A comparative study of multiple discriminant analysis and multinomial logistic regression applied to students' performance

T.S Nwanamidwa

orcid.org/0000-0003-1114-8097

Dissertation submitted in fulfilment of the requirements for the degree *Master of Statistics* at the North-West University

Supervisor: Prof N.D Moroke

Graduation April 2018

Student number: 16132661

NWU®
NORTH-WEST UNIVERSITY
NOORDWES-UNIVERSITEIT
UNIBESITI YA BOKONE-BOPHIRIMA

DECLARATION

I declare that the title *"A comparative Study of Multiple Discriminant Analysis and Multinomial Logistic Regression on Students Performance"* towards the award of the M.Com degree is my own work, that it has not been submitted for any degree or examination in any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full names.. Date....................................

Signed...

Signature Supervisor... Date.....................................

ACKNOWLEDGEMENTS

I would like to thank God for making sure that I had the courage to see this through. My supervisor Prof N.D Moroke, without your guidance and patience, I wouldn't have made it. I really appreciate all the words of encouragement and the wisdom you shared with me along this journey. I will be forever grateful. To all my colleagues who embarked on this quest with me, your assistance and advice came in handy when required. For that I am thankful. And lastly, to my family, thanks for the support you gave me, for making sure that my kids are well taken care of while I was busy with this study.

ABSTRACT

This study compared the performance of two of the most recommended statistical techniques in classification. The study sought to determine the classification accuracy of Multiple Discriminant Analysis (MDA) and Multinomial Logistic regression (MLR) by classifying the students based on their performances on the modules offered in Statistics Department at the North West University. The data used comprised of the performance of the third year students who majored in Statistics from 2013 and 2016 academic years. The preliminary analysis were performed to evaluate the descriptive statistics and test for assumptions. In order to achieve the MDA results, the Wilk's lambda was used to test the significance of the model. The canonical discriminant function coefficients were utilised to construct a discriminant model while the classification table was used to generate the overall classification of the students' performances. The model fitting table MLR showed that the model was useful. The deviance and the Pearson's statistic were used to indicate if the model fits the data or not and the parameter estimates were used to obtain the fitted model for MLR. The results from the classifications showed that the MDA was able to classify 58.3% and the MLR was able to classify 56.7%. These results showed that the techniques do not differ much in classification. Therefore both the techniques can be used for future studies in education.

Keywords: Multiple Discriminant Analysis, Multinomial Logistic Regression, students' Performance.

Table of Contents

List of Tables

List of Figures

Acronyms

ACCR	Apparent correct classification rate
ACT	American college test
AER	Apparent error rate
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AUROC	Area under receiver operating characteristics
BIC	Bayesian Information Criterion
CART	Classification and Regression Trees
CRL	Coefficient of Racial Likeness
CRT	Classification and regression tree analysis
D	Deviance
DA	Discriminant Analysis
EFO	Ezemvelo Farmers' Organisation
FCS	Fully conditional specification
GPA	Grade point average
KZN	KwaZulu-Natal
LDA	Linear discriminant analysis
LR	Logistic Regression
LRM	Logistic regression model
MANOVA	Multivariate Analysis of Variance
MDA	Multiple Discriminant Analysis
MLE	Maximum Likelihood Estimation

MLR	Multinomial Logistic Regression
NMT	New Mexico Tech
NN	Neural networks
OLS	Ordinary Least Squares
P-P plot	Probability-probability plot
Q-Q plot	Quantile-quantile plot
RBF	Radial Basis Function
RBS	Radial basis system
SA	South Africa
SEM	Structural equation modelling
SEN	Sensitivity
SMOTE	Over, Under & Synthetic Minority Over-Sampling
SPE	Specificity
SPSS	Statistical Package for the Social Sciences
TOL	Tolerance
VIF	Variance inflation factor

CHAPTER 1

STUDY ORIENTATION

1.1. Introductory background

A categorical variable has a measurement scale consisting of a set of categories. The development of methods for categorical variables was stimulated by the need to analyze data generated in research studies in both the social and biomedical sciences. Many categorical variables have only two categories often given the generic labels success and failure. Agresti (2013) calls these binary variables. In selecting a more relevant analytical technique, one occasionally encounters difficulties that include a categorical dependent variable and several metric independent variables. The most widely statistical methods used for analysing categorical outcome variables include linear discriminant analysis and logistic regression (Pohar, et al., 2004; El-habil & El-Jazzar, 2014).

Discriminant analysis (DA) provides a powerful technique for examining differences between two or more groups of objects with respect to the several variables that have to be factored in simultaneously. Although the initial study of DA involved applications in the biological and medical sciences, considerable interest was aroused by statisticians in areas of study such as business, education, engineering and psychology. This has led to the writing of textbooks that cover DA from various perspectives (Huberty & Olejnik, 2006). DA is used to classify an observation into one or several of a priori groupings dependent upon the individual characteristics (Gwary et al., 2012).

One other well-known method used for analysing a categorical dependent variable is logistic regression (LR). The use of LR in biomedical studies was to determine whether or not subjects have particular condition. The technique has since been used in social science for modeling opinions and behavioural decisions as well as in business areas. LR analysis is the most popular technique available for modeling dichotomous (binary) dependent variables (Kleinbaum et al., 1998). Multinomial Logistic Regression (MLR) (also referred to as polychotomous) is an extension of binomial logistic regression that allows for more than two categories of the dependent or outcome variable. MLR is often used in analysing categorical response data with continuous or categorical explanatory variables (Bull & Donner, 1987). This technique is used when the dependent variable has more than two nominal or unordered categories and dummy coding of dependent variables is very common. The logistic model assumes that data in each independent variable has a single value for each case (Sanabila et al. 2010).

DA and LR, according to Hair et al. (2010), are the statistical techniques most suitable when the dependent variable is categorical (nominal or nonmetric) and the independent variables are metric variables. Consequently, DA is capable of handling either two groups or more. When three or more classifications are identified, the technique is called Multiple Discriminant Analysis (MDA) and this study focused on MDA and MLR. Maiprasert & Kitbumrungat (2012) caution about lots of assumptions associated with DA. Some of the most important assumptions include interrelationships of groups and linearity in predictive variables. This means that predictive variables should be spread together in a dichotomous dimension with less interrelationship to prevent a dichotomous linearity.

In order to influence the nature of the research questions being posed, Mertler & Vannatta (2002) mention that one needs to choose an appropriate statistical technique. The choice of a technique is dependent on whether the variable is categorical or quantitative. The number of dependent and independent variables chosen also plays a significant role in the selection of the appropriate statistical technique. Techniques like multiple regression and simple correlation measures have been used in education institutions to study the retention of students, enrollment management and performance prediction in determining whether the students might graduate or not. According to Kabakchieva (2013), universities operate in a very complex and highly competitive environment and the challenge with these modern universities is to intensely analyse their performance, to identify their uniqueness and to build a strategy for further development and future actions. Correlation analysis is one of the few techniques that neither provide information about the relative properties of multiple measures nor provide simple means of evaluating numerous variables when permitted to work in combination.

Multiple regression analysis is built on the assumption that dependent variables are continuous in nature. This implies that the condition under which this technique is applied is erroneous since most variables in educational studies are not continuous (Thomas et al., 1996). This study applies MDA and MLR, as suggested by Lewicki & Hill (2006), to predict whether students could proceed with their post graduate studies or not. The variable of interest is non-metric and the use of these techniques is applicable only for this instance. These techniques are suitable for predicting categorical outcome variables and are widely adopted due to their flexibility for describing relationships between multiple independent variables and a single dependent variable (Prempeh, 2009). Hosmer & Lemeshow (2000) warn against some of the linear models such as multiple regression analysis due to their limitations and applications in only some limited situations. This study focused on the MDA and MLR, examining the main attributes or group differences that could affect the performance of students.

1.2. Significance of the study

There is paucity in literature on the comparison of MDA and MLR in predicting student's performance. Therefore this research adds to the few existing studies in the field of education. This could stimulate interest on the application of these techniques in other study areas. The findings of this study might also be useful to administrators in planning ahead for enrolment, retention and could be used to help in recruitment of students and also predict more accurately the characteristics of students likely to enroll in their institutions. The outcome could also help in estimating the number of students to expect at the next higher level as the students successfully complete studies in the current year of study. At-risk students are also likely to be more accurately identified and such estimations are likely to help in the initiation and implementation of effective strategies to help them with their studies.

1.3. Problem Statement

There is increasing research interest in using different linear models in education for predictive analysis. The problem associated with some of these models is the fact that they are incapable of handling categorical dependent variables such as whether the student will pass or fail. In this kind of situation, it becomes inappropriate to use ordinary linear models in determining a linear relationship between a categorical dependent variable of interest and the risk factors associated. Making predictions of the categorical outcome variable with such models is also inappropriate as such models are assumed efficient only in cases where the linearity assumption is anticipated.

Techniques such as the MDA and MLR have been developed to handle such situations and have been widely used by scholars in various fields. The proposed techniques in this study share many common similarities and can handle the categorical dependent variable and complex independent variables such as those analysed in this study. It should be noted that most of the researches done in comparing MDA and MLR focused on social sciences and medical research. There is still a gap in utilising the two techniques in educational research. A lot of literature concentrates more on the techniques individually, but this study focuses on a comparison of the techniques in predicting the performance of students.

1.4. Objectives of the study

The study addressed the following set of objectives in order to:

- Fit MLR and MDA of students' performance
- Explore the efficiency of MLR and MDA in determining and predicting students' performance.
- Identify the model with more predictive power.
- Use the findings of the study in formulating recommendations for further studies and policy purposes.

1.5. Organization of the study

This study consists of five chapters which are presented as follows:

Chapter 1: The initial chapter provided the introductory background to the study and defined the problem statement, demonstrating the relevance of the study. The rest of the chapter defined the aim and objectives of the study and culminated in outlining the significance of the study.

Chapter 2: This chapter presents the review of the theoretical framework of the MDA and MLR techniques and reviews the literature of studies done previously where the two techniques were used for predictive purposes. The review also provides motivation why this study is conducted.

Chapter 3: The chapter provides and justifies the methodology adopted for this study. It consists of the detailed theory of statistical tests that are used to analyse data.

Chapter 4: This chapter reports on the data analysis and presents the statistical results of both the MDA and MLR.

Chapter 5: This chapter discusses the findings of the study. It also provides the conclusions to the whole study and provides recommendations for future studies and policy planning for educational institutions.

1.6. Summary

This chapter introduced the study by providing a background to the study, motivation for and significance of the study. The objectives were outlined on the basis of the research problem identified. The next chapter outlines the theoretical framework of MDA and MLR and empirical review of the two predictive methods.

CHAPTER TWO

THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1. Introduction

This chapter covers some of the theoretical framework of DA and LR and reviews studies that have been carried out by researchers in different fields of study. It aims at identifying areas that DA and LR have been applied. Section 2.2, 2.3 and 2.4 cover the theoretical framework and background of both DA and LR while Section 2.5 reviews antecedent literature.

2.2. Theoretical framework

Linear discriminant analysis and LR are two of the most popular methodologies for solving classification problems involving dichotomous class variables (Yarnold et al., 1994, Balogun et al., 2015). Both methods have widespread application in situations in which the primary objective is to identify the group to which an object belongs. The following segments discuss the theoretical background for the two proposed techniques.

2.3. Discriminant Analysis

2.3.1. Brief History of DA

According to Huberty & Olejnik (2006), the ideas associated with DA go back to around 1920 where an English statistician, Karl Pearson (1857-1936), proposed what was called the coefficient of racial likeness (CRL) which was a type of intergroup distance index. The CRL was then studied extensively by Morant (1899-1964) in that year. Again, in that year, a study of another distance index was initiated in India, which was then formalized by P. C. Mahalanobis (1893-1972) later in the 1930s. Following that, R.A Fisher translated the idea of multivariable intergroup distance to that of a linear composite of variables derived for the purposes of two-group classification. Then in 1935, M. M. Barnard applied two-group (predictive) discriminant analysis involving seven Egyptian skull characteristics. The extension of two-group classification to multiple groups was given by C.R. Rao in 1948. Many other extensions and refinements of Fisher's ideas have appeared since the 1940s and methodologists from Harvard University played a vital role in the application of DA in education and psychology in the 1950s and 1960s. According to Huberty & Olejnik (2006), the application of DA since its inception has been utilized in various fields of study.

2.3.2. Overview of DA

According to Abledu et al. (2016), Sen (2010), DA is a statistical technique for predicting and categorizing a set of observations into predefined classes. This technique is used to determine which continuous variables best discriminate between two or more naturally occurring groups. It is regarded as multivariate analysis of variance (MANOVA) turned around and the two share many of the same assumptions and tests. According to Tabachnick & Fidell (2007), in MANOVA one may ask whether or not group membership is associated with statistically significant mean differences on a combination of dependent variables. If the answer is yes, then the combination of variables can be used to predict group membership. DA differs from MANOVA because the latter considers groups as independent variables and the dependent variables are the predictors. However, the variables are interchanged to suit DA settings.

DA focuses on the association between multiple independent variables and a categorical dependent variable by forming a composite of the independent variables. This type of multivariate analysis can determine the extent of any of the composite variables which discriminates between two or more pre-existing groups of subjects. The multivariate analysis can also derive a classification model for predicting the group membership of new observations (Antonogeorgos et al. 2009).

According to Lattin et al. (2003), the purpose of DA in any study is to use the information from the independent variables in order to achieve the clearest possible separation or discrimination between or among groups. There are two types of DA namely: linear discriminant and multiple discriminant analysis. Antonogeorgos et al. (2009) recommends linear discriminant analysis to determine which set of variables discriminates between two naturally occurring groups and to classify an observation into these known groups. Multiple discriminant analysis is a more complex version which is used when three or more classifications are involved in the analysis. Hair et al. (2010) suggest that discrimination is achieved by calculating the variates weight for each independent variable to maximize the differences between groups (i.e. the between-group variance relative to the within-group variance).

In general form, the discriminant function is represented by the following equation:

$$Z = a + W_1X_{1k} + W_2X_{2k} + \cdots W_nX_{nk} \qquad (2.1)$$

where, Z = discriminant score (dependent variable); a = constant;

W_i $(i = 1,2, \dots n)$ = discriminant weights or coefficients; X_{ik} $(i = 1,2, \dots n), j$ = independent variables or predictive variables

The discriminant function analysis is responsible for calculating coefficients that maximises the between-groups variance, given the within-groups variance. The centroids are known as the average of weights of all cases within one group and DA is used to test the null hypothesis that these group means are equal (Sen, 2010). The centroids indicate the most typical location of any member from a particular group, and a comparison of the group centroids shows how far apart the groups are in terms of that discriminant function.

According to Hair et al. (2010), DA can be considered either a type of profile analysis or an analytical predictive technique. This is because DA mainly provides an objective assessment of differences between groups on a set of independent variables. This technique is also appropriate in situations with a single categorical dependent variable and several metrically scaled independent variables.

2.3.3. Theoretical Structure of DA

According to Hair et al. (2010), the application of DA can be viewed from the six stage model building perspective. The following diagram shows the stages that need to be covered in carrying out the DA:

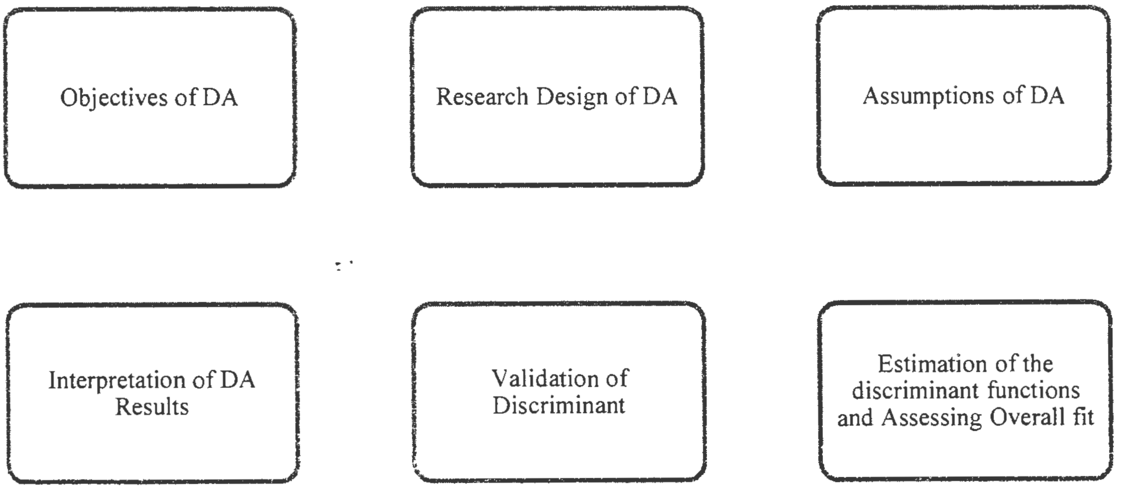

Figure 2.1 Framework for DA, adopted from Hair et al. (2010)

The discussion that follows is in line with Figure 2.1. Each of the steps is discussed accordingly.

2.3.4. Objectives of DA

Tabachnick & Fidell (2007) and Hair et al. (2010) recommended what is called a decision-tree for most of the multivariate techniques. The process lessens the time taken by an analyst and further helps in deriving lōgical sense out of the analysis. Theoretically, according to Gabor et al. (2011), Al-Jazzar (2012) and Prempeh (2009), DA sought to achieve the following objectives:

- Building objectives: building the discriminant functions, represented as linear combinations of independent variables (explanatory or predictor variables) that best discriminate between categories of dependent variables.
- Identify variables that contribute most to explaining the differences between groups. Undertake discriminant classification analysis which classifies cases into groups by allocation to a certain group, according to values of independent variables.
- Examine whether or not significant differences exist among groups.
- Determination of linear correlations of predictor variables to separate the groups by maximization of inter-group variations compared to intra-group variation.

