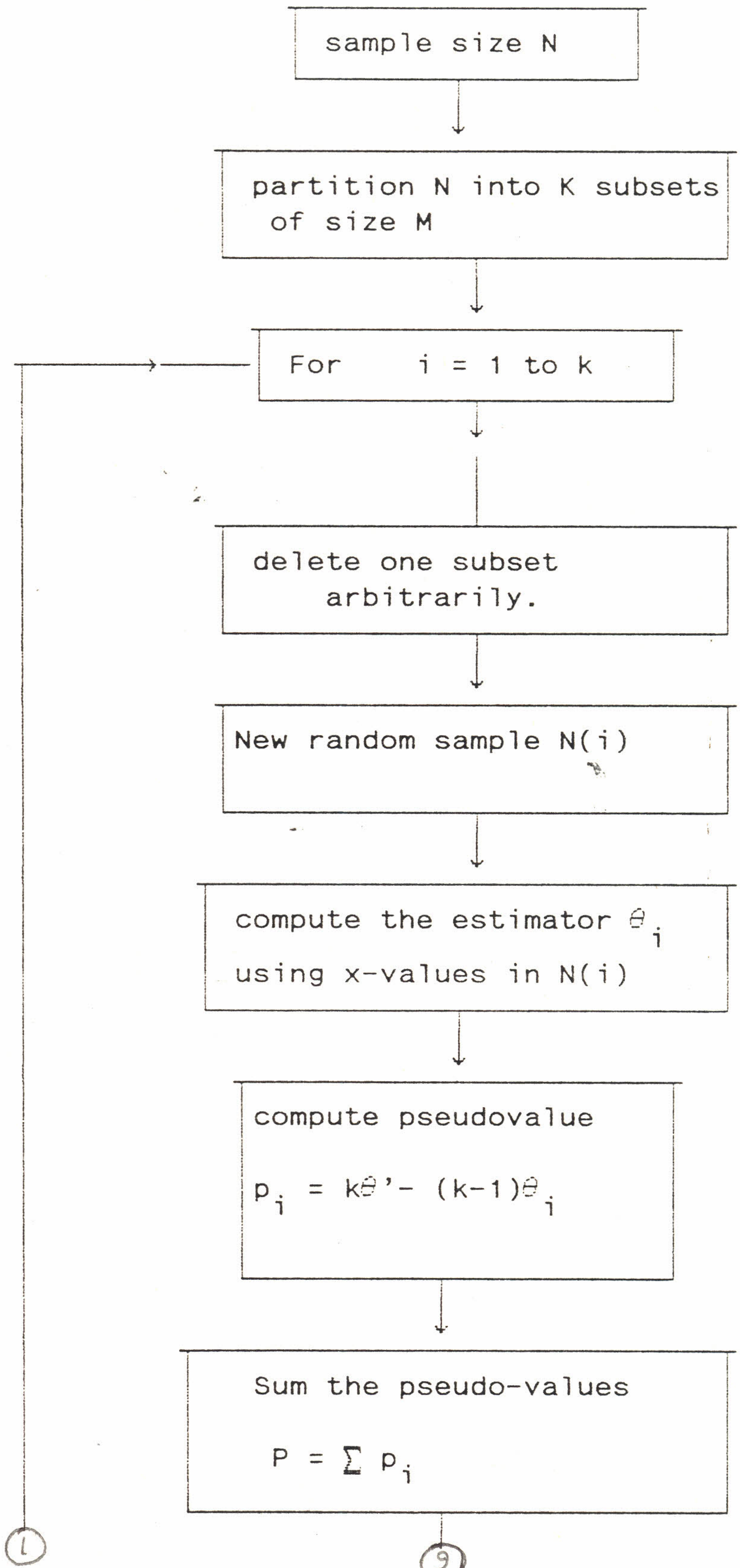
The Jackknife method developed by Tukey(1958) enables an assessment to be made of the stability and significance of results of discriminant analysis without requiring a large sample. This is made possible because of sample re-use. That is a number of discriminant functions are computed each by omitting a different subsample, then pseudovalues are calculated by weighing and subtracting these coefficients, a jackknife coefficient is arrived at by averaging these pseudovalues.

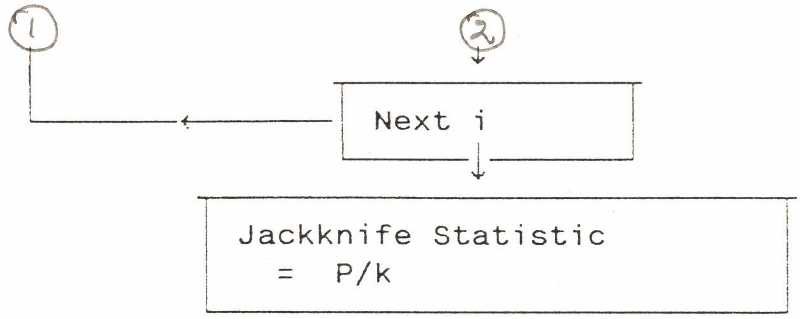
The stability of the coefficients is evaluated directly by computing the traditional standard errors for each set of pseudovalue coefficients. Because the Jackknife coefficients approximate t-distribution (Crask and Perreault 1977) each coefficient can be divided by its associated standard error to give a t-value.

To test for significance of the coefficients, the pseudo values and subsequent Jackknife estimators have important properties Pseudo values can be treated as independent, identically distributed random variables. Hence they can be used to obtain approximate confidence intervals for the jackknife estimate. This confidence interval can be estimated with student t having K - 1 degrees of freedom.

The Jackknife method has its own disadvantages. To compute the Jackknife equations, many successive computer runs are needed, each with selective deletion of subsamples from the analysis. This tends to be cumbersome and expensive.

FIGURE 2.1: THE JACKKNIFE PROCEDURE




1
2
Next i
Jackknife Statistic
= P/k

The U-Method

The U-Method developed by Lachenbruch and Mickey is a similar procedure to the Jackknife method. It is mainly used in the estimation of error rates. One observation is omitted from the sample and a discriminant function is computed by using the remaining observations. This function is then used to determine the group membership of the omitted observation.

By repeating this procedure for all observations in the sample, an estimate of misclassification can be obtained for each group by totalling the number of misclassifications in the group and dividing by the total number of cases in the group.