- The relative significance of the independent variables is assessed by classifying the dependent variable.
- Evaluation of classification precision.
- Developing procedures to set new objects whose profile is known but the group they belong to is not known.

2.3.5. Research Design of DA

A successful application of DA, according to Hair et al. (2010), requires consideration of numerous issues. These issues involve the selection of dependent and independent variables, the sample size required for estimation of the discriminant functions and the division of the sample for validation purposes.

Selection of dependent and independent variables.

According to Sen (2010), the number of dependent variable groups (categories) can be two or more. These groups must be mutually exclusive and exhaustive. When three or more categories are created, the possibility of examining only the extreme groups in a two group DA arises. Therefore the polar-extreme approach is conducted and this involves comparing only the extreme two groups and excluding the middle group from the DA. The independent variables in research design are selected by identifying variables from previous research or from intuition where no previous research exists.

According to Sen (2010) and Hair et al. (2010), DA like any other multivariate technique, is very sensitive to the proportion of sample size and the number of predictor variables. Therefore a more reliable DA according to Sen (2010), is when the sample size is greater than 30 for each independent variable in the analysis. At a minimum, the smallest group must exceed the number of independent variables.

Division of the sample

Division of the sample is a preferred means of validating a DA. In order to estimate the discriminant function and validate, the sample can be divided by creating sub-samples. This sample is then used to develop the discriminant function. The holdout sample is used to test the discriminant function (also referred to as the split-sample validation or cross validation).

2.3.7. Assumptions of DA

In any statistical analysis, all parametric tests assume some certain characteristics about the data. This includes testing assumptions and if these assumptions are violated, then the results could change the conclusion of the research and interpretation of the results. Like any other statistical technique, DA is often guided by the apparent model assumptions. Lei & Koehly (2003) outlined that when the DA assumptions are satisfied, the technique is recommended because its application has been warranted and the classification accuracy is likely to become more precise. Furthermore, Mihalovic (2016) mentions that testing of discriminant model assumptions might indicate whether or not the discriminant model is usable. If these assumptions are violated, then logistic regression becomes the most preferred technique. Therefore DA performs very well under certain assumptions such as those listed below.

Multivariate normality – this test checks a given set of data for similarity to the multivariate normal distribution. The independent variables are assumed to follow a multivariate normal distribution, thus allowing only continuous or ratio variables to enter the analysis and excluding all forms of categorical variables. The simplest method of assessing normality is by producing a histogram. The quantile- quantile also known as normal Q-Q plot can also be used to assess the normality of a distribution (Steven, 2001, Mertler & Vannatta, 2013). It is also possible to apply the Kolmogorov-Smirnov test if a sample is greater than 50. The convention is that a significant value greater than 0.05 indicates normality of the distribution (Balogun, 2015, Normadiah & Yap (2011), Garson, 2012).

Homogeneity of variance-covariance matrices - According to Mihalovic (2016) and Balogun (2015), this assumption verified with the Box's M test determines if two or more covariance matrices are equal. According to Burtner (2005), the Box's M test is used to verify the null hypothesis of equal population covariance matrices. Another way to test the covariance matrices is to do an eye-ball technique. This technique is used to compare groups by the categorical dependent variable and create a scatterplot matrix for all the independent variables. The scatterplots are then compared across the groups for the same independent variables (Garson, 2012). This test is only possible in solutions resulting in more than one discriminant function. Mertler & Vannatta (2013) mention that Tabachnick &Fidell (2007) confirmed that rough equality in the overall size of the scatterplots provides evidence of homogeneity of variance-covariance matrices.

Linearity - The discriminant model assumes linear relationships among all pairs of predictors within each group. This can be achieved by plotting the standardised residuals against standardised estimates (fitted values) of the dependent variable (Garson, 2012). This assumption, according to Mertler & Vannatta (2013), is best assessed through an inspection of bivariate scatterplots. The shape of the scatterplot is oblique if both variables in the pair of predictors for each group on the dependent variable are normally distributed and linearly related. The relationship is not considered to be linear if one of the variables is not normally distributed. The scatterplot between the two variables will not appear oval-shaped. According to Tabachnick & Fidell (2007), violation of the linearity assumption is often seen as less serious than violations of other assumptions in that they tend to lead to reduced power as opposed to an inflated Type I error.

2.3.8. Estimation of the Discriminant Functions

Estimation of DA can be done using either simultaneous or stepwise method. Simultaneous estimation involves computing the discriminant function so that all independent variables are considered concurrently. The simultaneous estimation method is only appropriate when the researcher wants to include all independent variables in the analysis and is not interested in seeing intermediate results. On the other hand, stepwise estimation involves entering the independent variables into the discriminant function one at a time on the basis of their discriminating power. This estimation method can be useful when the researcher wants to consider a relatively large number of independent variables for inclusion in the function. The Wilks lambda, Hoteling's trace and Pillai's criteria are all measures that evaluate the statistical significance of the discriminatory power of the discriminant functions (Hair et al. 2010, Tabachnick & Fidell, 2007, Johnson & Wichern, 2007, Rencher, 2002).

2.3.9. Assessing overall model fit

The assessment of model fit involves three tasks: the tasks include calculating discriminant Z scores for each observation, evaluating group differences on the discriminant Z scores and assessing group membership prediction. In order to achieve the three tasks above, the researcher needs to address the following: rationale for classification matrices, cutting score determination, costs of misclassification, creating classification matrices, and assessing classification accuracy and casewise diagnostics (Hair et al., 2010).

2.3.10. Interpretation of DA results

There are several approaches to the interpretation of linear discriminant functions. Hair et al. (2010) mention that the interpretation of results includes inspection of the discriminant functions to determine the relative significance of each independent variable in classifying between groups. This is achieved by standardised discriminant weights, discriminant loadings and partial F values. The discriminant score describes the position of a case in the discriminant space along the axis defined by the discriminant function. The scores can be interpreted as measures of the distance between the grand centroid and each particular case. The unstandardized discriminant weights indicate the absolute contribution of each predictor to the value of the discriminant score. According to Kislaya (2012), in order to assess relative importance of discriminating variables, the standardised discriminant function coefficients should be analysed. These standardised coefficients are stated in the following way:

$u^{s}_{ki} = u_{ki} \cdot \sqrt{\frac{w_{ii}}{n-g}}, \quad i = 1,2,\ldots,p$ where w_{ii} is the i-th main diagonal element of the covariance matrix.

According to Huberty (2006), the use of standardised discriminant function weights to assess the relative importance of a covariate has serious limitations. Klecka (1980) mentions that if there is multicollinearity, the standardised coefficients may hint at misleading conclusions. This is because these coefficients take into account the joint contribution of all variables. Huberty (2006) and Klecka (1980) suggests that the alternative in judging the relative importance of variables in linear discriminant analysis is through the correlation between each discriminating variable and the different discriminant functions. The structure coefficients show how a particular variable and discriminant function are related. The high loading reveals that the variable shares the most variation with a given function, but a loading close to zero tells that the variable and function have less in common. Kislaya (2012) cautions that this structure has been widely criticised because they fail to provide multivariate information. Therefore Rencher (2002) recommends the use of standardised coefficients in the interpretation of discriminant functions because they are able to assess the joint contributions of the discriminating variables.

2.3.11. Validation of discriminant results

The validation of results is a critical step that comprises validating the discriminant results to provide assertions that the results have external as well as internal validity. According to Frank et al. (1965), Montgomery, (1975) & Morrison, (1969), several alternatives for validation of results have been introduced. The holdout method is one that has been the most frequently suggested validation approach and is appropriate when the data is large. This approach is used when the sample is randomly split (Crask & Perreault, 1977). The one sub-sample is used to develop estimates of the discriminant coefficients and these coefficients are then applied to the observations in the other sub-sample for classification purposes. However, Frank et al. (1965) suggests the Monte Carlo simulations as a mechanism for evaluating discriminant results. This approach works when synthetic data is generated and discriminant functions are derived with the same degree of freedom as the original data. According to Hair et al., (2010), the most common approach for validation is by assessing hit ratios. This can be achieved by using either a separate sample (holdout sample) as explained earlier by Crask & Perreault (1977) or a cross-validation procedure using the jackknife approach.

2.4. Logistic Regression

2.4.1. Brief history of Logistic Regression

According to Cramer (2002), the logistic function was conceived in the 19^{th} century to describe the growth of populations and the course of autocatalytic chemical reactions. The population growth was described easiest by exponential growth but led to impossible values. Therefore the logistic function was then the solution to a differential equation that was examined from trying to dampen exponential population growth models. Quetelet & Verhulst showed how the logistic models agreed very well with the actual course of the populations of France, Belgium, Essex, and Russia for periods up to the early 1830s.

Professor David R. Cox then later in 1958 developed logistic regression (although much work was done in the single independent variable case almost two decades earlier). The author is widely acknowledged as among the most important scientists of the second half of the twentieth century. He inherited the mantle of statistical science from Pearson and Fisher, advanced their ideas, and translated statistical theory into practice and forever changed the application of statistics in many fields, but especially biology and medicine. The logistic and proportional hazards models that Professor Sir David R. Cox substantially developed are arguably among the most influential biostatistical methods in current practice (Zeger et al., 2004). This technique was introduced as an alternative to Ordinary Least Squares (OLS) and has found wide application in statistical software programs (Memic, 2015).

2.4.2. Overview of Logistic Regression

Logistic Regression, sometimes called the logistic model or logit model, analyses the relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data into a logistic curve (Hyeoun-Ae., 2013).

There are two types of logistic regression, binary logistic regression and multinomial logistic regression also known as polychotomous. The binary logistic regression is typically used when the dependent variable is dichotomous and the independent variables are continuous or categorical. When the dependent variable is not dichotomous and comprises of more than two categories, a multinomial logistic regression can be employed.

LR is thought of when the analyst intends to provide predictions of group membership. LR calculates the probability of success over the probability of failure and therefore the results of the analysis are presented in the form of an odds ratio. This technique also provides knowledge of the relationships between dependent and independent variables and the strengths among these variables.

In logistic regression, a mathematical model of a set of explanatory variables is used to predict a transformation of the dependent variable. This is called the logit transformation (NCSS.com). Adopting Agresti (2013), mathematically, the logit transformation for multinomial logistic regression is written below:

$$l = logit[\pi(\boldsymbol{x})] = ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \qquad (2.2)$$

Also written in probability form:

$$\pi(\boldsymbol{x}) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)} \qquad (2.3)$$

where

$\left(\frac{p}{1-p}\right)$ = the ratio of the probability of success to the probability of failure;

$\beta_1, \beta_2, \ldots \beta_p$= parameter estimates or logistic regression coefficients;

α = constant; $x_1, x_2, \ldots, x_p$ = independent variables

$\pi(\boldsymbol{x})$ = the response probability.

The predicted values in LR lie between 0 and 1 regardless of the values of the explanatory variables (Suleiman et al., 2014). According to Sen (2010), the LR estimates the odds of a certain event occurring and calculates changes in the log odds of the dependent variable, not changes in the dependent variable itself as ordinary least squares(OLS) regression does. Furthermore, Izenman (2013) explains that the log odds ratio involves the ratio of two odds and a summary measure of the relationship between two variables. The use of the log odds ratio provides a more simplistic description of the probabilistic relationship of the variables and the outcome in comparison to a linear regression whereby linear relationships and useful information can be drawn.

According to Kasper (2006), LR does not assume linearity of relationship between the independent and the dependent variable. The relationship between the dependent and independent variable may be linear or non-linear. LR can handle all sorts of relationships because it applies a non-linear log transformation to the predicted odds ratio. Although the multivariate normality yields a more stable solution, the independent variables in LR do not need to be multivariate normal and also the error terms do not need to be multivariate normally distributed. Furthermore, LR does not need variances to be heteroscedastic for each level of the independent variables. The technique is capable of handling ordinal and nominal data as independent variables. The independent variables do not need to be metric (interval or ratio scaled).

2.4.3. Assumptions of Logistic Regression

LR is one technique which does not face strict assumptions and is considered to be more robust when these assumptions are not met. However, the following have to be adhered to when using this technique:

Mutually exclusive and exhaustive: the groups (categories) must be mutually exclusive and exhaustive. This simply means that a case can only be in one group and every case must be a member of one of the groups. The model should be fitted correctly. Neither over fitting nor under fitting should occur i.e. only the meaningful variables should be included. A good approach to ensure this is to use a stepwise method to estimate the LR. The error terms need to be independent and each observation is required to be independent.

Multicollinearity: according to Mertler & Vannatta (2013), LR is sensitive to high correlations among predictor variables which can result in multicollinearity. Therefore, the model should have little or no multicollinearity. If multicollinearity is present then centering the variables might resolve the issue, i.e. deducting the mean of each variable. If this does not lower the multicollinearity, a factor analysis with orthogonally rotated factors should be done before the LR is estimated.

Linearity: LR assumes linearity of independent variables and log odds. Whilst this technique does not require the dependent and independent variables to be related linearly, it requires that the independent variables are linearly related to the log odds.

Sample size: Larger samples are needed than for linear regression because maximum likelihood coefficients are large sample estimates. A minimum of 50 cases per predictor is recommended (Burns & Burns, 2008).

2.4.4. Assessment of Logistic Regression

The success of the LR can be assessed by viewing the classification table that shows the classifications of the dichotomous, ordinal and polychotomous dependent variables. The goodness-of-fit tests such as the Chi-square are available as indicators of model appropriateness and the Wald statistic to test the significance of individual independent variables (Kasper, 2006; Abledu et al., 2016). The classification table in this instance applies to the generated regression model to predicting group membership. These predictions are then compared to the actual values and the percentage of participants correctly classified is calculated and serves as an indicator of the model fit (Mihalovic, 2016). The Wald statistic serves as a measure of significance for β and represents the significance of each variable in its ability to contribute to the model. This statistic is quite sensitive therefore a more liberal significance level ($p < 0.05$ or $p < 0.1$) should be applied when interpreting this value (Tabachnick& Fidell, 2007; Mertler & Vannatta, 2013). The goodness-of-fit test is one of the tests that shows how well a model fits a set of observations. This measure typically outlines the discrepancies between observed values and the values expected under the model in question (Wu, 2010).

2.5. Literature review

This section provides a review of literature around the application of DA and LR. The section further identifies areas in which the DA and LR have been applied. Important note should be made that the relevance of these techniques is not limited only to the field of statistics as already explained in the previous section. They are applicable to all areas of studies which intend to achieve the objectives listed in the previous sections. It is evident that these techniques have been widely concentrated in medical research. Reviewed below are some of the studies which adopted the DA and LR models.

Balogun et al. (2015) did a comparative study of the two techniques to predict mode of delivery for expectant mothers. The study examined 184 cases of the mothers giving natural birth or delivering by a caesarian section. The results from the DA classified mothers who had natural birth to 64.6% and those who had a caesarian section to 64.7%. LR on the other hand classified 76.8% and recorded 52.9% of natural birth and caesarian birth respectively. The two techniques yielded similar results with a classification performance of nearly an equal value (i.e. 64.7% and 65.8%). The test for multivariate normality and equal covariance matrices of the DA were not met, therefore LR was preferred.

Furthermore, Balogun et al. (2014) also compared the two techniques using drug offenders' data. The study showed that LR recorded 92.4% and 97.9% success rate in classifying the drug peddlers and non-drug peddlers respectively while DA recorded 56.3% and 84.6%. The overall predictive performance of the two models was high with LR having the highest value of 95.4% and 71.8% of DA. Both the techniques selected the same set of variables. The results from the study showed that the multivariate normality test and that of equal covariance matrices of the DA was not met. Therefore the use of LR was preferred again in this study.

Again Balogun et al. (2012) used DA with the main aim of constructing a discriminant function that can classify the drug peddlers and non-drug peddlers in Kwara state of Nigeria. The variables that were used in the study were type of exhibit, age, weight of exhibit and gender. The results from the study showed that these variables contributed to building a discriminant function and the misclassification rate obtained was 28.2%.

In a more recent study, Elgohari (2017) used multivariate logistic regression (MLR) and linear discriminant analysis (LDA) for the detection of anemic children with chronic kidney disease. The missing data was remedied first by using fully conditional specification (FCS) and thereafter conducted the comparison of the MLR and LDA. The comparison depended on the apparent error rate (AER) and the apparent correct classification rate (ACCR). The study had 344 observations and was divided into two groups. The study results of this study showed that LDA was significantly more efficient than MLR.

Furthermore, Mahfouz (2016) used LR in the prediction of factors associated with Khat abuse among students in Jazan region in Saudi Arabia. The study's main focus was to determine contributing factors to students' Khat abuse. The researchers were mainly focused on the first middle school, secondary schools and second year at university level of Jazan region. The data was obtained from surveys conducted by the Jazan Substance Abuse Research Centre during 2011-2012 academic year. The three stage cluster random sampling technique was utilised for the selection process and a questionnaire was administered. The study consisted of 56.9% males and 43.1% females. The results from the study revealed that the variables which were significant involved the following factors: student's smoking status, gender, friend's use of tobacco and friend's use of Khat. The researchers emphasized the relevance of health education as a strategy for changing customs and practices in the Jazan population.