Unlike the Jackknife technique, it is not possible to determine the stability of coefficients of the discriminant function in this case. This is because the U-Method does not have the bias reduction properties of the Jackknife. The U-Method can be used to cross validate discriminant results by predicting the group membership depending on the error rates of classification.

2.3 Findings of Chirchir

In his M.B.A project submitted in 1989, Chirchir identified the variables that distinguish between students who perform well in the Accounting option as opposed to those who perform well in the non-accounting options, which comprise the Business Administration option and the Insurance option. A sample of sixty students who had passed with at least an Upper Second Class Honours degree in each of the two categories was used to develop a discriminant model.

The predictive variables selected were O-Level and A-Level aggregate scores, and the first year individual subjects, that is:

Economics

Business Law

Fundamentals of Accounting

Quantitative Methods

Business Studies

Business Law

and the Grade Point Average for the first year.

He found out that four courses taught in first year, namely Fundamentals of Accounting, Business Law, Quantitative Methods and Introduction to Economics are the optimal discriminating factors between the two groups. Thus, he concluded these four courses can form a basis upon which students can be assisted in the choice of an option on entering second year.

Validation was done by testing the performance of the model using another sample of 60 in each group from the academic years 1984/85 to 1987/88. A successful prediction rate of 80.8% was achieved using the model developed.

2.4 Findings of Nderitu

In her project also submitted in 1989, Nderitu sought to examine whether certain pre-entry characteristics of M.B.A. students could help distinguish between eventual successful and unsuccessful academic performance in the first year of their studies. The variables she selected were sex, marital status, nationality, sponsor, degree class, first degree, and the number of years that have elapsed since the first degree.

She selected and analysed data for 45 students. She divided them into three groups. These were

Proceed directly, Proceed after supplementary examinations and Discontinue.

She found out that the variables could be used to distinguish among the three groups of performance, but with differing abilities. The two most important factors were the number of years after the first degree and the type of first degree. Marital status, Nationality and the degree class were found to be of moderate importance while sex and the sponsor were of little importance.

On plotting the results she found out that observations for those discontinued were found to be quite distinct as compared to the other two groups. Infact for the other two groups, that is proceed directly and proceed after
supplementary examinations, she found that they were overlapping.

She also tested the performance of the model and found out that it was a very weak model. This is because it could only classify correctly 49%. This could have been due to the kind of classification rule used and the choice of variables.

CHAPTER 3 RESEARCH DESIGN

3.1 Data Collection Procedures

All Master in Business and Administration (M.B.A.) Projects in the University of Nairobithat have used discriminant analysis as a tool of analysis will form the population of the study. The criterion used to select the sample to be used in the study is that the project should have retained the original data used in the analysis. Only two projects met this criterion. These projects are; one written by Chirchir M.K. and submitted in 1989 and the other is by Nderitu, M. submitted the same year.The basic data for the two projects are as shown in the table

Table 3.1 Basic data

Project	No. of Variables	No. of respondents
Chirchir's	9	60
Nderitu's	7	45

Data Collection Forms

Data was collected with the help of data collection forms as shown below:

a) Raw Data

Chirchir's project

	variables 1	2	3	4	5	6	7	8	9
respondents									
1									
2									
.									
.									
.									
120									

Nderitu's project

	variables 1	2	3	4	5	6	7
respondents							
1							
2							
.							
.							
.							
45							

b) Discriminant Scores

respondent	score

c) Discriminant Coefficients

Variable	coefficients Standardised	Unstandardised

3.2 Data Analysis Methodology

There will be three major tests to be carried out. These are:

1) To test whether the assumptions were met.

2) To test the stability and significance of the coefficients.

3) To determine the classification error rate.

1) Testing for the assumptions

Out of the four assumptions only two need testing. These are test for Equality of group dispersion and Multivariate Normality. The other two requirements can be checked through observation. Therefore this sub-section will comprise testing for the two major assumptions viz: Equality of group dispersion and Normality.

(i) Equality of group dispersion.

Normally it is important that this test is done before carrying out the discriminant analysis as it helps in the choice of an appropriate discriminant function. In both projects under study this was not done.

Green (1978) has suggested two tests that can be used, one developed by Bartlett (1947) and the other by Box (1949).Bartlett provided a chi-square approximation statistic for testing the equality of two or more covariance matrices. Box developed a more sophisticated (but more complex) procedure for the same test based on an *F* test. For the purpose of this paper the test developed by Bartlet will be used as it is not time-consuming.

Using Bartlet's test one begins by setting the following null

hypothesis: That covariance matrix over groups are equal.

$COV_1 = COV_2 = \ldots\ldots = COV_G$

where COV_j = covariance matrix for group j, j = 1...G

G = number of groups

The Bartlet's statistic is computed by use of the following formula:

$$B = (m - G) \ln |COV_W| - \sum_{g=1}^{G} (m_g - 1) \ln |COV_g|$$

Where COV_W = a pooled within-groups covariance matrix

m = total number of observations.

This is then compared with the critical chi-square statistic at some predetermined confidence level.

The degrees of freedom is given by:

$$d.f = \frac{1}{2}\left[(G - 1)(n)(n+1)\right]$$

where n = th number of variables

A conclusion can then be reached on whether to reject or fail to reject the null hypothesis.

(ii) **Test for Multivariate Normality:**

b) **Test for Multivariate Normality:**

As noted by Dillon (1979), the problem of testing for multivariate normality has been virtually ignored. This is mainly because most available normality tests are appropriate only for the univariate case. However various authors have proposed various mathodsto test for multivariate normality. Fujikoshi (1978) proposes that a test for multivariate normality should proceed in an indirect way. That is by testing whether the implications of multivariate normality hold true for a given set of data. These implications are:

i) That the exponential of the multivariate normality function: $(\bar{Z}-\bar{U})^{T}\Sigma^{-1}(\bar{Z}-\bar{U})$ follows a chisquare distribution.

ii) That the regression of any component on any other set of components is linear.

iii) That the conditional distribution of any subvector of the data given any other subvector is homescedastic.

If the data does not support these implications then multivariate normality is disproved, however if they support any of the implications then normality may be inferred to some degree.