Mansour et al. (2013) explored the use of the two techniques by comparing them with Artificial Neural Network (ANN) in determining the risk factors of Type 2 diabetes with data collected from 17 rural health centers in Kermanshah city using a group of 100 diabetic and pre-diabetic patients. ROC curve was used to compare prediction powers of the models. The Radial Basis Function (RBF) and wrapper method were applied in an ANN model to specify a model with the highest prediction. Although the DA model indicated significant correlation, the results showed that the RBF method had a higher level of accuracy and sensitivity with LR on the other hand proving to be significantly powerful in distinguishing between diabetic and pre-diabetic patients. Therefore the application of the three methods was recommended for medical studies.

Al-Jazzar (2012) used and compared the predictive performance of MDA and MLR using diabetes dataset. Both methods converged on similar results, implying that a similar conclusion was drawn from them. The estimated coefficients were found to be both statistically significant and the classification of diabetic women was done successfully. However, the MLR was reported to be more significant in terms of the error rates associated with the classification of diabetics.

In their paper, Maiprasert & Kitbumrungat (2012) used Multinomial logistic regression (MLR) and DA on 680 samples of women in predicting the stages of breast cancer. The study showed that the MLR model had a classification of 55.50% higher than the DA model with 54.10%. Furthermore, the results showed that the overall classification of MLR was higher than DA classification. Therefore the authors concluded that MLR was more capable of correctly predicting the stages of breast cancer than DA.

Antonogeorgos et al. (2009) did a divergence and similarity study of the two techniques in evaluating factors associated with asthma prevalence among 10 to 12 year old children in Greece. The epidemiologic data used was from 700 students from 18 schools located in several areas of Athens. The results show that the overall correct classification rate for DA was 77.4% and 79.2% for LR. The ROC curve indicated that the logistic model was similar to the DA, hence the authors recommended the use of both techniques.

The study by Panagiotakos (2006) investigated whether or not DA and LR techniques result in similar findings when evaluating categorical health outcomes. The overall correct classification rate for the prediction of the in-hospital mortality of patients presented with acute coronary syndromes was found to be good for both methods. The results of both LR and DA were similar, even though the assumption of equal covariance was not conducted in this study. In conclusion, the two techniques brought about the same findings with a better correct classification rate.

Other studies such as Geller et al. (2009), Smith (2005), Betensky & Williams (2001), Takahashi et al. (1998), Ito et al. (1997), Clark et al. (1989), Gordon et al. (1984), Kennedy et al. (1980), effectively applied such techniques individually and comparatively in the clinical medicine. Important findings and conclusions were drawn from these studies.

Memic (2015) did a comparison of MDA and LR in a financial sector. The author conducted a study to predict credit default, creating a prediction model that could distinguish between defaulted and non-defaulted companies based on the financial data obtained from the financial statements of Bosnia and Herzegovina banking markets. The results showed that the models had high predictive ability and some variables seemed to be more influential on the default prediction than others. In conclusion, LR exhibited better predictive ability than MDA.

Mihalovic (2016) did a performance comparison of MDA and logit models in bankruptcy prediction. The study was based on a sample of 236 firms operating in Slovakia with the main focus on failed and non-failed firms. The variables that were most significant were impending firms, net income to total assets, current ratio and current liabilities to total assets. The results suggested that the LR model outperformed the classification accuracy of the discriminant model.

Suleiman et al. (2014) used the two techniques by factoring principal component analysis as a reducing agent. The data was collected from a sample of 200 applicants of which 163 were considered creditworthy based on the application forms from First Bank PLC in Nigeria. The results showed that using principal component analysis as a reducing agent provided more accurate results than original data and also decreased the model complexity. The authors used Box M test and Wilk's lambda to confirm the equality of covariance matrices and confirmed the significance of the canonical correlation respectively. Furthermore, the results showed that the LR gives slightly better results than DA. Therefore LR was the preferred model since it achieved less cost of misclassification in the Suleiman study.

A similar study to Suleiman et al. (2014) was undertaken by Kong-lai & Jing-jing, (2010). The study sought to establish the credit evaluation models of China's listed companies. The study's total sample included failed managed groups and normal-managed groups of 130 listed companies from 2009 in Shanghai and Shenzhen stock exchange. The principal component analysis was used to achieve dimensionality reduction and minimize the loss of information contained in the original data. The authors selected six principal components from 14 financial ratios and the cumulative contribution rate reached 85.401%. The study concluded that the LR and DA models proved good for investors when determining the credit risk of the investment subjects with the aim of reducing the risk. But due to the limited sample size, the DA assumptions were not met, leaving the LR as the superior model in this instance.

The use of data mining techniques has been widely applied in bank classification, mainly for the purpose of determining risk factors. Albayrak (2009) used classification and regression tree analysis (CRT), LDR and LR to develop a classification for predicting the group membership of domestic and foreign banks onto two pre-defined groups. The dataset used in this study involved 18 domestic and 14 foreign commercial banks operating in Turkey in the period of 2002-2006. The variables involved ratios based on the financial statements of banks. The results from the analysis revealed that CRT outperformed DA and LR analysis in terms of bank classification accuracy and thus provided an effective alternative for implementing bank classification tasks.

Sen (2010) conducted a study focusing on factors that discriminate between domestic and foreign banks operating in Turkey during 2009. The study was based on financial ratios and aimed at determining if there were any operational differences between these banks. The data used in this study was from 32 domestic and foreign commercial banks of which only 31 were used for the analysis. The results reported that both techniques yielded almost the same results.

Vuran (2009) used LR and DA to study the prediction of business failure from a sample of 122 publicly and open firms from 1999-2007. The aim of the study was to construct a bankruptcy prediction model, identify which financial ratios were of interest and comparing those prediction models. The two techniques were applied on a two-year period prior to failure. The results showed that the predictive accuracy of the discriminant model amounted to 84.4% at the first year and 80.1% at the second year before failure. The LR model achieved 84.4% at the first year and 82% at the second year. In conclusion, both the techniques had similar results and therefore confirmatory conclusions were drawn from them.

Another study regarding corporate failure was conducted by Court & Radloff (1990) in South Africa using MDA and LRA. The study was aimed to report on investigated public companies which had failed during the period 1965 and 1986. In this study, similar to the studies mentioned above, the use of principal component was to achieve dimensionality reduction and minimise the loss of information. Once the significant predictor variables were established, they were subjected to the MDA and LRA. The researchers stipulated that although there were shortcomings in the data and inconsistent predictions were obtained, they were able to remedy that by using estimated missing values for the results to be consistent. Therefore they concluded that LRA gives better results one year prior to failure.

Afolabi (2008) did a study using DA to conduct an analysis of loan repayment among small scale farmers in Oyo state, Nigeria. The study was aimed at identifying the socio-economic characteristics that discriminate between loan defaulters and non-defaulters using data collected from a cross section of small-scale farmers who borrowed for farming activities during 2002/2003. The results from the analysis showed that the discriminant function revealed six factors (age of farmers, gross income, non-farm income, net farm income, interest rate and farming experience) and these factors were significant in discriminating between defaulters and non-defaulters. Therefore the authors recommended that credit institutions or lending agencies should look out for the socio-economic characteristics that significantly discriminate between defaulters and non-defaulters before granting loans and advances to small scale farmers in order to reduce the incidence of loan delinquencies and defaults.

The use of DA and LR is also evident in performance analysis. This is where Haq et al. (2015) conducted a study to investigate classification performance of DA, LR and artificial neural networks (ANN) with radial basis system (RBS) in rationalizing the establishment of more private academies or not. The authors used a sample of 256 academies where 52% were metric, 34/% intermediate and 14.1% were graduate students. Exploratory factor analysis in this study was used with principal component analysis as precursors to the three techniques. The use of principal component analysis resulted in six extracted factors which explained 65.58% of total variation. The factors that were extracted consisted of poor education quality of public sector institutions, draw-backs of academy culture, career counseling in academies, competition among academies to survive, joining academies is a fashion of the day and benefits of academy culture. The results of the study showed that the classification performance of DA, LR and RBS was 96.5%, 96.1% and 96.5% respectively. Therefore the authors concluded that ANNs with RBS have slightly higher discriminative performance than DA and LR.

Similarly, Lin et al. (2009) used LR, neural networks (NN), DA and structural equation modeling (SEM) to predict first year university students' retention in engineering after one year. The authors developed twenty retention modeling systems of which five contained cognitive and non-cognitive factors. The study included 1 508 incoming first year engineering students during 2004-2005 academic year as participants. The prediction results from these twenty modeling systems show that among the four techniques used in this study, NN produced the best prediction results consistently and models combining both cognitive and non-cognitive data performed better than cognitive only or non-cognitive only models.

Lim et al. (1999) did a comparison study consisting of thirty-three classification algorithms with various data sets. The study employed the use of CART, LR, and both linear and quadratic discriminant analyses. This study required the researchers to empirically investigate the accuracy and the relative time needed to build each model. The data used in this study consisted of a total of thirty-two data sets and fourteen of these data sets were taken from real-life studies and two consisted of simulated data. The data sets ranged from 3,772 to 151 observations. These data sets were then doubled by adding noise to each of the original data sets. The results showed that amongst all thirty-three classification algorithms in this study, LR and LDA performed exceptionally well at correctly predicting class outcome. The two versions of CART performed marginally well and finally quadratic discriminant analysis performed very poorly in classification accuracy. Furthermore, it was reported that none of these algorithms had median running times in hours. Therefore, the LR had the longest median running time of four minutes whilst the other algorithms, CART and DA had a median running time of less than a minute. Despite the fact that the linear discriminant analysis requires the variables to be normally distributed, this technique performed well.

Luna (2000) used the same techniques predicting student retention and academic success in New Mexico. The objective of the study was to find the model that could best predict the elusive outcome of fall to fall persistence, to gain insight into the factors that lead to students' academic success and what might influence the students to remain at NMT (New Mexico Tech) or leave. The focus again was to find a clear answer to the question about the goal program's effectiveness. The main focus of the study was on the new, incoming freshmen. The authors iterated that the objectives above were not realised, therefore numerous variables were reviewed to observe the variables that give information about the retention and the academic outcome after three semesters. The dependent variables were useful in predicting the academic outcome. The study concluded and recommended that different variables were required to predict fall to fall persistence or some type of measure of persistence must be used in order to establish a student retention model.

In a more recent study, Teshnizi & Ayatollahi (2015) focused on the comparison of the LR model and artificial neural network (ANN) in predicting students' academic failure. The data used in this study was collected in 2013 using a stratified random sampling from 275 undergraduate students in schools of nursing, midwifery and paramedic schools in Hormozgan University of medical sciences. The results showed that among the nine ANNs, the ANN with the 15 neurons in hidden layer was better compared to LR. The area under receiver operating characteristics (AUROC) of the LR and ANN were estimated respectively. The ANN proved to be significantly greater than the LR. The LR model classified 77.5% accuracy while the ANN model classified 84.3% of the students correctly. Based on these results, it was apparent that the ANN model proved to be much better than the LR model.

Thammasiri et al. (2014) mentioned in their study that over the last decade, the use of comprehensive models was developed to address the student retention problem in higher education. This is apparent because the authors used the very same techniques in their study. They used LR, decision trees, NN and support vector machines to predict freshmen student attrition. The data used in this study was collected during the 2005-2011 academic years. The main objective of this study was to assess the efficiency of both the balancing techniques as well as the prediction models. The results of their study showed that the support vector machine combined with over, under & synthetic minority over-sampling(SMOTE) (data balancing technique) achieved the best classification performance. The overall accuracy was 90.24% on the 10-fold holdout sample. The LR, decision trees, NN and the support vector machines seem to have improved the prediction accuracy of the minority class and the variables that identify the prediction of student attrition. Therefore, the researchers concluded that the application of these models have the potential to accurately predict at-risk students and also assist with dropout rate of students.

Similarly, Schreiner (2009) conducted a study on student retention focusing on students' academic experiences. The study was focused on empirically linking student satisfaction to retention, arising from the postulation that there is some positive relationship between the two. The models focused on determining whether or not student satisfaction is predictive of retention the coming year. A LR analysis was conducted on each class level and the actual enrolment status was used as the dependent variable. The results of the analysis indicated that the first year students who did not find college enjoyable are 60% less likely to return as sophomores. Furthermore, the results showed that the students who lack a sense of belonging were 39% less likely to return and the ones with the difficulty of contacting their advisor were 17% less likely to return.

In a related research, Garton & Ball (2002) conducted a study on the academic performance and retention of college students of agriculture. The study was mainly focused on the students in the college of agriculture, food and natural resources at the University of Missouri. The stepwise method was adopted to build predictive models to determine the linear combination of student experiences along with their learning style, ACT(American college test) score, high school rank, and high school score GPA to determine the likelihood of persistence. The results showed that in the linear models, the high school score GPA was the best predictor of college academic performance for freshmen students. Furthermore, the learning style was not a significant predictor of students' academic performance during their first year of enrolment. Therefore, the traditional criteria used in predicting students' retention was found to have limited value.

Dey & Astin (1993) did a comparative study using logit, probit and linear regression. These techniques were used as alternatives for studying college student retention. The main focus of the study was to use logistic regression and linear regression to predict whether or not first-time, full-time community college freshmen who intend to earn a two-year degree would graduate on record time. The researchers also attempted to predict a reduced amount of stringent expectations of the students such as completing two years of college or being enrolled for a third consecutive fall semester upon admission. The variables in this study included students' concerns about ability to finance their education, motives for attending college, how many hours they spent per week at various activities on their first year, and high school grade point average. These variables were shown to strongly predict retention among students at four-year colleges and universities.

The study by Sule & Saporu (2015) used LR to investigate the factors that influence students' performance in calculus denoted as MTH101. The data used in this study consisted of the grades of the 200-400 level students in MTH101. This data was obtained from the department's examination records and a questionnaire was also administered to the students. The results from the Hosmer Lemeshow test showed that the model was a good fit and only three variables made a significant contribution to the predictive ability of the model. The results from the classification table indicated that the model was able to correctly classify 70.8% of the overall cases which was also an indication that the model was a good fit.

On the other hand, Tektas (2014) did a comparison of DA and LR in determining the rate of satisfaction with university provisions. The application data used in this study was collected using face-to-face surveys with 163 students attending Marmara University. The survey consisted of two sections: the first part were questions linked to the demographic characteristics of the students and the second part consisted of 52-Likert scale statements targeted at determining their level of satisfaction. The researcher also implemented factor analysis to determine the primary factors influencing the students' satisfaction levels. The results showed that DA not only enabled better prediction rates but also revealed the real relation between the variables better; therefore the researcher recommended the use DA for similar studies.

Kirschenbaum et al., (2000) did a comparative study to examine the differences between two groups of Israeli workers. The LR results provided hit ratio of 70% and this technique was identified as the better predictive model compared to the DA.

Aromolaran et al. (2013) conducted a study using binary LR on student's academic performance in Nigeria. The objective of the study was to determine the socio-economic factors influencing the performance of students and if an association existed between the demographic factors and economic factors. The study used non-random sampling and quota sampling method to administer 600 questionnaires to seven schools in the area. Statistical data analysis was performed using SPSS and the findings from the study showed five factors that influence the performance of students of which four were significantly effective in building the predictive logistic regression model.

Metcalfe (2012) examined the factors on which student veterans with disabilities differed from their student veteran peers without reported disabilities by using univariate tests of significance, a LR, and a discriminant function analysis. Univariate tests of significance revealed that students with disabilities had a significantly lower mean GPA, were more often male than female, tended to favor certain academic majors over others, more often enrolled in bachelor's degree versus associate and certificate programs, and had a lower risk of attrition based on their index of risk. Major degree program type and risk index proved to be the most significant predictors of disability status in LR and discriminant function analysis.

The study by Erimafa et al. (2009) applied a DA to classify students who might be at risk of graduating with poor class and students who would graduate with better class within their two years of study. SPSS was used for analysing student's records of 100 and 200 levels from 2004-2007 academic years. The findings revealed that a linear discriminant function successfully predicted about 87.5% of the graduating student's class of degrees with a hit ratio of 88.2% of students with unknown group membership. The use of DA for predictive purposes enabled researchers to identify students who might be at risk of graduating or students who would graduate with better class. This is an indication that DA is suitable for predicting performance of students for preparation of the future.

On the contrary, Peng et al. (2002) mentioned that many educational researches call for the analysis and prediction of a dichotomous outcome. He further explained that ordinary least square and linear discriminant analysis are useful techniques in addressing statements such as whether a student will succeed or not, or whether a child has a learning disability or not. The authors recommend these techniques when handling binary outcomes because of their stringent statistical assumptions.

Kisaka-Lwayo (2007) used DA to identify the characteristics that distinguish between fully-certified organic, partially-certified organic and non-organic farmers in Umbumbulu district, KwaZulu-Natal (KZN), South Africa (SA), during October-November 2004. The study interviewed 200 farmers who were drawn by purposively selecting 151 organic farmers who were members of the Ezemvelo Farmers' Organisation (EFO) and by random sampling 49 non-organic farmers in neighbouring wards. The results from the two estimated discriminant functions suggested that farmers with higher household sizes, incomes, input costs per hectare and number of chickens owned, locations further from innovators and less risk aversion were more likely to be certified as organic. The author recommended that the household location should be considered in delineating target domains for introducing new technologies especially where resources are limited.

Pohar et al. (2004) considered the problem of choosing between the LDA and MLR by using several simulated datasets, and they concluded that LDA is a more appropriate method when the explanatory variables are normally distributed, and LR overcomes DA only when the number of categories is small and the results of LDA and MLR are close whenever the normality assumptions are not too badly violated.