Green(1978) suggests a chi-square test to find out whether an observation vector has arisen from a multi-variate normal distribution. This is the same as the first implication that Fujikoshi(1978) had proposed to be tested. The first thing to be computed is the chi-square value by substituting the specific values in the expression $(\bar{Z}-\bar{U})^{T}\Sigma^{-1}(\bar{Z}-\bar{U})$ which is the exponential of the multi-variate normality function.

where $\bar{Z}$ is the observation vector

$\bar{U}$ is the mean vector for the observations

Σ^{-1} is the inverse of the variance-covariance matrix.

Then the computed Chi-square statistic is compared with a critical Chi-square. The degrees of freedom is the same as n (the number of variables) and a conclusion as to whether the data is multivariate normality is arrived at.

3.2.1 Test for Stability and Significance of the Estimate

To apply the Jackknife technique the sample under consideration is partitioned into K subsamples. The standard discriminant function is first computed by combining all of the subsamples. Then a discriminant function is computed by using K-1 of the subsamples (i.e.,holding out one subsample). This process is repeated K times, each time omitting a different subsample. The pseudovalues (ρ) are then arrived at by use of the following equation:

$$\rho_i = K(\beta) - (K-1)\bar{\beta}_i$$

where β = coefficient of the full discriminant function.

$\bar{\beta}_i$ = coefficient of a K-1 discriminant function where i = 1 .... K

The Jackknifed coefficients are then computed by averaging the pseudovalues. This Jackknifed coefficients will then be used to test for the significance and stability of the coefficients.

3.2.3 Determining the Classification Error rate

The U-Method

One observation is omitted from the sample and a discriminant function is computed by using the remaining observations. This function is then used to determine the group membership of the omitted observation.

By repeating this procedure for all observations in the sample, an estimate of misclassification can be obtained for each group by totalling the number of misclassifications in the group and dividing by the total number of cases in the group.

CHAPTER 4

DATA ANALYSIS AND FINDINGS

4.1 Introduction

This chapter is divided into two main sections. Section A is on testing the validity of the findings of Chirchir. Section B is on testing the validity of the findings of Nderitu.

Each section will have four subsections corresponding to the four tests of validity alluded to in chapter 3.

4.2 Section A: Findings on Chirchir's Project

The variables used in Chirchir's project will be represented by the following symbols throughout section 1 of the data analysis chapter.

VARIABLE	SYMBOL
O-level aggregate	OLEV
A-level aggregate	ALEV
Introduction to Economics	ECON

Business Law	BLAW
Fundamentals of Accounting	FACT
Quantitative Methods 1	QM
Business Studies	BSTUD
Behavioural Science 1	BSCI
Grade Point Average (First Year)	GPA1

The Accounting option is referred to as Group 1 and the Non-Accounting options as Group 2.

4.2.1 Test for Normality

The statgraphic package was used to find out whether the expression $(\bar{Z}-\bar{U})^T\Sigma^{-1}(\bar{Z}-\bar{U})$ was approximately chi-square distributed. Preliminary computations were carried out with the aid of The LOTUS-123 release 2.1 package. The results are as shown in the appendix.

$\bar{Z}$ = raw data vector for each student.

$\bar{U}$ = mean vector for each variable.

Σ = Variance-Covariance matrix.

It was found out the value of $(\bar{Z}-\bar{U})^T\Sigma^{-1}(\bar{Z}-\bar{U})$ for each observation followed a chi-square distribution at 95% level of confidence. This was confirmed by the Kolmogorov test.(see appendix). Therefore according to this test, the raw data originated from a multivariate normal population.

Test for the assumption that Covariance matrices over groups are equal.

H_o : COV_1 = COV_2 =....... = COV_G H_a : At least one of the COV_G 's to another.

$\alpha = 0.05$

critical $B^*(0.95,45)$ =

Decision rule: we will reject H_o if $B > B^*$

Calculated B (from appendix) = 59.50291

Statistical Conclusion

Since $B > B^*$ then we reject the H_o at α = 0.05

Administrative conclusion

The data used by Chirchir had did not satisfy the assumption of equality of covariance over groups.

Test 3

Testing for the stability and significance of the coefficients.

The Jackknife procedure:

The Jackknifed discriminant coefficients were computed using the original data. This can be shown in appendix 3.

To evaluate the stability of the jackknifed discriminant coefficients, standard errors for each set of pseudovalue coefficients were computed. The Jackknifed coefficients approximate the t-distribution. Each coefficient therefore was divided by its associated standard error to give a t-value; the degrees of freedom for the T-value are based on the number of the subsample minus 1. There were a total of 30 subsamples therefore the degree of freedom is 29. At 97.5% confidence level, the critical t-value is 2.045. Therefore this was compared to the pseudovalues and the most significant variables were found out to be:

Introduction to Economics

Fundamentals of Accounting

Quantitative Analysis.

This can be summarised in the table below.

TABLE 4.1 JACKKNIFE RESULTS I

Variable	Jackknife Coeff.	Computed t-value
Constant	-239.27	-3.8957
OLEV	1.84	1.3196
ALEV	7.13	1.7963
ECON	10.32	7.0390
BLAW	3.55	1.5576
FACT	7.00	3.5279
QM	2.66	2.1621
BSTUD	4.59	1.6030
BSCI	3.09	1.6535
GPA1	-1.73	-0.4703

Classification Error Rate:

The students were grouped randomly into 30 pairs, each pair consisting of a successful accounting student and a successful non-accounting student. By holding one of this sub-sample, a discriminant function with the remaining 29 subsamples was computed and then the hold-out subsample was classified by computing its discriminant score. The results are as shown in appendix 4.

13 students were misclassified. These means that the

model has a classification error of 21.6%. Compared to the confusion matrix computed by Chirchir where he found the classification error to be 19.2%, the U-method in this case has provided a more conservative estimate of classification error.

4.3 Section B: Findings on Nderitu's Project

The following symbols will be used throughout section B of the data analysis chapter whenever encountered.

X_1 = Sex

X_2 = Marital Status

X_3 = Nationality

X_4 = Sponsor

X_5 = First Degree

X_6 = Degree Class

X_7 = Number of years the First Degree.

Those who Proceed directly, Proceed after Supplementary Examination and those who are Discontinue will be referred to as Group 1, Group 2 and Group 3 respectively.

4.3.1. Test for Normality

It was not possible to carry out a test of Normality due to the kind of data used in this project.

4.3.2. Test for the assumption that Covariance matrices over groups are equal.

H_o : COV_1 = COV_2 =....... = COV_G H_a : At least one of the COV_G 's to another.

α = 0.05

critical $B^*(0.95,56)$ = 71.5

Decision rule: we will reject H_o if $B > B^*$

Calculated B (from appendix) =129.50291

Statistical Conclusion

Since $B > B^*$ then we reject the H_o at α = 0.05

Administrative conclusion

The data used by Nderitu did not satisfy the assumption of equality of covariance over groups. This was expected because she used dummy variables.

Test 3

Testing for the stability and significance of the coefficients.

The Jackknifed discriminant coefficients were computed using the original data(see appendix)

For each Jackknife coefficient a t-value was computed. This was compared with the critical t-value at 97.5% confidence level and the results are as shown.Only one variable was found to be significant. This was variable x7, the number of years after first degree.

This can be summarised in the table below.

TABLE 4.2 JACKKNIFE RESULTS II

Variable	Jackknife Coeff.	Computed t-value
Constant	-4.156	-1.3019
X1	3.986	1.5963
X2	1.039	1.5032
X3	-0.004	-0.0091
X4	-0.412	-0.8210
X5	-2.584	-5.2477
X6	0.996	0.7361
X7	0.564	4.3150

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

In order to evaluate the validity of discriminant analysis it is important to note that the underlying assumptions play a crucial role. If these assumptions are not met the validity of the method is jeopardised. However, it is important to note that there are some assumptions that are more crucial than others and they have to be met. The two main assumptions are the multivariate normality and equal variance over groups. These two assumptions have been the subject of this study.

If the multi-variate normality assumption is violated, test of significance and estimating classification error rates may be biased. On the other hand if the covariance over groups is not equal then the significance test for the difference in group means and the choice for appropriate form of classification rule is jeopardised.

The major objective of the study was to evaluate the validity of discriminant analysis as used in the two projects. This was done by determining whether the data used in the analysis met the underlying assumptions of discriminant analysis, testing for significance and stability of the coefficients and by determining the classification error rate.

Conclusions on Chirchir's Findings

The various tests done on the findings of Chirchir indicated that the two main assumptions were not violated. However when testing for the significance and stability, it was found out that Introduction to Economics, Fundamentals of Accounting and Quantitative Methods were quite significant and could be used to be discriminators of the two groups. This could be explained by the fact that over time Fundamentals of Accounting and Quantitative Analysis has been used as a basis for selection for the accounting/non-accounting groups.

It was also found out that the classification error rate is approximately 21.6%, given the rigour of the U-Method,this is not very poor. Compared with the error rate of 19.2% of the test of performance that Chirchir had done.

Policy Recommendations

The Faculty of Commerce can use Fundamentals of Accounting as the basis for selection for students who wish to pursue either of the options. With the introduction of 8-4-4 system of education, there will not be any significant difference. Those students who would wish to specialise on accounting courses and those who would not can be discriminated on this basis.

However, a more detailed study on this field is needed. This study would identify all the possible variables that may affect performance, for example, using subject-to-subject performance in the O-Level exam as criterion variables.

Conclusion on Nderitu's Findings

It was not possible to test all the data in this project for multivariate normality for Nderitu's study. This was because in her study she used dummy variables. As noted earlier on, for dummy variables it is usually better to use a quadratic classification rule rather than linear.

The tests carried out showed that the findings of Nderitu did not satisfy the assumptions. The hypothesis of equal covariance group dispersion was rejected. This has an adverse effect on the test for the equality of group mean vectors.

Policy Recommendations

Given the importance of Nderitu's study it would be of major interest if one would replicate the same study and use a more appropriate classification rule say the quadratic discriminant function. Dillon (1978) observes that for dummy variables it is more appropriate to use a quadratic function.

5.4 Limitations of the Study

Results of this study should be interpreted in light of the following limitations:

a) There are more tests which would have been carried out but due to pressure of time and non-availability of some assessment tools, the study concentrated on only those tests that were possible.

b) There were few projects that satisfied the sample selection criteria.

APPENDIX 1
RAW DATA (CHIRCHIR'S PROJECT)

OLEV	ALEV	ECON	BLAW	FACT	QM	BSTUD	BSCI	GPA1
14	12	63	61	75	73	57	58	65.9
24	13	56	56	65	65	48	51	58.3
15	13	58	48	63	69	77	65	61.8
19	13	65	59	62	70	59	65	63.6
8	22	70	70	75	81	64	54	71
21	12	59	52	54	61	64	60	57.6
23	14	67	57	60	71	62	56	63
11	16	70	58	63	73	61	61	65
20	14	64	57	57	66	58	64	61
16	16	62	49	60	63	75	44	58.7
16	18	71	60	74	79	67	61	69.6
16	13	64	55	56	66	65	63	61
18	20	75	54	76	76	56	58	67.6
25	15	70	62	58	57	54	57	60.5
18	16	68	61	71	68	57	70	66.2
15	18	73	61	78	82	60	62	71
15	15	64	53	65	62	53	55	59.6
22	13	66	53	65	84	64	68	66.8
17	14	62	52	63	83	50	49	61.9
14	14	70	51	60	74	52	49	61.1
12	14	65	45	66	68	53	56	59.7
23	15	77	65	56	68	53	54	63.9
23	14	62	53	63	52	51	47	55.8
13	14	66	51	73	77	53	50	63.7
26	16	64	43	73	59	56	59	59.3
11	15	58	53	66	64	47	57	58.6
14	17	66	51	76	78	61	55	65.8
14	15	61	71	70	62	61	74	66.3
16	13	71	51	61	59	54	60	59.8
9	18	69	64	74	75	53	54	67.1
22	13	62	60	64	65	48	65	61.5
24	13	61	53	66	62	44	62	59
21	13	71	53	72	71	53	61	64.8
14	14	66	51	73	77	53	50	63.7
14	17	66	51	76	78	61	55	65.8
24	11	70	69	54	69	60	60	64.4
13	11	65	65	65	72	54	56	64.4
26	12	68	62	49	69	70	61	62.7
6	25	65	62	83	86	61	59	71.2
13	13	66	51	59	66	51	49	58.4
23	11	73	55	64	62	50	53	61.1
13	14	78	59	81	80	70	66	73.2
12	17	71	69	68	69	74	69	69.7
12	13	56	60	62	45	70	58	57.4
17	14	66	54	66	63	74	62	63.4
14	14	69	60	70	58	62	60	63.6
19	12	71	52	68	72	67	56	64.9
16	17	62	46	72	74	61	60	62.9
18	14	72	59	76	76	71	61	69.8
15	15	79	65	79	70	63	61	71
8	14	67	56	63	59	67	56	61.3
19	14	63	56	58	62	71	60	60.9
16	13	62	63	79	74	65	59	68
12	13	70	63	77	73	69	60	69.7
16	12	65	54	82	78	61	63	68.2
10	18	63	67	71	77	69	71	69.6
21	15	54	49	65	73	67	67	61.6
17	14	59	48	65	75	65	74	63.3