Similarly Hyeoun-Ae (2013) wrote an introduction on how to use and report LR models in the nursing domain. The study was conducted by using twenty-three articles published between 2010 and 2011 in the *Journal of Korean Academy of Nursing*. Peng et al (2002) also used LR to provide insight and guidelines on how to interpret the results when using this technique. This was achieved by evaluating eight articles published in the *Journal of Educational Research* between 1990 and 2000. The authors concluded that all the eight studies met or exceeded the recommended criteria and therefore the results were considered stable.

Bayaga (2010), on the other hand, explored the usage of MLR in risk analysis, using data from a hundred risk analysts of a historically black South African University. In a study by Montgomery et al (1987), the authors compared the two methods in veterinary data using stepwise linear discriminant analysis and LR. The aim of the study was to predict coliform mastitis in dairy cows. The authors then concluded that both techniques selected the set of variables as important predictors and were of approximately equal value in classification performance.

Easa & Alkarkhi (2008) used DA and LR to analyse the concentration of arsenic and heavy metal contents in cockles. The data of this study consisted of five samples of cockles that were collected from the 40 estuarine sites in 2005, Malaysia. The results from the study showed that both techniques exhibited results that were close in discriminating the two locations and the authors concluded that neither of the two techniques could be used.

2.6. Summary

The above mentioned studies give a clear indication that DA and LR can be applied in many fields of study. The various studies showed that when one is conducting a study using DA, careful attention needs to be given to the assumptions because any violations of these assumptions can lead to erroneous results. Alternatively, a researcher can utilise LR as it is not so rigid when assumptions are involved. Evaluating the studies in the field of finance, the LR technique seems to give better results than DA and therefore becomes the preferred technique. On the other hand, the use of principal component analysis as a reducing agent in studies by Suleiman et al. (2014), Kong-lai & Jing (2010) and Haq et al. (2015) seem to enhance the overall performance of the two techniques. On the whole, the two techniques showed that they are more than capable of handling categorical data and give results that are useful for interpreting the challenges in different fields of study. This current study adds to the body of knowledge and already existing literature by deriving insight on the performance of the two techniques combined.

CHAPTER 3

RESEARCH METHODOLOGY

3.1. Introduction

The focus of this chapter is on the methodology of both MDA and MLR. Section 3.2 begins by providing a full description of the data used in this study. The study also covers the different methods employed to achieve the objectives stipulated in Chapter 1. Sections 3.3 and 3.4 holistically cover both the preliminary and primary data analysis based on Figure 3.1:

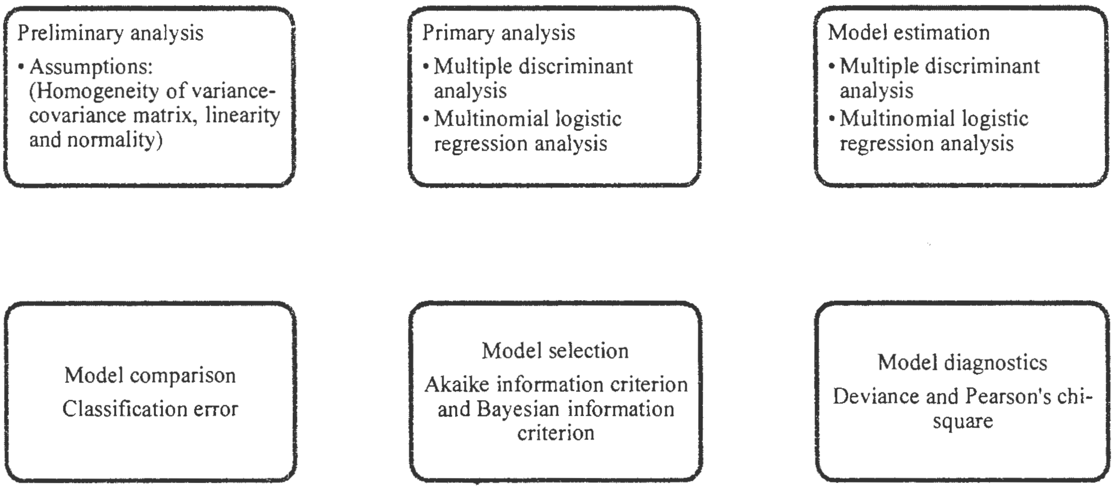

Figure 3.1: MDA and MLR decision process

3.2. Data description and source

The data used in this study was obtained from the North-West University database. It comprises the performance of all third year students majoring in statistics in the School of Economic and Decision Sciences. This study only focused on the performance of students for 2013 to 2016 academic years. The performance was on the four final year modules offered in the first and second semester and comprises the following modules: Module 1 (Multivariate techniques), Module 2 (Time series Analysis), Module 3 (Econometrics methods) and Module 4 (Forecasting methods). The sample size for this current study is shown in Table 3.1 below. The Software Packages for Social Sciences (SPSS) version 22 was used for data analysis.

Table 3.1: Description of data

Year	Module	Number of students
2013	Module1	50
	Module2	50
	Module3	50
	Module4	50
2014	Module1	44
	Module2	44
	Module3	44
	Module4	44
2015	Module1	44
	Module2	44
	Module3	44
	Module4	44
2016	Module1	49
	Module2	49
	Module3	49
	Module4	49

3.3. Preliminary Data Analysis

Marshall & Gretchen (1999) describe data analysis as the process of bringing order, structure and meaning to the mass of collected information. This section discusses the measures and tests used for obtaining preliminary data analysis results which ensures that the data is well prepared for the main analysis.

3.3.1. Descriptive Statistics

According to Garson (2012), all forms of statistical analysis assume sound measurement, relatively free of coding errors. Descriptive statistics, according to Tabachnick & Fidell (2007), describe samples of subjects in terms of variables or a combination of variables. Garson (2012) stipulates that it is good practice to run descriptive statistics on data so that one is confident that the data is generally as expected. The author suggests descriptive measures such as the mean and standard deviation and out-of-bounds entries. These statistics also show relevant information such as the number of cases, missing cases and generally the direction that the responses tend towards.

Following the definition by Rencher (2002), the sample mean of a random of n observations $y_1, y_2, \ldots, y_n$ is calculated as follows:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{3.1}$$

where $\sum y$ represents the sum of all values and n represents the sample size. The mean represents the gravitational center of a distribution and the standard deviation can be calculated as follows:

$$S = \left(\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2 \right)^{1/2}, \tag{3.2}$$

where N is the total number of elements in the population , y_i is the ith observation of the variable y and $\bar{y}$ is the sample mean of y. The standard deviation measures how concentrated the data is around the mean. The distributions with big standard deviations have more variability than distributions with small standard deviations.

3.4. Assumptions

This section reviews the proposed tests that are employed to address the assumptions of the MDA and MLR. These assumptions include multivariate normality, homogeneity of covariance matrices and multicollinearity.

3.4.1. Normality

According to Liong & Foo (2013) DA requires that the independent variables be normally distributed with equal covariance matrices. The authors further explain that when multivariate normality assumption is met, LR is able to give the same level of accuracy as DA. However, if the normality assumption is not fulfilled, the use of LR has more benefits compared to DA because the results of LR are not affected by the degree of normality of the independent variables since it is a non-parametric method.

There are several statistics that are utilised to prove if the variables are normally distributed or not. According to Ghasemi & Zahediasl (2012), visual inspection of the distribution may be used for assessing normality. The frequency distribution (histogram), stem-and-leaf plot, boxplot, P-P plot (probability-probability plot), and Q-Q plot (quantile-quantile plot) are used for checking normality visually. The frequency distribution plots the observed values against their frequency, provides both a visual judgment about whether or not the distribution is bell shaped and provides insights about gaps in the data and outlying values. Another way of testing normality is by using the Kolmogorov-Smirnov test or the Shapiro-Wilks test (Balogun, 2015). This study considered the use of the Q-Q plots and the Kolmogorov-Smirnov test.

3.4.2. Homogeneity Review Test

The multivariate procedures assume that the individual group covariance matrices are equal or homogeneous across groups. According to Liong & Foo (2013), having unequal matrices of covariance could impact negatively on the classification process. However, Abledu et al. (2016) mentions in their study that several studies that involved extant data sets did not suggest that LDA's performance would suffer appreciably because the assumption was violated. Furthermore, Tangen (2008) indicates that when the group samples are equal or large, the DA is robust to violation of this assumption. Therefore despite the violation, one continues with the analysis.

The Box's M test is employed in this study for testing the homogeneity of the variance-covariance matrices. Due to its sensitivity, this test must be non-significant and an alpha level of 0.001 is recommended (Jihad, 2015). The non-significant value of this test indicates that there is equal dispersion. The hypothesis for the Box's M test is tested under the following conditions:

H_0: The covariance matrices are equal

H_1 : The covariance matrices are not equal

Decision rule: Reject H_0 if $p < 0.001$ otherwise do not reject H_0 at 0.01% level of significance.

According to Suleiman et al. (2014), the test statistic is:

$$M = \frac{|S_L|}{|S_s|}, \tag{3.3}$$

where s_L is the larger variance and S_S is the smaller variance.

3.4.3. Multicollinearity

Multivariable analysis, according to Yoo et al. (2014), is a commonly used statistical method when multiple predictive variables are considered to estimate the association with study measurements. Furthermore, Yoo et al. (2014) mention that the efficiency of multivariable analysis highly depends on correlation structure among predictive variables since inference for multivariable analysis assumes that all predictive variables are uncorrelated. Therefore, when the covariates in the model are not independent from one another, collinearity or multicollinearity problems arise in the analysis. This can lead to biased coefficient estimation and a loss of predictive power. Collinearity increases the estimate of standard error of regression coefficients, causing wider confidence intervals and increasing the chance to reject the significant test statistic (Tu, et al., 2005 & Dohoo, et al., 1997). This leads to imprecise estimates of regression coefficients with wrong signs and large differences in regression coefficients. Such a scenario causes a loss in power and makes interpretation more difficult since there is a lot of common variation in the variables. Like in multiple regression, multicollinearity in DA is identified by examining tolerance values or the variance inflation factor (VIF). According to Tabachnick & Fidell (2007), these tests are calculated as follows:

$$Tol = 1 - R_j^2 = \frac{1}{VIF_j}. \tag{3.4}$$

A tolerance of < 0.1 is the same as $VIF > 10$ thus indicating excessive multicollinearity. Though debatable, a tolerance of 0.50 or higher is generally acceptable and some scholars accept values as low as 0.20. In the case where data displays multicollinearity, then one can standardise the predictors or run a completely different analysis using principal component analysis.

3.5. Discriminant Analysis

This section highlights the steps that were adhered to when conducting the DA in this study.

3.5.1. DA model specification

According to Osita (2010), the linear discriminant function which was first introduced by Sir Ronald Fisher provides a rule of classifying an individual into one of several populations on the basis of a set of independent variables. The populations of independent variables p variables are assumed to be multivariate normal with a common covariance matrix but with different means for the groups. A linear function of the p variables according to Fisher is:

$$Y = a + b_1 x_1 + b_2 x_2 + \ldots + b_p X_p. \tag{3.5}$$

According to Press and Wilson (1978), linear discriminant analysis produces a linear combination of variables that maximises the separation between the two groups and then classifies new observations as belonging to one group or the other based on their score on the weighted combination of variables chosen as the discriminant function. Suppose that there are n_0 and n_1

individuals in group 0 and 1, respectively. Let X_{ijk} denote the value of the j_{th} individual in the i_{th} group can be specified as:

$$Y_{ij} = a + b_1 X_{Module1} + b_2 X_{Module2} + b_3 X_{Module3} + b_3 X_{Module4} \tag{3.6}$$

For each individual, the value of the p independent variables can be transformed into an index or a score Y_{ij}. On the basis of Y_{ij}, the individual is classified into group 0 and 1. As a matter of fact, DA may be conceptualised in terms of evaluating the centroid of a group or cases (Akomolafe & Amahia, 2015). In the context of the current study, the performance of a student is evaluated on the basis of the final examination score in a particular group. The researcher as such evaluates a mean value of a discriminating variable for a student in a specific group. Depending on the difference between the mean values of the independent variables, the bigger the difference, the more discriminating is that variable. The use of DA enables a researcher to determine the predictors which are more effective in predicting students' performance.

3.5.2. Estimation of Coefficients

To determine the coefficients $b_1,...,b_p$ so that the discriminant function Y_{ij} has the maximal discriminatory power, Prempeh (2009) proposed $\overline{X}_{0k}$ and $\overline{X}_{1k}$ as the observed mean values of the k th independent variable in group 0 and 1 respectively, that is:

$$\overline{X}_{ik} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ijk}, \qquad i = 0,1 \text{ and } k = 1...p. \tag{3.7}$$

The pooled sample covariance matrix of the p variables is then:

$$S = \begin{pmatrix} a_{11} & a_{12} \cdots\cdots a_{1p} \\ a_{21} & a_{22} \cdots\cdots a_{2p} \\ \vdots & \\ a_{p1} & a_{p2} \cdots\cdots a_{pp} \end{pmatrix} \tag{3.8}$$

where

$a_{kk} = \frac{\sum_{i=0}^{1}\sum_{j=0}^{n_i}(X_{ijk} - \overline{X}_{ik})^2}{n_0 + n_1 - 2}$ is the variance of the kth variable. The inverse of **S** can be computed as:

$$S^{-1} = \begin{pmatrix} S_{11} & S_{12} \cdots\cdots S_{1p} \\ S_{21} & S_{22} \cdots\cdots S_{2p} \\ \vdots & \\ S_{p1} & S_{p2} \cdots\cdots S_{pp} \end{pmatrix} \qquad (3.9)$$

Further, let d_k be the difference of the two group means for variable K.

$d_k = \overline{X}_{0k} - \overline{X}_{1k}$, therefore, the coefficients $b_1, b_2, ..., b_p$ of the discriminant function are obtained as follows:

$$\begin{aligned} b_1 &= S_{11}d_1 + S_{12}d_2 + ...S_{1p}d_p \\ b_2 &= S_{21}d_1 + S_{22}d_2 + ...S_{2p}d_p \\ &\vdots \\ b_p &= S_{p}d_1 + S_{p_2}d_2 + ...S_{pp}d_p \end{aligned}$$, so that

$$Y = (\overline{X}_0 - \overline{X}_1)' S^{-1} X. \qquad (3.10)$$

For a new individual, an index or a score that is the value of y in (3.5) can be computed on the basis of $X_1 ... X_p$ *and* $b_1 ... b_p$ values. This value of Y is used to classify the individual into one of the groups. In order to do that, a classification rule has to be established:

Let $\overline{Y}_0 = b_1\overline{X}_1 + b_2\overline{X}_2 + ... + b_p\overline{X}_{0p}$ i.e. $\overline{Y}_0 = (\overline{X}_0 - \overline{X}_1)S^{-1}\overline{X}_0$ (3.11)

$\overline{Y}_1 = b_1x_{11} + b_2x_{12} + ... + b_px_{1p}$ i.e. $\overline{Y}_1 = (\overline{X}_0 - \overline{X}_1)S^{-1}\overline{X}_1$ (3.12)

where $\overline{Y}_0$ and $\overline{Y}_1$ are the mean scores of groups 0 and 1 respectively. A symmetrical dividing line between the groups is the average of the two mean scores:

$$Y_c = \frac{1}{2}(\overline{Y}_0 + Y_1) = \frac{1}{2}b_1(\overline{X}_{01} + \overline{X}_{11}) + ... + \frac{1}{2}(\overline{X}_{0p} + \overline{X}_{1p})$$

Since $\bar{Y}_0 > Y_1$, assign a new individual to group 0 if $Y > Y_C$ and if $Y < Y_C$ to group 1. The standard linear discriminant model assumes that the conditional distribution of x / y is multivariate normal with mean vector μy and common variance matrix Σ. Adopted from Pohar (2004), it can be shown by assigning x to group 1 as:

$$P(1 \mid x) = \frac{1}{1 + \left(e^{\alpha + \beta x}\right)^{-1}} \tag{3.13}$$

where α *and* β coefficients are

$$\beta = \left(\mu_1 - \mu_0\right)^T \Sigma^{-1}$$

$$\alpha = -\log \frac{\pi_1}{\pi_0} + \frac{1}{2}\left(\mu_1 + \mu_0\right)^T \Sigma^{-1}\left(\mu_1 - \mu_0\right) \tag{3.14}$$

π_1 *and* π_0 are prior probabilities belonging to group 1 and group 0. In practice, the parameters $\pi_1, \pi_0, \mu_1, \mu_0$ *and* Σ will be unknown, so they are replaced by sample estimates. The estimates are as follows:

$$\hat{\pi}_1 = \frac{n_1}{n}, \hat{\pi}_0 = \frac{n_0}{n},$$

$$\hat{\mu}_1 = \bar{x}_1 = \frac{1}{n} \sum_{y_i = 1} x_i, \hat{\mu}_0 = \bar{x}_0 = \frac{1}{n_0} \sum_{y_i = 0} x_i, \tag{3.15}$$

$$\hat{\Sigma} = \left[\sum_{y_i = 1} (x_i - \bar{x}_1)(x_i - \bar{x}_1)^T + \sum_{y_i = 0} (x_i - x_0)(x_i - x_0)^T \right] / n$$

3.5.3. Assessing the Goodness-of-fit of the estimated model

According to Garson (2012), the Kolmogorov-Smirnov test (also known as the K-S Lilliefors test) can be used to test the goodness-of-fit against any theoretical distribution, not just the normal distribution. According to Lilliefors (1967), this statistic provides a means of testing whether or not a set of observations are from some completely specified continuous distribution. The test is more powerful than the chi-square test for any sample size. The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The hypothesis associated with the Kolmogorov-Smirnov test statistic is:

$H_0 : \varepsilon_i$ *is normally distributed*

$H_1 : \varepsilon_i$ *is not normally distributed*

If $prob\ (D) < 0.05$ then reject H_0

If $prob\ (D) > 0.05$ then fail to reject H_0

These tests require a relatively large number of data points to properly reject the null hypothesis.

The Kolmogorov-Smirnov (D) is defined as:

$$D = \sup_x |F_n(x) - F(x)| \tag{3.16}$$

where

$F(x)$ is the sample cumulative distribution function

$F_n(x)$ is the cumulative normal distribution function with $\mu = \bar{x}$, the sample mean, and $\sigma^2 = s^2$, the sample variance is defined with denominator $n - 1$.

3.5.4. Wilk's Lambda

This test is relevant in testing the significance and the strength of relationship for each discriminant function and whether or not group membership can be predicted using discriminant function therefore providing information about the statistical significance of the entire model using all discriminant functions and predictor variables. The Wilk's Lambda (also known as the U statistic) is a multivariate technique used to compare the variation within the samples to the variation both within and between the samples and tests for significant differences between groups. This statistic tells us the variance of a dependent variable that is not explained by the discriminant function. Wilk's lambda is also used to test for significant differences between the groups on the individual predictor variables. It tells which variables contribute a significant amount of prediction to help separate the groups. The Wilk's Lambda is used to test the significance of the discriminant functions. Unlike the F-statistics in linear regression, when the value lambda for a function is small, the function is significant. A small Λ (lambda) indicates a low within variation in comparison with the total variation, providing evidence that the sample does not come from entities with same mean vector. The likelihood ratio test according to Rencher (2002) is given by:

$$\Lambda = \frac{|E|}{|E+H|} = \prod_{i=1}^{p} \frac{1}{(1+\lambda_i)} \tag{3.17}$$

which is known as Wilks' Λ. where

p = number of variables

H = degrees of freedom of hypothesis

E= degrees of freedom of error

The Wilks' Λ. is tested under the hypothesis that:

$H_0: \beta_1 = \beta_2 = \ldots = \beta_k$ (group means are the same)

H_1 :some $\beta_i \neq \beta j$ for i, $j = 1, 2 \ldots k$

Decision rule: Reject H_0 if $p < 0.05$ otherwise fails to reject H_0 at the 5% level of significance.

3.6. Interpretation of Discriminant Functions

According to Jihad (2015), the correspondence between interpreting discriminant functions and determining the contribution of each variable is very close. When it comes to interpretation, the signs of the coefficients are taken into account and in determining the contribution, the signs are then ignored, and the coefficients are classified in absolute value. The examination of the standardised discriminant function coefficients and calculation of partial F-test for each variable is the most common approach in assessing the contribution of each variable or discrimination.

3.6.1. Standardised Coefficients

The standardised canonical discriminant coefficients can be used to rank the importance of each variable. A high standardised discriminant function coefficient might mean that the groups differ a lot on that variable. This approach examines the sign and the scale of the standardised discriminant weight assigned to each variable in computing the discriminant functions. The independent variables with a relatively large weight contribute more to the discriminating power of the function than variables with smaller weights. According to Rencher (2002), the standardised coefficients must be of the form:

$$a_j^* = a_j s_j \; r = 1,2,\cdots,p \tag{3.18}$$

The mean standardized discriminant weight for each variable is calculated by:

$$\bar{b}_j = \sum_{k=1}^{K} \left| b_{jk}^* \right| . \; EP_k, \tag{3.19}$$

where

b_{jk}^* = standardised discriminant weight for variable j and discriminant function k

EP_k = Eigenvalue proportion of DA function k.

3.6.2. Partial F-values

In order to calculate a partial F-statistic, Tabachnick & Fidell (2007) suggested that firstly, cross-products matrices for between-group differences and within-groups differences should be created: $SS_{total} = SS_{bg} + SS_{wg}$. The determinants are calculated for these matrices and used to calculate a test statistic, either Wilks' Λ or Pillai's trace. The procedure for calculating approximate F is based on Wilks' Λ and the various degrees of freedom associated with it.

The approximate $F(df_1, df_2) = \left(\frac{1-y}{y}\right)\left(\frac{df_2}{df_1}\right)$

with df_1 and df_2 defined as the degrees of freedom for testing the F ratio and $y = \Lambda^{\frac{1}{s}}$.

$df_1 = pv_H$, $df_2 = wt - \frac{1}{2}(pv_H - 2)$ and $w = v_E + v_H - \frac{1}{2}(p + v_H + 1)$, $t = \frac{p^2 v^2{}_H - 4}{p^2 + v^2{}_H - 5}$.

s is defined as follows: $s = \min(p, df_{effect})$, where p is the number of predictor variables and d_{effect} is the degrees of freedom for the effect being tested. Therefore

$df_1 = p(df_{effect})$ And $df_2 = s\left[(df_{error}) - \frac{p - df_{effect} + 1}{2}\right] - \left[\frac{p(df_{effect}) - 2}{2}\right]$, where df_{error} = the degree of freedom associated with the error term. This test is useful in determining whether classification improves as a new set of predictors is added to the analysis. It is used together with the Wilk's Lambda to test which variables contribute to the significance in the discriminant function.

3.7. Logistic regression

This section outlines the methods that should be followed in LR. This includes the assumptions, LR model, estimating the parameters of LR model and goodness-of-fit tests.

3.7.1. Estimating the Logistic Regression Model (LRM)

According to Agresti (2013), Sule & Saporu (2015), the Logistic Regression Model (LRM) is one of the important models applied to analyse categorical data. MLR provides an effective and reliable way to obtain the estimated probability of belonging to a specific population and the estimate of odds ratio. It is also a procedure by which estimates of the net effects of a set of explanatory variables on the response variable can be obtained. According to Garson (2012), MLR is used to predict a response variable on the basis of continuous and/or categorical explanatory variables to determine the percent of variance in the response variable explained by the explanatory variables. This procedure is also relevant in ranking the relative importance of independents to assess interaction effects and understand the impact of covariate control variables. He furthermore stipulates that this technique allows the simultaneous comparison of more than one contrast, which is the log odd of three or more contrasts.

Odds ratio is defined as the probability that an event will occur divided by the probability that the event will not happen. According to Hair et al. (2010), the logistic transformation ensures that estimated values do not fall outside the range of 0 and 1. This can be accomplished by restating a probability as odds and calculating the logit value. The odds ratio can be calculated by the ratio of the probability of the two outcomes or events, written as $\Pr ob_i \div (1 - \Pr ob_i)$. In order to retain the odds values from going below 0, which is the lower limit of the odds, we then compute the logit value. The odds function can also be written as follows:

$$Odds_i = \left(\frac{prob_{event}}{1 - prob_{event}} \right) = e^{b0 + b_1 X_1 + \ldots + b_n X_n}. \tag{3.20}$$

The odds value can be converted back into a probability that falls between 0 and 1. This can be written as:

$$\Pr obability_i = \frac{odds}{1 + odds} \tag{3.21}$$

In order to keep the odds values from going below 0, which is the lower limit of the odds (no upper limit), the logit value should be computed. This is achieved by calculating the log odds and the odds with less than 1 have a negative logit value. Similarly, the odds ratio greater than 1.0 have positive logit values and the odds ratio of 1.0 have a logit value of 0.

According to Muchabaiwa (2013), the formula for a logistic regression model is given by:

$$\pi(x_i) = P(y_i = 1 : x_i) = [1 + \exp(-X'\beta)]^{-1} \tag{3.22}$$

where, $y_i = \begin{cases} 1 \\ 0 \end{cases} \quad i = 1, 2, \ldots, n$

$$X^T \beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{p-1} X_{p-1} \tag{3.23}$$

and $x_1, x_2, \ldots, x_k$ are the independent variables with,

β_0 as the coefficient of the constant term

$\beta_1, \beta_2, \ldots, \beta_{p-1}$ are the coefficients of the p independent variable

$\pi(x_i)$ is the probability of an event that depends on p-independent variables

Since $\pi(x_i) = [1 + \exp(-X^T \beta)]^{-1}$

$$= \frac{1}{1 + \exp(-X^T \beta)}$$

$$= \frac{\exp(-X^T\beta)}{1+\exp(-X^T\beta)}$$

$$\frac{\pi(x_i)}{1+\pi(x_i)} = [\exp(-X^T\beta)]^{-1}$$

thus

$$\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \text{logit}[\pi(x_i)]$$

$$= X^T\beta \qquad (3.24)$$

3.7.2. Interpretation of coefficients using odds

According to Tabachnick and Fidell (2007), odds ratio greater than 1 reflect the increase in odds of an outcome of 1 (considered as the response category) with one-unit increase in the predictor and odds ratios less than one reflect the decrease in odds of that outcome with a one-unit change.

3.7.3. Model estimation

According to Prempeh (2009) and Yeboah (2012), LR uses maximum likelihood estimation to compute the coefficients for the LR equation. The method of Maximum Likelihood Estimation (MLE) chooses values for parameter estimators (regression coefficients) which makes the observed data to be maximum likely. The MLE procedure continuously attempts to get closer and closer to the correct answer and iterates until the absolute value of the largest parameter change is less than the value specified for tolerance on the LR modelling (Prempeh, 2009).

Agresti (2013) indicates that when more than one observation occurs at a fixed x_i value, it is sufficient to record the number of observations n_i and the number of successes. Let y_i refer to the success count rather than to an individual binary response. Then $\{Y_1,\ldots,Y_N\}$ are independent binomials with $E(Y_i) = n_i\pi(x_i)$, where $n1 + \cdots + nN = n$. The MLE is derived from the probability distribution of the dependent variable:

$$\prod_{i=1}^{N} \pi(x_i)^{y_i} [1-\pi(x_i)]^{n_i - y_i} \qquad (3.25)$$

$$= \left[\prod_{i=1}^{N} \exp\left[\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)^{y_i}\right]\right]\left\{\prod_{i=1}^{N}[1-\pi(x_i)]^{n}\right\}$$

$$= \left\{\exp\left[\sum_{i=1}^{N} y_i \log\frac{\pi(x_i)}{1-\pi(x_i)}\right]\right\}\left\{\prod_{i=1}^{N}[1-\pi(x_i)]^{n_i}\right\}.$$

The exponential term is equal to

$$\exp\left[\sum_i y_i (\sum_j \beta_j x_{ij})\right] = \exp\left[\sum_j (\sum_i y_i x_{ij})\beta_j\right] \qquad (3.26)$$

since

$$[1-\pi(x_i)] = \left[1+\exp\left(\sum_j \beta_j x_{ij}\right)\right]^{-1} \qquad (3.27)$$

The log likelihood equals

$$L(\beta) = \sum_j \left(\sum_i y_i x_{ij}\right)\beta_j - \sum_i n_i \log\left[1+\exp\left(\sum_j \beta_j x_{ij}\right)\right] \qquad (3.28)$$

This depends on the binomial counts only concluded by the satisfactory statistics for the model parameters $\{\sum_i y_i x_{ij}\}$, $j = 0,1,\ldots,p$.

The likelihood equations result from setting $\partial L(\beta)/\partial\beta = 0$ since:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_i y_i x_{ij} - \sum n_i x_{ij} \frac{\exp(\sum_k \beta_k x_{ik})}{1+\exp(\sum_k \beta_k x_{ik})}, \qquad (3.29)$$

the likelihood equations are:

$$\sum_i y_i x_{ij} - \sum_i n_i \hat{\pi} x_{ij} = 0 \quad \text{j=0, 1,..., p} \qquad (3.30)$$

where

$$\hat{\pi} = \frac{\exp\left(\sum_k \hat{\beta}_k x_{ik}\right)}{\left[1 + \exp\left(\sum_k \hat{\beta}_k x_{ik}\right)\right]} \quad (3.31)$$

is the MLE of $\pi(x_i)$.

Let X denote the matrix of values of $\{x_{ij}\}$ with N rows for the binomial observations and a column for each parameter. The likelihood equation is:

$$X^T y = X^T \hat{\mu}, \quad (3.32)$$

where $\hat{\mu}_i = n_i \hat{\pi}_i$ and the likelihood equations equate the sufficient statistics to the expected values. The main focus of using MLE is to yield unbiased estimators (estimate true $\beta's$), asymptotically efficient estimators, normally distributed estimators and consistent estimators.

3.7.4. Assessing the Goodness-of-fit of the estimated model

To evaluate the goodness-of-fit of the LR model, three common methods are used namely: the Deviance method, the Pearson method and the Hosmer-Lemeshow (2000) method. This study concentrates on the Deviance and the Pearson method. The hypothesis to test is:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$ (Model is not good)

$H_1 : \beta_j \neq 0$ (Model is good)

The two test statistics for evaluating the above hypotheses are discussed in subsequent sections.

3.7.5. The Deviance Method

The observed values of the dependent variable must be compared with the estimated values obtained from models with and without the variable in question. This comparison is based on the log-likelihood function given by:

$$\sum_{i=1}^{n} \{y_i \ln(\pi(x_i)) + (1 - y_i)\ln[1 - \pi(x_i)]\}. \quad (3.33)$$

A comparison, according to Prempeh (2009), has to be made between a saturated model and the current model where a saturated model is one that contains as many parameters as the number of data points and the current model is one that contains only the variables being assessed. The comparison of the current to saturated model is based on the likelihood ratio:

$$D = -2\ln\left[\frac{likelihood\ of\ the\ current\ \text{model}}{likelihood\ of\ the\ saturated\ \text{model}}\right]$$

Using the two equations above, the test statistic is:

$$D^2 = 2\sum_i \left\{ Y_i \ln\left(\frac{Y_i}{n_i p_i}\right) + (n_i - Y_i)\ln\left(\frac{n_i - Y_i}{n_i(1-p_i)}\right)\right\} \sim \chi^2_{\alpha,n-p} \qquad (3.34)$$

Deviance (D) follows a chi-square distribution with q-degrees of freedom, where q is the number of covariates in the equation. A goodness-of-fit test generally refers to measuring how well observed data correspond to the fitted (assumed) model.

At some α-level of significance, H_0 is rejected if $G^2 > \chi^2_{\alpha,n-p}$ or $prob(G^2) < \alpha$. According to Agresti (2007), large deviance values and p-values less than 0.05 are an indication of lack of fit of the current model.

3.7.6. The Pearson Method

Pearson's method is another way of assessing model fit. This test statistic is a classical summary measure of goodness-of-fit.

The test statistic is:

$$\chi^2 = \sum \frac{(Y_i - n_i\hat{p}_i)^2}{n_i\hat{p}_i(1-\hat{p}_i)} \sim \chi^2_{\alpha,n-p}$$

At some α-level H_0 is rejected if $\chi^2 > \chi^2_{\alpha,n-p}\ or\ prob(\chi^2) < \alpha$

3.8. Testing for significance of the coefficients

This section summarises the Wald test which is relevant for testing the significance of the coefficients.

3.8.1. Wald Test

The Wald test statistic was named after the Hungarian statistician, Abram Wald. This test is used to test the true value of the parameter based on the sample estimate. The Wald test is a squared coefficient divided by its squared standard error (Tabachnick & Fidell, 2007). The hypotheses that are tested are:

$H_0 : B_j = 0 \; vs \; H_1 : B_j \neq 0$ where $(j = 1,2,\ldots, p)$

Under the null hypothesis H_0 the Wald test statistic is:

$$Wald = \frac{\hat{\beta}_j}{SE(\hat{\beta}_i)} \sim N(0,1) \tag{3.35}$$

and

$$Wald^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \sim \chi^2_{\alpha,1} \tag{3.36}$$

H_0 is rejected if $Wald^2 > \chi^2_{\alpha,1}$ or prob($Wald^2$)<α.

The Wald statistic is chi-square distributed with 1 degree of freedom. The null hypothesis is rejected if the p-value of the test is less than 0.05 (significance level). A coefficient with a p-value of the Wald statistic less than 0.05 implies that the variable is important in the model.

3.9. Comparison Criteria between DA and LR

This section focuses on using the classification error (table) for comparison of both the techniques.

3.9.1. Classification Table

The classification table is simply one in which the rows are the observed categories of the dependent and the columns are the predicted categories. When prediction is perfect all cases will lie on the diagonal. The percentage of cases on the diagonal is the percentage of correct classifications.

Table 3.2: Classification table

Classification		*Expected*		*Total*
		P	N	
Observed	P	PP (TP)	PN (FN)	P
	N	NP (FP)	NN (TN)	P'
	Total	Q	Q'	1

Table 3.2 above shows a representation of the results obtained by means of classification.

PP: represents the actual positive and classified as positive and PN: represents the actual positive, but classified as negative.

NP: actually negative, but classified as positive and NN: actually negative, and classified as negative.

TP: True positive, FP: false positive, TN: true negative and FN: False negative

According to Cizek & Fitzgerald (1999), the classification table uses the sensitivity and the specificity statistic. Sensitivity is known as the percentage of successes that have been correctly classified as success and can be denoted by SEN. The percentage of failures that have been classified is called specificity, denoted as SPE. The failures that are incorrectly classified as success are referred to as false positive and the successes that are incorrectly classified as failures are referred to as false negatives. SEN is calculated from this equation:

$$SEN = \frac{TP}{TP + FN} = \frac{TP}{P} \qquad (3.37)$$

and $$SPE = \frac{TN}{FP + TN} = \frac{TN}{P'} \qquad (3.38)$$

Ratio of correct classification (also known as Hit Ratio) is defined as the probability value of the correct classification or efficiency ratio (EF) and is calculated from the equation:

$EF = TP + TN$.