APPENDIX 2

DISCRIMINANT COEFFICIENTS (CHIRCHIR'S PROJECT)

pair	C	OLEV	ALEV	ECON	BLAW	FACT	QM	BSTUD	BSCI	GPA1
0	-216.8	1.725	6.798	2.891	3.403	3.06	2.08	1.79	2.82	-11.08
1	-208.1	1.012	5.765	1.921	3.62	2.98	2.11	1.28	2.72	-10.17
2	-227.9	0.997	5.234	1.87	3.02	3.37	1.98	1.17	2.09	-9.03
3	-197.7	1.091	6.07	2.03	3.52	1.99	2.01	1.87	2.11	-10.21
4	-219.3	2.08	5.091	1.99	2.67	1.75	2.78	1.01	2.81	-10.24
5	-258.1	1.886	6.99	2.01	2.25	2.97	2.45	1.23	2.37	-12.76
6	-221.5	2.01	4.96	2.22	3.13	3.07	2.67	1.08	2.78	-11.35
7	-212.7	2.1	4.48	2.61	3.09	3.21	2.95	0.97	2.95	-11.71
8	-211.3	2.2	4.56	2.13	3.19	3.81	2.15	0.56	2.22	-12.07
9	-202.9	1.9	4.23	2.07	3.01	3.03	1.98	0.78	2.17	-10.76
10	-201.8	2.12	5.58	2.44	3.12	3.68	1.56	1.21	1.98	-9.77
11	-227.3	1.734	6.89	2.19	1.97	2.97	1.95	2.41	3.12	-9.01
12	-217.9	1.12	5.67	2.89	3.76	2.86	1.75	0.61	2.16	-11.35
13	-195.8	0.979	4.25	2.98	4.21	3.16	2.78	1.35	2.98	-10.45
14	-231.6	1.19	6.94	1.77	3.1	3.58	2.36	1.26	3.07	-13.01
15	-228.5	2.8	5.01	2.31	2.08	3.76	3.01	2.05	1.56	-12.95
16	-220.7	1.3	6.12	2.91	3.05	3.87	1.97	2.31	1.8	-11.76
17	-212.9	2.19	5.43	2.34	3.97	2.99	2.01	2.47	1.96	-12.07
18	-209.1	1.78	5.48	2.56	4.09	3.01	1.86	2.33	1.74	-11.79
19	-164.5	1.87	4.78	2.81	3.21	2.79	1.34	2.56	1.98	-11.87
20	-215.01	1.27	3.97	2.91	4.02	2.98	1.78	2.97	0.97	-13.94
21	-245.9	2.98	3.78	3.1	5.76	0.917	1.19	0.895	1.79	-12.99
22	-198.7	1.79	3.56	1.21	4.97	1.313	1.76	1.07	1.01	-10.89
23	-176.34	1.25	2.99	1.76	3.97	1.23	1.89	2.71	1.21	-12.32
24	-158.9	2.13	5.21	1.31	2.56	3.04	2.09	1.25	3.11	-11.32
25	-209.8	1.76	3.11	1.98	3.79	2.65	1.92	2.14	2.19	-12.89
26	-214.3	1.29	3.96	1.93	4.19	2.99	1.27	2.76	1.98	-13.45
27	-237.9	2.62	3.14	2.61	1.97	2.71	2.53	2.31	1.73	-13.44
28	-211.3	1.51	4.17	2.97	3.11	2.54	1.36	0.76	1.22	-14.91
29	-264.4	1.35	4.28	3.45	4.78	1.43	1.73	1.11	2.78	-10.99
30	-253.7	1.19	2.63	3.52	2.6	2.71	1.97	1.22	2.07	-12.97

APPENDIX 3

PSEUDO-VALUES (CHIRCHIR'S PROJECT)

	C	OLEV	ALEV	ECON	BLAW	FACT	QM	BSTUD	BSCI	GPA1
1	-338.6	11.707	21.26	16.471	0.365	4.18	1.66	8.93	4.22	-23.82
2	-61.4	11.917	28.694	17.185	8.765	-1.28	3.48	10.47	13.04	-39.78
3	-484.2	10.601	16.99	14.945	1.765	18.04	3.06	0.67	12.76	-23.26
4	-181.8	-3.245	30.696	15.505	13.665	21.4	-7.72	12.71	2.96	-22.84
5	361.4	-0.529	4.11	15.225	19.545	4.32	-3.1	9.63	9.12	12.44
6	-151	-2.265	3.246	12.285	7.225	2.92	-6.18	11.73	3.38	-7.3
7	-274.2	-3.525	3.925	6.825	7.785	0.96	-10.1	13.27	1	-2.26
8	-293.8	-4.925	38.13	13.545	6.385	-7.44	1.1	19.01	1.12	2.78
9	-411.4	-0.725	4.275	14.385	8.905	3.48	3.48	15.93	1.92	-15.56
10	-426.8	-3.805	23.85	9.205	7.365	-5.62	9.36	9.91	1.45	-29.42
11	-69.8	1.599	5.51	12.705	23.465	4.32	3.9	-6.89	-1.38	-40.06
12	-201.4	10.195	22.59	2.905	-1.595	5.86	6.7	18.31	1.26	-7.3
13	-510.8	12.169	4.247	1.645	-7.895	1.66	-7.72	7.95	0.58	-19.9
14	-9.6	9.215	4.81	18.585	7.645	-4.22	-1.84	9.21	-0.68	15.94
15	-53	-13.325	3.183	11.025	21.925	-6.74	-10.94	-1.85	2.046	15.1
16	-162.2	7.675	-16.29	2.625	8.345	-8.28	3.62	-5.49	1.71	-1.56
17	-271.4	-4.785	1.95	10.605	-4.535	4.04	3.06	-7.73	1.486	2.78
18	-324.6	0.955	25.25	7.525	-6.215	3.76	5.16	-5.77	1.794	-1.14
19	-949	-0.305	-3.505	4.025	6.105	6.84	12.44	-6.99	14.58	-0.02
20	-241.86	8.095	4.639	2.625	-5.235	4.18	6.28	-14.73	2.872	28.96
21	190.6	-15.845	4.905	-0.035	-29.595	33.062	14.54	14.32	1.724	15.66
22	-470.2	0.815	4.13	26.425	-18.535	27.518	6.56	-11.87	2.816	-13.74
23	-783.24	8.375	-21.11	18.725	-4.535	28.68	4.74	-11.09	2.536	6.28
24	-1027.4	-3.945	1.03	25.025	15.205	3.34	1.94	9.35	-1.24	-7.72
25	-314.8	1.235	2.43	15.645	-2.015	8.8	4.32	-3.11	1.164	14.26
26	-251.8	7.815	4.643	16.345	-7.615	4.04	13.42	-11.79	1.458	22.1
27	78.6	-10.805	-10.17	6.825	23.465	7.96	-4.22	-5.49	1.808	21.96
28	-293.8	4.735	4.359	1.785	7.505	10.34	12.16	16.21	2.522	42.54
29	449.6	6.975	4.205	-4.935	-15.875	25.88	6.98	11.31	3.38	-12.34
30	299.8	9.215	6.51	-5.915	14.645	7.96	3.62	9.77	1.332	15.38

	C	OLEV	ALEV	ECON	BLAW	FACT	QM	BSTUD	BSCI	GPA1
JACKKNIFE COEFFICIENT	-239.27	1.842133	7.133393	10.32453	3.547666	6.998666	2.658666	4.587666	3.091266	-1.728
STANDARD ERROR	61.4188	1.39598	3.971	1.46676	2.27761	1.98376	1.22966	2.86196	1.8695	3.67398
COMPUTED T-VALUE	-3.89571	1.319598	1.796372	7.039006	1.557626	3.527980	2.162115	1.602980	1.653525	-0.47033
CRITICAL T(0.975,29)	2.045	2.045	2.045	2.045	2.045	2.045	2.045	2.045	2.045	2.045
DECISION	S	NS	NS	S	NS	S	S	NS	NS	NS

NOTE: S represents SIGNIFICANT at 97.5% confidence level
NS represents NOT SIGNIFICANT at 97.5% confidence level

APPENDIX 4

U-METHOD'S RESULTS (CHIRCHIR'S PROJECT)

GROUP 1		GROUP2	
0.977	C	-0.32476	C
0.91657	C	-2.86229	C
-1.00578	MC	-0.94212	C
1.062915	C	-0.95294	C
1.32638	C	-1.12440	C
-0.93099	MC	0.881574	MC
1.039275	C	-0.92513	C
1.081585	C	-2.09948	C
-1.05993	MC	-0.94739	C
0.940655	C	-0.90683	C
1.270735	C	-1.06476	C
-0.99066	MC	0.307607	MC
1.281145	C	-1.06149	C
1.123755	C	-0.95729	C
1.259115	C	-1.05095	C
-0.27432	MC	-1.07061	C
0.96945	C	-0.92473	C
1.102425	C	-0.95947	C
-0.85571	MC	-0.86206	C
0.851185	C	0.287227	MC
0.87418	C	-2.89153	C
1.102925	C	-0.95558	C
-0.29207	MC	-0.86843	C
0.89385	C	-0.89681	C
1.119235	C	-0.95100	C
-0.39142	MC	-0.14818	C
1.086505	C	-0.98487	C
1.26855	C	-1.07197	C
-0.69532	MC	0.912844	MC
1.12241	C	-1.02268	C

NOTE: C represent those that are correctly classified
MC represent those that are misclassified

RESULTS: 47 CORRECTLY CLASSIFIED
13 MISCLASSIFIED

21.6% MISCLASSIFIED

APPENDIX 5 : CHIRCHIR'S PROJECT

WITHIN GROUP COVARIANCE MATRIX BOTH GROUPS

	OLEV	ALEV	ECON	BSCI	BLAW	QM	FAC	BSTUD	GPA1
OLEV	27.5535	-3.933	-6.172	-1.602	-6.424	-9.263	-12.668	-5.494	-7.646
ALEV	-3.933	4.268	1.384	-0.063	1.996	5.209	5.542	1.436	2.949
ECON	-6.172	1.384	40.795	-3.281	13.03	14.283	13.053	2.963	16.101
BSCI	-1.602	-0.063	-3.281	35.1	4.491	8.686	8.758	18.212	8.941
BLAW	-6.424	1.996	13.03	4.491	42.902	4.575	7.836	8.519	14.964
QM	-9.263	5.209	14.283	8.686	4.575	70.77	32.083	10.538	26.047
FAC	-12.668	5.542	13.053	8.758	7.836	32.083	61.008	10.603	24.443
BSTUD	-5.494	1.436	2.963	18.212	8.519	10.538	10.603	56.695	14.004
GPA1	-7.646	2.949	16.101	8.941	14.964	26.047	24.443	14.004	18.503

INVERSE OF COVARIANCE MATRIX (BOTH GROUPS)

	OLEV	ALEV	ECON	BSCI	BLAW	QM	FAC	BSTUD	GPA1
OLEV	0.045363	0.031510	-0.02043	-0.01178	-0.02054	-0.02315	-0.01796	-0.00984	0.117597
ALEV	0.031510	0.302378	0.007551	0.011983	-0.00876	-0.01357	-0.01651	-0.00250	0.002388
ECON	-0.02043	0.007551	1.302286	0.627057	1.280022	1.271736	1.250708	0.650473	-6.41585
BSCI	-0.01178	0.011983	0.627057	0.339580	0.625474	0.619657	0.608271	0.307062	-3.13061
BLAW	-0.02054	-0.00876	1.280022	0.625474	1.333821	1.295986	1.271422	0.657781	-6.50370
QM	-0.02315	-0.01357	1.271736	0.619657	1.295986	1.301975	1.251940	0.653052	-6.44251
FAC	-0.01796	-0.01651	1.250708	0.608271	1.271422	1.251940	1.262870	0.642005	-6.33187
BSTUD	-0.00984	-0.00250	0.650473	0.307062	0.657781	0.653052	0.642005	0.355646	-3.28664
GPA1	0.117597	0.002388	-6.41585	-3.13061	-6.50370	-6.44251	-6.33187	-3.28664	32.37909