The hit ratio is then obtained as follows: $Hit\,ratio = \frac{EF}{Total} = \frac{TP + TN}{P + P'} = \frac{TP + TN}{Q + Q'}$ (3.39)

Both the sensitivity and specificity are usually given in percentages. A decision method is considered good if it simultaneously has a high sensitivity and a high specificity so there is a trade-off between sensitivity and specificity (El-habil & El-jazaar, 2014).

3.10. Summary

There are several stages that need to be followed when conducting a comparative study. The first stage is to begin with the preliminary tests which include the descriptive statistics and testing the assumptions of the techniques. The study then continues with the primary analysis, focusing on the tests for both MDA and MLR. In order to achieve the objectives of the study, these tests need to be followed and adhered to.

For MDA to be achieved there are several tests or measures that need to be followed like testing the significance of the model by looking at the Wilk's lambda. This is followed by interpretation of the discriminant functions by using the standardised coefficients and the partial F-values.

For estimating the MLR model, one looks at interpretation of coefficients using odds and MLE. This is followed by assessing the goodness-of-fit of the estimated model using the Deviance, Pearson method or the Hosmer Lemeshow. In this case, the Deviance and the Pearson are relevant in this study. Furthermore, the study tests the significance of the coefficients using the Wald test and model validation using the AIC values. Comparison of the two techniques is achieved by using the classification error table.

CHAPTER FOUR

RESULTS AND INTERPRETATION

4.1. Introduction

This chapter focuses on data analysis and presentation of statistical results. The data used was students' final marks on the four statistics modules offered to final year statistics students from 2013 to 2016 at North–West University in South Africa. The independent variables were the Module1, Module2, Module3 and Module4 and the Groups in this study were students registered for years 2013 to 2016. The outcomes of this data were generated using SPSS version 22. The results are presented in the form of descriptive statistics for both the MDA and MLR. This is followed by testing the assumptions of normality, homogeneity and multicollinearity prior to the application of the proposed techniques.

Firstly, the MDA results are presented according to the methodology discussed in Chapter 3 followed by the MLR results. The analysis in this chapter was carried out to achieve the following objectives:

- To fit MLR and MDA of students' performance.
- To explore the efficiency of MLR and MDA in students' performance.
- To identify the model with more predictive power.
- To use the findings of the study in formulating recommendations for further studies and policy purposes.

4.2. Preliminary data analysis results

This section presents and discusses the results to enable the researcher evaluate the data for correctness and eligibility of the MDA and MLR methods. The results are presented as descriptive and inferential statistics especially to evaluate the assumptions associated with these methods.

4.3. Descriptive Statistics

Table 4.1 presents the descriptive statistics which explain the performance of students on the four modules.

Table 4.1: Group statistics

Groups		*Mean*	*Std. Deviation*	*Weighted*
2013	Module1	75.32	15.641	50.000
	Module2	71.64	14.540	50.000
	Module3	73.22	14.457	50.000
	Module4	62.46	14.637	50.000
2014	Module1	65.80	16.756	44.000
	Module2	47.84	11.158	44.000
	Module3	67.25	12.006	44.000
	Module4	66.00	15.052	44.000
2015	Module1	70.43	9.936	44.000
	Module2	60.73	12.049	44.000
	Module3	65.11	15.226	44.000
	Module4	67.00	7.205	44.000
2016	Module1	69.04	7.765	49.000
	Module2	60.59	9.893	49.000
	Module3	60.92	9.119	49.000
	Module4	61.65	7.774	49.000

From Table 4.1, in 2014 and 2015 the Department of Statistics enrolled same number of students, six students less than in 2013 while in 2016 the same department only enrolled 49 students. A drop in enrolments could be due to the capping system that saw most of the programs losing students. Due to the limited resources available, the university decided to stick to a certain enrolment figure and trying to align the numbers with the resources as well. The enrolment also closely adheres to prescribed subsidy formula for student support from the government. It is clear, judging from the results, that the performance of students in the four modules was different, and that only Module1 met a minimum throughput rate of 75% as prescribed by the university. This is not a surprise finding as far as performance in any mathematical related subject is concerned. Usually, students perform better in other subjects than they do in Mathematics related subjects.

However, the average performance of students is generally good according to the results save for Module2 in 2014. Clearly the three modules provide insight into the varying performances across all the four groups. The standard deviations in the performance of students across all the groups are equally dispersed. Having gathered that students performed differently in the four modules, one question that remains to be answered is whether or not there is a significant difference between these performances over the years. This question was evaluated using the statistical test results discussed under primary data analysis.

The Case Processing Summary table simply shows how many cases or observations were in each category of the outcome variable (as well as their percentages) and also indicates if there was any missing data. In this study, all the observations in the dataset are confirmed to be valid according to the results presented in Table 4.2.

Table 4.2: Case Processing Summary

		N	*Marginal Percentage*
Groups	2013	50	26.7%
	2014	44	23.5%
	2015	44	23.5%
	2016	49	26.2%
Valid		187	100.0%
Missing		0	
Total		187	
Subpopulation		186	

4.4. Assumptions results

This section discusses the results for the assumptions of MDA and MLR methods. It is important to ensure that the data used conforms to normal distributions prior to applying the MDA and MLR methods. The issue of multicollinearity could also be a problem if it is not addressed at this stage.

4.4.1. Normality Test

Table 4.3: Tests of Normality

	Kolmogorov-Smirnov			*Shapiro-Wilk*		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Module1	.077	187	.009	.909	187	.000
Module2	.047	187	.200	.996	187	.875
Module3	.074	187	.015	.968	187	.000
Module4	.080	187	.005	.902	187	.000

Table 4.3 shows the results of the Kolmogorov-Smirnov test and the Shapiro Wilks test. The results from the Kolmogorov-Smirnov table show that only one variable is normally distributed (Module2). This is validated by the *p*-values greater than 0.05 and the Q-Q plots in Appendices (Figure 7.1). Absence of normality could be as a result of the difference in performance in a

particular module rather than attributable to some students performing substantially better or worse than others. Another reason for lack of normality could be due to sample size used. Even though Module1 and Module4 are not normally distributed, one continues with the analysis (Mordkoff, 2011) since the sample used was the actual number of students enrolled and the researcher neither decrease nor increase it. According to the central limit theorem, this assumption is overruled if a sample size of 30 or more observations is used. Logistic regression does not suffer the consequences of non-normality and as a result this technique is safe. The researcher cautiously interpreted the results of the MLR due to this normality issue.

4.4.2. Homogeneity of variance-covariance matrices

The other distinctive assumption of discriminant analysis is the equality of variance-covariance matrices across groups. The study used Box's M statistic to evaluate this assumption and the results are presented in Table 4.4.

Table 4.4 Log Determinants and Box's M Test

Groups	*Rank*	*Log Determinant*	*Box's M*	*172.004*
2013	4	19.560	F Approx.	5.506
2014	4	20.551	df1	30
2015	4	18.774	df2	89735.642
2016	4	16.543		
Pooled within-groups	4	19.757	Sig.	.000

Table 4.4 shows the results of the Box's M test and log determinants. These tests were used for testing the homogeneity of variance-covariance matrices. Looking at the table, the Box's M statistic has a value of 172.004 and the associated approximate F for this statistic is F (5.506) with a significant p-value of 0.000. The p-value is declared significant if it is greater than 0.05. Judging from the results, the observed p-value is less than 0.05 and therefore the variance-covariance matrices were not homogenous thus indicating a violation of this assumption. This is not a surprising finding since there is imbalance in terms of data in the four modules. One module had 50 observations, two have 44 while the other has 49. This is associated with incomplete data which is regarded as one of the pervasive problems in data analyses (Kim & Bentler, 1999). Tabachnick & Fidell (1996) highlighted that the extent to which incomplete data is problematic is dependent on its pattern and the magnitude of missing values. It should be noted that the author for the current

study does not have control over such a problem, hence the analyses was continued with the data at hand. Also, it should be noted that since the sample size is in excess of 30 observations for each module, hence the violation of normality assumption may not be a problem for the methods utilised.

4.4.3. Multicollinearity Test

The results presented in Table 4.5 are with regard to the multicollinearity in the variables. Generally, a certain degree of collinearity is allowed. However, if two variables are significantly collinear, one that is responsible for such multicollinearity should be dropped.

Table 4.5: Collinearity Statistics

	Tolerance	*VIF*
Module1	.706	1.417
Module2	.820	1.219
Module3	.976	1.025
Module4	.685	1.460

Table 4.5 shows tolerance values that are greater than 0.1 and VIF smaller than 10, therefore one concludes that there is no problem of multicollinearity, therefore further analyses using MDA and MLR were carried out and the results are discussed in subsequent sections.

4.5.Multiple Discriminant Analysis Results (MDA)

This section presents and discusses the results from MDA. The intention was to build a model using students' performance that may be useful in formulating suggestions to the Department of Statistics when making future enrolment and curriculum development plans.

4.5.1. Testing the significance of the model

This test is relevant in testing the significance and the strength of relationship for each discriminant function. In Table 4.6 the first row, '1 through 3', tests whether group membership can be predicted using discriminant function (1 and 3 combined; i.e. it provides information about the statistical significance of the entire model using all discriminant functions and predictor variables since there are three functions in the analysis).

Table 4.6: Wilks' Lambda

Test of Function(s)	*Wilks' Lambda*	*Chi-square*	*Df*	*Sig.*
1 through 3	.571	102.040	12	.000
2 through 3	.878	23.665	6	.001
3	.977	4.274	2	.118

The first Function 1 has a Wilk's $\Lambda = 0.571$ with a chi-square of 102.040, is significant at $p = 0.000$ (since $\alpha = 0.05$) and suggests that the group means differ. The second function has a Wilk's $\Lambda = 0.878$ with the chi-square of 23.665 and is significant at $p = 0.001$. The third function does not assist much in discriminating against the groups because it has a very high lambda of 0.977 indicating that the observed groups are also different to a certain degree. The chi-square statistic is relevant in testing the hypothesis that the means of the functions are equal across groups. Since the focus of this study was developing a model which provides a better discrimination against the four groups of students, Function 3 was not considered for further analysis because it was found not to be significant at other levels of significance.

4.5.2. Test of equality of group means

This test is relevant in establishing which variables contribute significance in discriminant function and therefore measures each independent variable's potential before the model is created. The Wilk's lambda and the F ratio are used to test the equality of group means. The hypothesis is rejected if the observed probability associated with the test statistic is less than the level of significance. The results are summarised in Table 4.7.

Table 4.7: Tests of Equality of Group Means

	Wilks' Lambda	*F*	*df1*	*df2*	*Sig.*
Module1	.933	4.363	3	183	.005
Module2	.668	30.341	3	183	.000
Module3	.887	7.801	3	183	.000
Module4	.964	2.299	3	183	.079

Table 4.7 presents the results of univariate ANOVA carried out for each independent variable. This table presents the Wilk's lambda and the univariate F-ratio with 3 degrees of freedom. The F test from the Wilk's lambda shows that Module1, Module2 and Module3 are highly significant at

$p < 0.05$ while Module4 is insignificant at $p > 0.05$. Therefore one can conclude that the first three modules are relevant in explaining the group membership well.

4.5.3. Relationship between the dependent and the independent variables

The canonical discriminant functions tables below show the overall functions and also evaluates the significance of the functions. In order to achieve this, it must be noted that the number of functions possible is equal to the number of groups minus one. Therefore one assumes that k = number of groups = 4 and p = number of discriminant variables. In this study there are four groups of students registered in 2013, 2014, 2015 and 2016, thus three functions are anticipated. Function 1, 2 and 3 are the only three canonical discriminant functions estimated in this analysis.

The eigenvalue is a ratio of the between-groups sum of squares to the within-groups or error sum of squares. The size of the eigenvalue is helpful in measuring the spread of the group centroids in the corresponding dimension of the multivariate discriminant space. The larger eigenvalues indicate that the discriminant function is more useful in distinguishing between the groups (Hair et al., 2010; Tabachnick & Fidell, 2007). Table 4.8 shows the eigenvalues λ_1, λ_2 and λ_3 which verify the relationship between the dependent variable and the independent variables.

Table 4.8: Eigenvalues

Function	*Eigenvalue*	*% of Variance*	*Cumulative %*	*Canonical Correlation*
1	.538	79.8	79.8	.592
2	.112	16.7	96.5	.318
3	.024	3.5	100.0	.152

Table 4.8 of eigenvalues provides information about the effectiveness of the discriminant functions. Function 1 is more important than Function 2 and 3 because it explains 79.8% of variance among the four groups whereas Function 2 explains 7.4% and Function 3 only explains 3.5%. Function 1 also has the largest eigenvalue (0.538) compared to Function 2 (0.112) and Function 3(0.024), indicating a fairly moderate function and useful in distinguishing between the groups. The degree of relationship between the predictors and groups (canonical correlation) due to Function 1 is 0.592 which is more than that due in Function 2(0.318) and Function 3(0.152).

The canonical correlation which is denoted by η can also be obtained follows: $\eta_1 = \sqrt{\frac{\lambda_1}{1+\lambda_1}}$

$$= \sqrt{\frac{0.538}{1+0.538}} = 0.592,\ \eta_2 = \sqrt{\frac{\lambda_2}{1+\lambda_2}} = \sqrt{\frac{0.112}{1+0.112}} = 0.318 \text{ and } \eta_3 = \sqrt{\frac{\lambda_3}{1+\lambda_3}} \sqrt{\frac{0.024}{1+0.024}} = 0.152$$

4.5.4. Interpretation of Discriminant Functions

The discriminant functions are interpreted by means of standardised coefficients and the structure matrix. When comparing the standardised coefficient, it is possible to identify which independent variable provides more discrimination than other variables. Therefore, these coefficients were used to assess the importance of each independent variable's unique contribution to the discriminant function. The standardised coefficients measured all the variables on the same scale and the weights were compared to determine the relative significance of each of the variables in explaining "group separation" (differences in student's performance). The higher the standardised discriminant coefficient, the higher the discriminating powers.

The structure matrix table reveals the correlations between each variable in the model and the discriminant functions. Generally, any variable with a correlation of 0.3 or more is considered to be important (OriginLab Corporation, 2016). The interpretation of the discriminant coefficients (or weights) is similar to that in Multiple Regression. Table 4.9 shows that Function 1 has a high positive loading in Module2 (0.958) and a high negative loading in Module4 (-0.161). Module1 and Module3 scores seem to be comparatively insignificant in describing the separation among the categories of student's performance.

Table 4.9: Standardised Coefficients and Structure Matrix

	Function				*Function*		
Standardised Coefficients	1	2	3	Structure Matrix	1	2	3
Module1	-.061	-.068	.258	Module2	.958	.131	.255
Module2	1.041	-.193	.115	Module1	.348	.219	.204
Module3	-.089	1.066	-.215	Module3	.195	.977	.066
Module4	-.260	-.007	.980	Module4	-.161	.141	.951

The structure matrix shows that Function 1 reveals a high correlation between Module2 (0.958) and Module1 (0.348) while Function 2 also reveals that Module3 (0.977) has a high correlation. This finding suggests that there is a strong relationship between the two modules or simply that the performance of students in these two modules is similar. It could either be good or bad performance of the students in the two modules.

Table 4.10: Functions at Group Centroids

	Function		
Groups	1	2	3
2013	.922	.338	-.045
2014	-1.122	.273	-.065
2015	-.040	-.135	.268
2016	.102	-.469	-.136

The standardised coefficients and the structure matrix do not give an indication which of the groups the respective functions discriminate. In order to identify the nature of the discrimination for each discriminant function, one looks at the means for the functions across groups, also known as centroids. The differences in location of the centroids show dimensions along which groups differ, thus projecting how the functions discriminate between groups by plotting a territorial map or a chart for the discriminant functions. Looking at Table 4.10 above, the 2013 group has a relatively high mean (0.922) while the other three groups have relatively low means (-1.122,-0.040 and 0.102) showing discrimination between the groups (2013, 2014, 2015 and 2016). Figure 7.2 also confirms the information in Table 4.10.

Table 4.11: Canonical Discriminant Function Coefficients

	Function		
	1	2	3
Module1	-.005	-.005	.020
Module2	.086	-.016	.009
Module3	-.007	.083	-.017
Module4	-.022	-.001	.083
(Constant)	-3.019	-4.139	-6.192

The discriminant functions in Table 4.11 were used to construct a discriminant model. The scores were as follows:

Function1: $D = -3.019 - 0.005\ Module1 + 0.086\ Module2 - 0.007 Module3 - 0.022 Module4$ (4.1)

Function2: $D = -4.139 - 0.005\,Module1 - 0.016\,Module2 + 0.083\,Module3 - 0.001\,Module4$ (4.2)

Function3: $D = -6.192 + 0.020\,Module1 + 0.009\,Module2 - 0.017\,Module3 + 0.083\,Module4$ (4.3)

4.5.5. Classification of Function Coefficients

The classification table, also known as a confusion table, is one in which rows are the observed categories of the dependent and the columns are the predicted categories. When the prediction is perfect, all cases lie on the diagonal. The percentage of cases on the diagonal resembles the correct classification percentages and the cross validated (also termed 'jack-knife') is a more reliable presentation of the power of the discriminant function than that provided by the original classifications.

Table 4.12: Classification Results

			Predicted Group Membership				
		Groups	1	2	3	3	Total
Original	Count	2013	34	3	8	5	50
		2014	0	33	9	2	44
		2015	8	11	14	11	44
		2016	9	7	5	28	49
	%	2013	68.0	6.0	16.0	10.0	100.0
		2014	.0	75.0	20.5	4.5	100.0
		2015	18.2	25.0	31.8	25.0	100.0
		2016	18.4	14.3	10.2	57.1	100.0
Cross-validated	Count	2013	34	3	8	5	50
		2014	0	32	9	3	44
		2015	9	11	12	12	44
		2016	9	7	5	28	49
	%	2013	68.0	6.0	16.0	10.0	100.0
		2014	.0	72.7	20.5	6.8	100.0
		2015	20.5	25.0	27.3	27.3	100.0
		2016	18.4	14.3	10.2	57.1	100.0

58.3% of original grouped cases correctly classified. 56.7% of cross-validated grouped cases correctly classified.

The classification results displayed in Table 4.12 summarises the association between the actual group (defined as the 'Original') membership and predicted group membership. The table shows that the discriminant model was able to identify and classify 34 out of 50 students who performed well in the 2013 examinations. Thus, it holds 68% classification accuracy of the students who performed within this academic year (2013). On the other hand, the same discriminant model was also able to classify 33 out of 44 students who performed well in the 2014 academic year. This group holds about 75% classification accuracy and 14 out of 44 students in 2015 with a classification accuracy of 31.8%. Lastly, 28 out of 49 students were correctly classified for the 2016 academic year and had a classification accuracy of 57.1%. The cross-validated results showed that 34 out of 50 students were classified in 2013, 32 out of 44 were classified in 2014, 12 out of 44 and 28 out of 49 were classified 2016. In conclusion, the discriminant model was able to generate the student's overall performance of 58.3% classification accuracy in the combined groups. The overall predictive accuracy of the discriminant function is known as the hit ratio and has fairly justified an acceptable goodness of fit by the MDA model.

4.6. Stepwise Discriminant Analysis

This section shows the results from the Stepwise Method. This method starts with a model that does not include any of the predictors (Step 0). At each step, the predictor with the largest *F* to Enter value exceeds the entry criteria (by default, 3.84) is added to the model. The table below shows that Module2 is the first predictor to be entered. The F to remove values are useful for describing what happens if a variable is removed from the current model (given that the other variables remain).

Table 4.13: Variables Not in the Analysis/in the analysis

Step		*Tolerance*	*Min. Tolerance*	*F to Enter*	*Wilks' Lambda*
0	Module1	1.000	1.000	4.363	.933
	Module2	1.000	1.000	30.341	.668
	Module3	1.000	1.000	7.801	.887
	Module4	1.000	1.000	2.299	.964
1	Module1	.827	.827	.314	.664
	Module3	.890	.890	7.122	.598
	Module4	.989	.989	2.969	.637
2	Module1	.779	.779	.055	.597
	Module4	.972	.875	2.674	.572
			Variables in the Analysis		
		Tolerance	F to Remove	Wilks' Lambda	
1	Module2	1.000	30.341		
2	Module2	.890	29.331	.887	
	Module3	.890	7.122	.668	

Tolerance is the proportion of a variable's variance not accounted for by other independent variables in the equation. A variable with very low tolerance contributes little information to a model and can cause computational problems. Therefore Table 4.13 shows that the tolerance values are greater than 0.1 and therefore useful to the model.

4.6.1. Relationship between the dependent and the independent variables

Table 4.14 of eigenvalues shows that Function 1 explains about 81.8% of variance among the four groups while Function 2 only explains 18.2%. Function 1 also has the larger eigenvalue (0.505) than Function 2 indicating a good function and usefulness in distinguishing between the groups. The degree of relationship between the predictors and groups (canonical correlation) attributable to Function 1 is 0.579. The Wilk's lambda reveals that all the predictors add some predictive power to the discriminant function as all are significant with $p < 0.05$

Table 4.14: Eigenvalues & Wilk's Lambda

Function	*Eigenvalue*	*% of Variance*	*Cumulative %*	*Canonical Correlation*
1	.505	81.8	81.8	.579
2	.112	18.2	100.0	.317
		Wilks' Lambda		
Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1 through 2	.598	94.195	6	.000
2	.899	19.432	2	.000

Table 4.15: Canonical Discriminant Function Coefficients

	Function	
	1	2
Module2	.086	-.018
Module3	-.011	.081
(Constant)	-4.469	-4.365

The discriminant functions in Table 4.15 were used and the scores were as follows:

Function1: $D = -4.469 + 0.086\,Module2 - 0.011\,Module3$ (4.4)

Function2: $D = -4.365 - 0.018\,Module2 + 0.018\,Module3$ (4.5)

Table 4.16: Classification Results

			Predicted Group Membership				
		Groups	2013	2014	2015	2016	Total
Original	Count	2013	33	2	7	8	50
		2014	3	33	1	7	44
		2015	10	10	13	11	44
		2016	8	7	8	26	49
	%	2013	66.0	4.0	14.0	16.0	100.0
		2014	6.8	75.0	2.3	15.9	100.0
		2015	22.7	22.7	29.5	25.0	100.0
		2016	16.3	14.3	16.3	53.1	100.0
Cross-validated	Count	2013	33	2	7	8	50
		2014	3	33	1	7	44
		2015	9	11	13	11	44
		2016	9	8	8	24	49
	%	2013	66.0	4.0	14.0	16.0	100.0
		2014	6.8	75.0	2.3	15.9	100.0
		2015	20.5	25.0	29.5	25.0	100.0
		2016	18.4	16.3	16.3	49.0	100.0

56.1% of original grouped cases correctly classified. 55.1% of cross-validated grouped cases correctly classified.

The classification results displayed in Table 4.16 represents the results from the Stepwise discriminant analysis and also summarises the association between the actual group (defined as the 'Original') membership and predicted group membership. The table shows that the discriminant model was able to identify and classify 33 out of 50 students who performed well in the 2013 examinations. Thus, it holds 66% classification accuracy of the students who performed within the 2013 academic year. On the other hand, the same discriminant model was also able to classify 33 out of 44 students who performed well in the 2014 academic year. This group holds about 75% classification accuracy, 13 out of 44 performed well in 2015 while 26 out of 49 students in 2016 were correctly classified. The cross-validated results showed that 33 out of 50 students were classified in 2013, a total of 33 out of 44 classified in 2014. Moreover, 13 out of 44 students were classified in 2015 and 26 out of 49 students were classified in 2016. In conclusion, the discriminant model was able to generate the student's overall performance of 56.1% classification accuracy in combined groups. There is not much improvement in terms of the predictive power from the first classification table.

4.7. Multinomial Logistic Regression Results (MLR)

This section reports and discusses the MLR results. The intention is to construct a multi logit model and use the model to make projections into the future.

4.7.1. Model Estimation

This section used the maximum likelihood estimation method of the beta coefficients to estimate parameters of the MLR model. The Wald test was used to establish whether or not the parameters associated with a group of explanatory variables are zero. The Parameter Estimates table is relevant in showing the logistic coefficient (B) for each predictor variable for each alternative category of the outcome variable. This table identifies some independent variables that are significantly related to the dependent variable and others that are not so strongly associated. The estimates from the parameters obtained through the maximum likelihood estimation method for the final model are summarised in Table 4.17.

Table 4.17: Parameter Estimates

Groups		B	Std. Error	Wald	Df	Sig.	Exp(B)	*95% Confidence Interval for Exp(B)* Lower Bound	Upper Bound
2013	Intercept	-5.708	1.822	9.811	1	.002			
	Module1	-.023	.022	1.147	1	.284	.977	.936	1.020
	Module2	.059	.023	6.370	1	.012	1.061	1.013	1.110
	Module3	.064	.023	7.627	1	.006	1.066	1.019	1.116
	Module4	-.012	.018	.432	1	.511	.988	.953	1.024
2014	Intercept	-.275	2.183	.016	1	.900			
	Module1	.002	.021	.014	1	.907	1.002	.961	1.045
	Module2	-.113	.025	21.310	1	.000	.893	.851	.937
	Module3	.062	.023	7.419	1	.006	1.064	1.018	1.113
	Module4	.035	.023	2.165	1	.141	1.035	.989	1.084
2015	Intercept	-3.506	1.878	3.488	1	.062			
	Module1	.000	.020	.000	1	.996	1.000	.961	1.040
	Module2	-.011	.021	.307	1	.579	.989	.949	1.030
	Module3	.024	.020	1.530	1	.216	1.025	.986	1.065
	Module4	.040	.021	3.678	1	.055	1.041	.999	1.084

The reference category is 2016

Table 4.17 show that for one-unit increase in Module 1 score, we can expect a 0.977 increase in the log-odds in 2013, holding all other independent variables constant. For Module 2 score, we can expect a 1.061 increase in the log-odds in 2013, Module 3 we can expect a 1.066 increase in the log-odds in 2013 and for Module 4, we can expect a 0.988 increase in the log-odds. Furthermore, for one-unit increase in Module 1 score, we can expect a 1.002 increase in the log-odds in 2014, holding all other independent variables constant. For Module 2 score, we can expect a 0.893 increase in the log-odds in 2014, Module 3 we can expect a 1.064 increase in the log-odds in 2014 and for Module 4, we can expect a 1.035 increase in the log-odds. Lastly, for one-unit increase in Module 1 score, we can expect a 1.000 increase in the log-odds in 2015, holding all other independent variables constant. For Module 2 score, we can expect a 0.989 increase in the log-odds in 2015, Module 3 we can expect a 1.025 increase in the log-odds in 2015 and for Module 4, we can expect a 1.041 increase in the log-odds. The fitted model for MLR is therefore obtained as follows:

$$\log it(P_{2013}) = -5.708 - 0.023 Module1 + 0.059 Module2 + 0.064 Module3 - 0.012 Module4 \quad (4.6)$$

$$\log it(P_{2014}) = -0275 + 0.002 Module1 - 0.113 Module2 + 0.002 Module3 + 0.035 Module4 \quad (4.7)$$

$$\log it(P_{2015}) = -3.506 + 0.000\, Module1 - 0.011\, Module2 + 0.024\, Module3 + 0.040\, Module4 \quad (4.8)$$

Alternatively, this can be presented as:

$$P_{2013} = \frac{e^{-5.708-0.023 Module1+0.059 Module2+0.064 Module3-0.012\ Module4)}}{1+e^{-5.708-0.023 Module1+0.059 Module2+0.064 Module3-0.012 Module4)}}\text{, and} \quad (4.9)$$

$$P_{2014} = \frac{e^{-0.275+0.002\ Module1-0.113\ Module2+0.002 Module3+0.035 Module4}}{1+e^{-0.275+0.002\ Module1-0.113\ Module2+0.002\ Module3+0.035\ Module4}} \quad (4.10)$$

$$P_{2015} = \frac{e^{-3.506+0.000\ Module1-0.011\ Module2+0.024\ Module3+0.040\ Module4}}{e^{-3.506+0.000\ Module1-0.011\ Module2+0.024\ Module3+0.040\ Module4}} \quad (4.11)$$

The significance test, in this case the Wald Statistic, is the statistical evidence of the presence of a relationship between the dependent variable and each of the independent variables. The independent variables should be less than or equal to the level of significance of 0.05 for this to hold the statistically significant relationships with the dependent variable.

4.7.2. Assessing the model fit results

This section reviewed the technique employed in assessing the overall fit of the model. This includes obtaining an overall measure of the model using the model fitting information. The coefficients of the model are also tested for statistical significance.

The Model Fitting Information table shows various indices for assessing the intercept only model and the final model which includes all the predictors and the intercept. The table also shows both the AIC and the BIC wherein values show information that is theory based for the model fit statistics. The presence of a relationship between the dependent variable and combination of independent variables is based on the statistical significance of the final model chi-square.

Table 4.18: Model Fitting Information

	Model Fitting Criteria			Likelihood Ratio Tests		
Model	AIC	BIC	-2Log Likelihood	Chi-Square	Df	Sig
Intercept Only	523.817	533.510	517.817			
Final	449.098	497.565	419.098	98.719	12	.000

According to the results shown in Table 4.18, one observes that the -2log likelihood value of the basic model only with intercept term was 517.817. This value decreased to 419.098 with the independent variables considered in the model. Again, Table 4.18 shows that the model is significant at $\chi^2 = 98.719$ (equivalent to significant value of 0.000) with $df = 12$ and $p < .001$. The cut-off value used for this analysis is 0.05. Therefore, this is an indication that there is a relationship between the independent variable and the dependent variables. The results are a further confirmation that the model is a useful predictor.

4.7.3. Assessing the goodness-of-fit of the estimated model

This section evaluated the goodness-of-fit of the LR model. The Pearson's and Deviance statistics were calculated for this purpose. The goodness-of-fit results provide evidence of the model significance or lack thereof. The ideal for these results is to interpret lack of significance as a good fit.

Table 4.19: Goodness-of-Fit

	Chi-Square	*Df*	*Sig.*
Pearson	662.785	543	.000
Deviance	419.098	543	1.000

This table shows whether or not the model adequately fits the data when $p > 0.05$. Table 4.20 shows that $p = 0.000$ which is not ideal because $p < 0.05$. The Cox and Snell and the Nagelkerke R^2 values provide an indication of the amount of variation in the dependent variable. That is, they determine the proportion of variation being explained by the model.

Table 4.20: Pseudo R-Square measurements

Measure	*Value*
Cox and Snell	.410
Nagelkerke	.438
McFadden	.191

Table 4.20 reveals the values 0.410, 0.438 and 0.191 as the measures of the three constructs respectively. These values suggest that about 41% and 44% of variability is explained by this set of variables used in the model. The Nagelkerke (modified form of Cox & Snell coefficient) value among the three (44%) indicate that the model may be good but not a robust predictor.

The likelihood ratio test evaluates the overall relationship between an independent variable and the dependent variable. Table 4.21 gives a summary confirming the contribution of each variable to the model.

Table 4.21: Likelihood Ratio Tests

	Model Fitting Criteria			*Likelihood Ratio Tests*		
Effect	AIC of Reduced Model	BIC of Reduced Model	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	455.792	494.565	431.792	12.693	3	.005
Module1	444.552	483.326	420.552	1.454	3	.693
Module2	503.102	541.875	479.102	60.004	3	.000
Module3	456.538	495.311	432.538	13.440	3	.004
Module4	451.183	489.957	427.183	8.085	3	.044

Table 4.21 shows that Module2, Module3 and Module4 play a significant role in the performance of the students with ($p = 0.000, 0.004\ and\ 0.044$), while Module1 shows no statistically significant relationship with ($p = 0.693$).

4.7.4. Model Validation

The classification table shows how well the full model correctly classifies cases. This table summarizes the observed group and the predicted group classification.

Table 4.22: Classification

	Predicted				
Observed	2013	2014	2015	2016	Percent Correct
2013	34	3	5	8	68.0%
2014	0	33	7	4	75.0%
2015	9	9	10	16	22.7%
2016	10	6	4	29	59.2%
Overall Percentage	28.3%	27.3%	13.9%	30.5%	56.7%

The classification results displayed in Table 4.22 shows that the MLR model was able to classify 34 out of 50 students who performed well in the 2013 examinations. Only a few students (3, 5 and 8) were misclassified in 2014, 2015 and 2016 respectively. Thus, the model holds 68% correct classification accuracy of the students who performed within this academic year (2013). Similarly, the same MLR model correctly classified 33 out of 44 students who performed well in the 2014 academic year. This group holds about 75% classification accuracy. In 2015, only 10 out of 44 students were classified for the academic year and had a classification accuracy of 22.7%. Lastly, 29 out of 49 students were correctly classified in 2016, with a classification accuracy of 59.2%. In conclusion, the average accuracy of the MLR model for the students' performance is 56.7% accurate, suggesting that the model could be somehow useful.

$$Sensitivity = \frac{34}{34+3+5+8} = 0.68 = 68\%, \; Specificity = \frac{33}{0+33++7+4} = 0.75 = 75\%$$

$$= \frac{10}{9+9+10+16} = 0.023 = 23\% \text{ and } = \frac{29}{10+6+4+29} = 0.59 = 59\%$$

$$False\ positive = \frac{10}{10+6+4+29} = 0.20 = 20\%, \; False\ negative = \frac{34}{34+3+5+8} = 0.68 = 68\%$$

4.8. Results from Stepwise MLR

This section reports on the stepwise results. The method used for stepwise was both Forward Entry and Backward Elimination. The results from the tables below are only from the Forward Entry because both methods yielded similar results. The table below shows the step summary of the predictors that were entered in the stepwise.