COVAR GRP1

	OLEV	ALEV	ECON	BSCI	BLAW	QM	FAC	BSTUD	GPA1
OLEV	26.27	-1.67	-4.09	4.47	-12.74	-1.79	-14.99	-0.13	-4.46
ALEV	-1.67	2.27	0.29	2.07	0.87	6.75	-1.38	0.8	1.96
ECON	-4.09	0.29	61.55	-8.11	20.61	19.19	15.07	3.62	19.95
BSCI	4.47	2.07	-8.11	71.36	-10.23	2.21	-6.86	1.01	11.49
BLAW	-12.74	0.87	20.61	-10.23	54.25	19.87	18.27	1.86	18.5
QM	-1.79	6.75	19.19	2.21	19.87	119.84	-19.46	1.84	30.65
FAC	-14.99	-1.38	15.07	-6.86	18.27	-19.46	54.04	10.27	7.73
BSTUD	-0.13	0.8	3.62	1.01	1.86	1.84	10.27	26.78	5.44
GPA1	-4.46	1.96	19.95	11.49	18.5	30.65	7.73	5.44	17.55

COVAR GRP2

	OLEV	ALEV	ECON	BSCI	BLAW	QM	FAC	BSTUD	GPA1
OLEV	25.81	-2.03	-4.71	0.49	-2.53	4.04	3.53	-2.18	-0.35
ALEV	-2.03	2.09	0.57	0.69	1.19	3.32	-0.55	0.37	1.13
ECON	-4.71	0.57	44.85	10.69	15.84	53.9	16.11	1.65	26.73
BSCI	0.49	0.69	10.69	42.61	-3.15	28.44	-6.36	-7.34	14.37
BLAW	-2.53	1.19	15.84	-3.15	47.18	19.81	24.08	13.03	19.61
QM	4.04	3.32	53.9	28.44	19.81	137.95	6.59	-3.66	48.24
FAC	3.53	-0.55	16.11	-6.36	24.08	6.59	64.85	5.73	15.06
BSTUD	-2.18	0.37	1.65	-7.34	13.03	-3.66	5.73	34.83	4.91
GPA1	-0.35	1.13	26.73	14.37	19.61	48.24	15.06	4.91	23.75

APPENDIX 6 - RAW DATA FOR NDERITU'S PROJECT

X1	X2	X3	X4	X5	X6	X7	X8
1	1	1	2	1	2	0	1
1	1	1	2	1	2	0	1
1	1	1	2	2	2	0	1
2	1	1	2	1	1	0	1
1	1	1	2	1	2	0	1
2	2	1	2	1	2	0	1
1	1	1	2	1	2	0	1
1	2	1	2	1	2	0	1
1	2	1	2	1	1	0	1
1	1	1	2	1	1	0	1
2	1	1	2	1	2	0	1
1	1	1	2	1	2	1	1
2	1	1	2	1	2	1	1
1	1	1	2	1	2	0	1
1	1	1	2	1	1	0	2
2	1	1	2	2	1	0	2
1	1	1	2	1	2	0	2
1	2	2	2	1	1	7	2
1	1	1	2	1	2	0	2
1	1	1	2	1	2	0	2
1	2	1	2	2	2	5	2
1	2	1	2	1	2	0	2
2	2	1	2	2	2	2	2
2	2	1	2	1	2	2	2
1	1	1	2	2	2	5	2
2	1	2	1	3	1	0	2
2	2	1	2	2	2	0	2
2	2	1	2	1	2	2	2
1	1	1	2	3	2	5	2
1	1	2	1	3	2	12	3
1	1	1	2	1	2	1	3
1	1	1	2	1	2	1	3
1	2	2	2	4	3	5	3
1	1	2	2	1	2	1	3
2	1	1	2	1	2	2	3
1	1	1	2	1	2	1	3
2	2	1	2	4	2	14	3
2	1	1	1	4	2	0	3
1	1	1	2	2	2	0	3
1	2	2	2	3	2	6	3
1	2	1	1	4	2	12	3
1	2	1	2	2	2	6	3

APPENDIX 9

U-METHOD RESULTS

APPENDIX 7

DISCRIMINANT COEFFICIENTS (NDERITU'S PROJECT)

pair	C	X1	X2	X3	X4	X5	X6	X7
0	-2.4937	-0.1216	-0.6531	0.7649	-0.3918	0.3523	1.2732	0.1786
1	-2.49408	-0.12871	-0.94808	0.84343	-0.35321	0.35758	1.32141	0.17241
2	-3.22097	-0.13244	-0.6669	0.85676	-0.23864	0.38063	1.35634	0.19409
3	-2.52859	-0.57384	-0.95845	0.725777	-0.06317	0.384244	1.245751	0.113388
4	-2.06290	-0.10940	-0.93956	0.711476	-0.47920	0.337903	1.339894	0.186007
5	-2.54320	-0.19920	-0.94900	0.718627	-0.40382	0.573469	1.299547	0.172478
6	-2.27009	-0.05722	-0.04815	0.793707	-0.05617	0.592778	1.326445	0.100825
7	-3.07376	-1.10456	-0.23229	0.933142	-0.55458	0.619810	1.360679	0.199536
8	-1.16179	-0.01571	-1.00566	0.761530	-0.57253	0.735663	1.262868	0.102757
9	-1.03260	-0.99936	-0.97733	0.740078	-0.54023	0.585055	1.242083	0.196959
10	-3.59927	-0.11508	-0.15203	0.872363	-0.55997	0.710561	1.190732	0.100503
11	-1.16835	-0.91205	-1.03399	0.782981	-0.35357	0.573469	1.238415	0.163458
12	-2.50776	-0.58910	-1.36449	1.033249	-0.67483	0.552230	1.213962	0.121119
13	-2.45289	-0.51493	-0.40698	1.065427	-0.07556	0.610156	1.339894	0.135614
14	-1.52888	-0.62592	-0.83569	0.632820	-0.55638	0.691253	1.288543	0.199858
15	-3.97902	-0.14727	-1.09065	0.825884	-0.37331	0.726008	1.368015	0.106700

PSEUDO-VALUES (NDERITU'S PROJECT)

pair	C	X1	X2	X3	X4	X5	X6	X7
1	-2.48838	-0.02206	3.47662	-0.33452	-0.93206	0.27838	0.59826	0.26526
2	7.68808	0.03016	-0.4599	-0.52114	-2.53604	-0.04432	0.10924	-0.03826
3	-2.00520	6.209873	3.621819	1.312613	-4.99253	-0.09491	1.657481	1.091567
4	-8.52478	-0.29233	3.357418	1.512827	0.831903	0.553857	0.339472	0.074890
5	-1.80066	0.964807	3.489618	1.412720	-0.22342	-2.74407	0.904333	0.264305
6	-5.62415	-1.02282	-9.12227	0.361594	-5.09052	-3.01439	0.527759	1.267447
7	5.627273	13.63991	-6.54437	-1.59049	1.887237	-3.39285	0.048482	-0.11451
8	-21.1403	-1.60397	4.282820	0.812077	2.138506	-5.01478	1.417843	1.240389
9	-22.9490	12.16716	3.886219	1.112398	1.686221	-2.90627	1.708832	-0.07843
10	12.98434	-0.21280	-7.66807	-0.73958	1.962618	-4.66336	2.427746	1.271957
11	-21.0485	10.94477	4.679420	0.511755	-0.92698	-2.74407	1.760183	0.390579
12	-2.29672	6.423422	9.306427	-2.99199	3.570744	-2.44672	2.102523	0.983333
13	-3.06496	5.385131	-4.09867	-3.44248	-4.81915	-3.25768	0.339472	0.780394
14	-16.0010	6.938886	1.903216	2.614007	1.912364	-4.39304	1.058386	-0.11902
15	18.30083	0.237855	5.472621	-0.08888	-0.65058	-4.87962	-0.05421	1.185197

	C	X1	X2	X3	X4	X5	X6	X7
CKKNIFE EFFICIENT	-4.15622	3.985866	1.038859	-0.00394	-0.41211	-2.58426	0.996386	0.564339
ANDARD ERROR	3.19244	2.49679	0.691065	0.433153	0.501944	0.492453	1.35355	0.130786
MPUTED T-VALUE	-1.30189	1.596396	1.503273	-0.00909	-0.82103	-5.24773	0.736128	4.314983
ITICAL T(0.975,14)	2.145	2.145	2.145	2.145	2.145	2.145	2.145	2.145
CISION	NS	NS	NS	NS	NS	S	NS	S

NOTE: S represents SIGNIFICANT at 97.5% confidence level
NS represents NOT SIGNIFICANT at 97.5% confidence level

APPENDIX 9

U-METHOD RESULTS

GROUP1	GROUP2	GROUP3
-0.3884	-1.4309	-0.2098
-0.3884	-0.3884	-0.2098
0.0361	-0.2996	2.9465
-1.7832	-0.3884	0.5551
-0.3884	-0.3884	-0.1528
1.1631	0.2038	-0.2098
-0.3884	-1.0415	2.3942
1.0415	-0.4536	0.9387
-2.3147	-0.8059	-0.0361
1.6616	0.8569	1.4996
-0.51	0.0781	2.5504
-0.2098	-0.8108	0.3824
0.3314	-0.8059	-2.4937
-0.3884	1.2092	-2.4937
1.6616	3.6161	0.938950

CUTTING SCORE:

PROCEED DIRECTLY less than or equal to -0.6184
PROCEED AFTER SUPP between -0.6184 and 0.4142
DISCONTINUE greater than 0.4142

MISCLASSIFICATION ERROR RATE = 0.444

BIBLIOGRAPHY

Bruce Cooil, et al:"Cross-Validation for Prediction", Journal of Marketing Research, Aug.1987., pp.271.

Chirchir, M.K. :" A Discriminant Model To Distinguish Between Successful Accounting And Non-Accounting Students In The Faculty Of Commerce Of The University Of Nairobi", Unpublished M.B.A. Project , Faculty Of Commerce, University Of Nairobi.

Efron, Bradley : "Improvement on Cross-Validation", Journal Of The American Statistical Association, 78 June, 1983, pp 316-31

Farrar D.E and Glauber R.R. : Multicollinearity in Regression Analysis Rev. Economics and Statistics, vol.49,1967, pp. 92-107

Finlay, N.P et al:" The Validation of Expert System", Journal of The Operational Research Society Vol.

Finlay N.P. and Wilson M.J. :" The Paucity of Model Validation In Operational Research Projects", Journal of The Operational Research Society Vol.38, No.4, pp 303-308, 1987

Frank, R.E., et al. "Bias in Multiple Discriminant Analysis" Journal of Marketing Research, 2(aug.1965), 250-8

Fujikoshi, Yasunori :"Selection of Variables in Discriminant Analysis and Canonical Correlation Analysis" in

Multivariate Analysis IV -Proceedings of the sixth International Symposium on Multivariate Analysis,(edited by Krishnaiah, P.R.)

Goldfeld, S.M. and Quandt, R.E. : " Some Tests for Homoscedasticity" Journal of The American Statistical Association Vol.60, 1965, pp.539-47.

Harris, J.R. :A Primer of Multivariate Statistic , Academic Press, 1981, pp.87.

Hawkins, M.D. :Topics in Applied Multivariate Analysis, Cambridge University Press, 1982 pp. 1-52.

Kendall, M.G.: Multivariate Analysis, Griffin, London, 1975..

Kiilu, P.N. " An Evaluation of The Validity of Regression and Correlation Analysis Among Some Selected M.B.A. Projects, Faculty Of Commerce, University Of Nairobi." Unpublished M.B.A Project, University of Nairobi 1990.

Koutsoyiannis, A. Theory of Econometrics. Macmillan Publishers Ltd. Hong Kong,1977.

McCabe, G.B. :"Computations for variable selection in Discriminant Analysis; Technometrics , 17, 1975. pp.103-109.

Mosier Charles I (1951),Problems and Designs of Cross-Validation." Educational and Psychological Measurement, 11 (spring), 5-11

Nderitu, M.M.:"An Evaluation Of The Validity Of Regression and Correlation Analysis Among Some Selected M.B.A. Projects, Faculty Of Commerce, University Of Nairobi." Unpublished M.B.A. Project, UNiversity Of Nairobi, 1990.

Wasserman, W, et al, Applied Linear Statistical Models.,
Irwin,
Homewood, Illinois, 1985,pp392

Wittink, D.R., The Application of Regression Analysis. Allyn and Bacon, Inc.Boston, 1988, pp176.