Table 4.23: Step Summary

			Model Fitting Criteria			Model Fitting Criteria		
Model	Action	Effect(s)	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
0	Entered	Intercept	523.817	533.510	517.817			
1	Entered	Module2	454.047	473.434	442.047	75.769	3	.000
2	Entered	Module3	446.440	475.520	428.440	13.607	3	.003
3	Entered	Module4	444.552	483.326	420.552	7.888	3	.048
Stepwise Method: Forward Entry								

4.8.1. Model Estimation

Table 4.24: Parameter Estimates

								95% Confidence Interval for Exp(B)	
		B	Std. Error	Wald	Df	Sig.	Exp(B)	Lower Bound	Upper Bound
2013	Intercept	-6.365	1.706	13.925	1	.000			
	Module2	.050	.022	5.233	1	.022	1.052	1.007	1.098
	Module3	.055	.022	6.514	1	.011	1.056	1.013	1.102
	Module4	-.009	.018	.288	1	.592	.991	.957	1.025
2014	Intercept	-.144	1.930	.006	1	.940			
	Module2	-.111	.023	22.881	1	.000	.895	.855	.936
	Module3	.061	.022	7.463	1	.006	1.063	1.017	1.110
	Module4	.035	.024	2.167	1	.141	1.035	.989	1.084
2015	Intercept	-3.562	1.777	4.017	1	.045			
	Module2	-.011	.020	.328	1	.567	.989	.951	1.028
	Module3	.024	.018	1.651	1	.199	1.024	.988	1.062
	Module4	.041	.021	3.858	1	.050	1.042	1.000	1.085

Table 4.24 show that for one-unit increase in Module 2 score, we can expect a 1.052 increase in the log-odds in 2013, holding all other independent variables constant. For Module 3 score, we can expect a 1.056 increase in the log-odds in 2013, Module 4 we can expect a 0.991 increase in the log-odds in 2013.Furthermore, for one-unit increase in Module 2 score, we can expect a 0.895 increase in the log-odds in 2014, holding all other independent variables constant. For Module 3 score, we can expect a 1.063 increase in the log-odds in 2014, Module 4 we can expect a 1.035

increase in the log-odds in 2014. Lastly, for one-unit increase in Module 2 score, we can expect a 0.989 increase in the log-odds in 2015, holding all other independent variables constant. For Module 3 score, we can expect a 1.024 increase in the log-odds in 2015, Module 4 we can expect a 1.042. The fitted model for MLR was therefore obtained as follows:

$$\log it(P_{2013}) = -6.365 + 0.005Module2 + 0.055Module3 - 0.009Module4 \quad (4.12)$$

$$\log it(P_{2014}) = -0.144 - 0.111Module2 + 0.061Module3 + 0.035Module4 \quad (4.13)$$

$$\log it(P) = -3.562 - 0.011Module2 + 0.024Module3 + 0.041Module4 \quad (4.14)$$

Alternatively, this can be presented as:

$$P_{2013} = \frac{e^{-6.365+0.050Module2+0.055Module3-0.009Module4}}{1+e^{-6.365+0.050Module2+0.055Module3-0.009Module4}}, \quad (4.15)$$

$$P_{2014} = \frac{e^{-0.144-0.111Module2+0.061Module3+0.035Module4}}{1+e^{-0.144-0.111Module2+0.061Module3+0.035Module4}} \text{ and} \quad (4.16)$$

$$P_{2015} = \frac{e^{-3.562-0.011Module2+0.024Module3+0.041Module4}}{1+e^{-3.562-0.011Module2+0.024Module3+0.041Module4}}$$

4.8.2. Assessing the model fit results

Table 4.25 Model Fitting Information

	Model Fitting Criteria	*Likelihood Ratio Tests*		
Model	-2 Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	517.817			
Final	420.552	97.264	9	.000

According to the results shown in Table 4.25 one observes that the -2log likelihood value of the model only with intercept term was 517.817. This value decreased to 420.552 with the independent variables factored into the model. Again, Table 4.26 shows that the model is significant at χ^2 = 97.264 (equivalent to significant value of 0.000) with $df = 9$ and $p < .001$. The cut-off value used for this analysis is 0.05. Therefore, this is an indication that there is a relationship between the independent variable and the dependent variables. The results are a further confirmation that the model is useful in predicting student success and retention rates.

4.8.3. Assessing the Goodness-of-fit of the estimated model

Table 4. 26 Goodness-of-Fit

	Chi-Square	*Df*	*Sig.*
Pearson	651.424	546	.001
Deviance	420.552	546	1.000

Table 4.26 gives the statistics used for confirming the goodness-of-fit of the model. This table shows that $p = 0.001$ showing that the model has exceeded the significance levels of 1% and 5%.

Table 4.27: Pseudo R-Square

Cox and Snell	.376
Nagelkerke	.423
McFadden	.215

Table 4.27 reveals the values 0.406, 0.433 and 0.188 of the three measures respectively. These values suggest that about 40% and 43% of variability is explained by this set of variables used in the model. The Nagelkerke (modified form of Cox & Snell coefficient) value among the three (43%) indicate that the model may be good but not great.

Table 4.28: Likelihood Ratio Tests

	Model Fitting Criteria			Likelihood Ratio Tests		
Effect	AIC of Reduced Model	BIC of Reduced Model	-2 Log Likelihood of Reduced Model	Chi-Square	Df	Sig.
Intercept	456.908	485.988	438.908	18.356	3	.000
Module2	504.502	533.582	486.502	65.950	3	.000
Module3	451.020	480.100	433.020	12.468	3	.006
Module4	446.440	475.520	428.440	7.888	3	.048

Table 4.28 shows that only Module2, Module3 and Module4 play a significant role in the performance of the students with ($p = 0.000, 0.006$ *and* 0.048).

4.8.4. Model Validation

This section outlines the techniques for model validation. This is to ensure that the model built is a good model or a bad one. The classification table in this case is used to reflect (looking at the overall classification) if the model stands to be good or not.

Table 4.29: Classification

	Predicted				
Observed	2013	2014	2015	2016	Percent Correct
2013	34	3	5	8	68.0%
2014	1	31	8	4	70.5%
2015	9	9	10	16	22.7%
2016	10	7	5	27	55.1%
Overall Percentage	28.9%	26.7%	15.0%	29.4%	54.5%

The classification results displayed in Table 4.29 show that the MLR model was able to classify 34 out of 50 students who performed well in the 2013 examinations. Only a few students (3, 5 and 8) were misclassified in 2014, 2015 and 2016 respectively. Thus, the model holds 68% correct classification accuracy of the students who performed within this 2013 academic year. Similarly, the same MLR model correctly classified 31 out of 44 students who performed well in the 2014 academic year. This group holds about 70.5% classification accuracy and in 2015 only 10 out of 44 students were correctly classified with a classification accuracy of 22.7%. Lastly, 27 out of 49 students were classified for the 2016 academic year and had a classification accuracy of 55.1%. In conclusion, the average accuracy of the MLR model for the students' performance is 54.5%.

4.9.Comparison of the MDA and MLR

This section highlights the criteria used for evaluation of both MDA and MLR. This was achieved by comparing the two techniques using the classification error table.

Table 4.30 Classification Results of MDA and MLR

	MDA	*MLR*
2013	68%	68%
2014	75.0%	75.0%
2015	31.8%	22.7%
2016	57.1%	59.2%
Overall classification	58.3%	56.7%

The classification error criterion is the most frequently used criterion for comparison between the two methods. Table 4.30 shows the classification results of the two statistical methods (MDA and MLR). The results in Table 4.30 indicate that both the MDA and MLR model can correctly classify the first category of 2013 with an accuracy of 68%. The second category, which is 2014, correctly classified 75.0% for both the models. The MDA model for 2015 was 31.8% more than the MLR model which classified at only 22.7%. Lastly, the 2016 category classified that the MLR model was 59.2% and the MDA model was 57.1%. As we can see from the above results, the correct classification rates for all categories (2013-2016) by the MDA model was a bit more efficient than the MLR model. The overall performance of the students was 58.3% and 56.7%.

4.10. Concluding Remarks

This chapter presented the empirical results of both the MDA and MLR discussed in Chapter 3. The purpose of this study was to compare the two most widely used statistical methods for analysing categorical outcome variables. The intention was to explore how the models perform using the marks of the students (offering statistics) in their final year of study. A summary of findings and recommendations for future research are presented in the next chapter.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1. Introduction

This chapter summarises the findings of this study, conclusions are drawn and recommendations are also made for further studies. The rest of the chapter is organised as follows: Section 5.2 entails the discussion with regards the objectives outlined in Chapter 1 and deductions thereof, Section 5.3 outlines the findings and Section 5.4 offers the recommendations based on the findings. Furthermore, Section 5.5 highlights the study limitations while Section 5.6 provides the chapter summary.

5.2. Research objectives and conclusions

This section is aimed at outlining the objectives and provides the summary of the findings obtained in Chapter 4. The following were observed.

Objective 1: To ensure that the data is ready for multivariate analyses

The study proposed two sections in Chapter 3, the preliminary and primary data analysis. The two techniques (MDA and MLR) were applied to the marks of the students between 2013 and 2016 inclusive. In the preliminary data analysis, the diagnostic tests were performed. These included the normality test, homogeneity and multicollinearity tests using the tolerance and VIF. The assumptions for MDA looking at the Kolmogorov-Smirnov and the Shapiro Wilks test were confirmed and the results showed that the data was not normally distributed (except for Module2 only). Although the assumption was violated, the large sample according to Memic (2015) makes this violation not too relevant for further interpretation and classification ability of the model.

Objective 2: To fit MLR and MDA of students' performance

In order to derive the model for MDA, the discriminant canonical discriminant function coefficients were reviewed. The model below shows both the MDA and the stepwise discriminant model. The stepwise discriminant gives a clear indication of the variables that played a significant role in the performance of the students.

Function1: $D = -3.019 - 0.005\ Module1 + 0.086\ Module2 - 0.007 Module3 - 0.022\ Module4$

Function1: $D = -4.469 + 0.086\ Module2 - 0.011\ Module3$

The MLR Model was derived from the parameter estimate tables.

$$P_{2013} = \frac{e^{-5.708-0.023 Module1+0.059 Module2+0.064 Module3-0.012\ Module4)}}{1+e^{-5.708-0.023 Module1+0.059 Module2+0.064 Module3-0.012 Module4)}},$$

$$P_{2013} = \frac{e^{-6.365+0.050 Module2+0.055 Module3-0.009 Module4}}{1+e^{-6.365+0.050 Module2+0.055 Module3-0.009 Module4}},$$

The MLR model showed that the modules that contributed to the building of the model were Module2, Module3 and Module4. This was also validated from the results of the stepwise logistic regression.

Objective 3: To explore the efficiency of MLR and MDA in students' performance

Several tests were conducted to see which model performs better than the other. The tests for MDA included the Wilk's lambda, the canonical discriminant function coefficients and the classification table. The results from the Wilk's lambda showed that only Function 1 and Function 2 were able to discriminate against the groups. Similarly for the MLR, the Pearson and Deviance show if the model fits the data well. In order to get a good indication, the significance value needs to be greater ($p > 0.05$). The results from both the actual (MLR) and the Stepwise (MLR) show that the model was not a good fit ($p = 0.000$). The likelihood ratio test evaluated the overall relationship between the independent variable and the dependent variable. The modules that showed a significant role in the performance of the students was Module2, Module3 and Module4. These modules were validated by their significance values. Parameter estimates were utilised to construct the MLR model.

Objective 4: To identify the model with more predictive power.

In terms of determining the factors that discriminate between the performances of the students, the results of both the MDA and MLR were quite similar. The two techniques showed that both Module2 and Module3 were the most important variables. On the other hand, the MLR pointed out another variable (Module4), where only slight differences were observed. Both MDA and MLR have different models but these models are all formulated with the basic objective of showing the predictive capability of the individual independent variables and also classifying new cases into any of the predefined groups. A reliable model should be able to come up with highly correct classification with less or few misclassification possibilities. Therefore the predictive accuracy for both the techniques was done using the classification table. The overall classification rate was good for both. When using the stepwise approach, both the techniques selected the same variables.

Objective 5: To use the findings of the study in formulating recommendations for further studies and policy purposes

The results of the study through MDA and MLR approach are bound to aid policy makers to enroll more students or not. Judging from the intake of the students through the years (2013-2016), the number seems to fluctuate. This could be due to the lack of interest in studying statistics or the stereotype that the course itself is more difficult than other fields of study.

With statistics being a scarce skill, more still needs to be done. The problem starts at the lower levels because not many people know what a statistician's role in society is, therefore more awareness campaigns and marketing are critical to inform students starting from high school of the possible career options when one chooses statistics field. This study could assist the university in establishing relationships with both the private and public sector to ensure that students are enticed to enroll for a degree in statistics. This could also motivate students to finish their degrees and enroll more for postgraduate studies so that these postgraduates could be absorbed by the different institutions brought aboard by the university.

5.3. Findings of the study

The current study evaluated the students' performance at the North-West University by conducting a comparison between MDA and MLR. The analysis took into consideration the objectives as outlined in Chapter 1 and a quantitative research approach was used to conduct the study. The variables used in this study were the modules that the students had to take for the completion of their degrees. The primary analysis was conducted and the results from the normality test in Table 4.3 revealed that only one variable was normally distributed (Module2) while the other three (Module1, Module3 and Module4) were not normally distributed. The Box's M test and the Wilk's lambda were used to confirm the equality of the covariance matrices and the significance of the canonical correlation. The study also reviewed the multicollinearity test which resulted in the conclusion that there was no problem of multicollinearity since the tolerance values were greater than 0.1 and the VIF values were smaller than 10.

In order to see which variables contribute a more significant role in the discriminant function, Table 4.7 of the equality of group means was employed. The interpretation of the discriminant functions included the use of standardised coefficients to identify which independent variable discriminates more than the other variables and the structure matrix which provides information about the correlation between each discriminating variable and a discriminant function. The group centroids were used to identify the nature of the discrimination for each discriminant function. The canonical discriminant functions in Table 4.11 was employed to construct a discriminant model. And lastly, the classification table was used to assess the cases that were correctly and incorrectly classified, giving the overall percentage of the students' performance of 58.3%.

For the MLR results, assessing the model using the model fitting information which included the AIC/BIC and the chi-square was vital in establishing if the model was significant and useful. The results for assessing the goodness-of-fit were presented in Table 4.19 and variability was explored using the Pseudo R-squared table. The likelihood ratio tests in this study were utilised to evaluate the overall relationship between the independent variable and the dependent variable and showed that only three variables (Module2, Module3 and Module4) played a significant role in the students' performance. In order to build a fitting model, the parameter estimates were utilised. The classification table, again similar to MDA, was used to classify cases and give the overall percentage of the students' performance (56.7%).

The study further performed the stepwise method for both the MDA and MLR. This method was applied to identify the most important variables and their contribution to the model step by step.

5.4. Recommendations

The results from the classification table show that the two techniques do not differ much in classification. The results obtained from MDA and MLR indicate that the two techniques gave almost the same percentage of correct classification and identified the same parameters responsible for discriminating the performance of the students. The MDA was able to give the overall classification of 58.3% and MLR was 56.7%. Looking at the overall classification, the margin of difference was 1.6%, and logically, one would conclude that both techniques converged in reaching similar results. The findings in this current study confirm and are consistent with the findings of previous studies done by Antonogeorgos et al. (2009), Balogun et al., (2015) and Alkarkhi & Easa (2008). Therefore one concludes that these techniques can be recommended for future studies and can be explored more in education.

5.5. Limitations of the study

There are several classification techniques used for studying the performance of students within institutions of higher learning. Artificial neural network (ANN) is one such technique that could be used for prediction of students' performance. The study conducted by Mansour et al. (2013) showed that ANN was used together with DA and LR. However, this current study's scope was limited only to MDA and MLR as classification techniques. Even though these techniques were used to achieve similar objectives, not all the assumptions were inferred from the techniques.

5.6. Summary

This chapter provided the summarised conclusions and recommendations of this study with specific reference to the objectives and research problem. The main purpose was to compare the two models and their predictive powers by using different tests and measures. The results in Chapter 4 showed that the two techniques converged in reaching similar predictive results. The question of whether the application of MDA and MLR on the same data set would yield the same result, or independent variables of the same data set would be significant in the cases is clarified by the empirical analysis. Therefore with all the tests conducted in Chapter 4, the overview of this research confirmed that these two techniques are well-suited for this study and could be used comparatively.

References

Abledu, G. K., Buckman, A., Adade, T., Kwofie, S. 2016. Comparison of logistic regression and linear discriminant analyses of the determinants of financial sustainability of rural banks in Ghana. *American Journal of Theoretical and Applied Statistics*, Vol. 5(2): 49-57.

Afolabi, J.A. 2008. Analysis of loan repayment among small scale farmers in South Western Nigeria, a discriminant approach. *Journal of Social Science*, Vol. 17(1), 83-88.

Agresti, A. 2007. *Categorical Data Analysis*, 3rd Edition. John Wiley & Sons Inc., New York.

Agresti, A. 2013. *Categorical Data Analysis*, 3rd Edition. John Wiley & Sons Inc., New York.
Akomolafe, A.A., 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*, Vol 16(1), 201-205.

Albayrak, A.S. 2009. Classification of domestic and foreign commercial banks in turkey based on financial efficiency, a comparison of decision tree, logistic regression and discriminant analysis models. *The Journal of faculty of economics and administrative sciences*, Vol. 14(2), 113-139.

Al-Jazzar, M.F. 2012. A comparative study between linear discriminant analysis and multinomial logistic regression in classification and predictive modeling. A thesis submitted to Al-Azhar University in Gaza.

Alkarkhi, F.M., Easa A.M. 2008. Comparing discriminant analysis and logistic regression model as a statistical assessment tools of arsenic and heavy metal contents in cockles. *Journal of sustainable development*, Vol 1 (2), 102-106.

Antonogeorgos, G., Panagiotakos, D.B., Priftis, K. N., Tzonou, A. 2009. Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10 to 12 years old children, divergence and similarity of the two statistical methods. *International Journal of Pediatrics*, 1-6.

Aromolaran, Adeyemi, D., Oyeyinka, Isaiah, K., Olukotun, Oluseyi, O., Benjamin, E. 2013. Binary logistic regression of student's academic performance in tertiary institution in Nigeria by socio-demographic and economic factors. *International Journal of Engineering Science and Innovative Technology*, Vol. 2(4), 590-596.

Balogun, O.S., Akingbade, T.J., Oguntunde, P.E., 2015. An assessment of the performance of discriminant analysis and the logistic regression methods in classification of mode of delivery of an expectant mother. *Mathematical Theory and Modeling*, Vol. 5(5), 147-154.

Balogun, O.S., Balogun, M.A., Abdulkadir, S.S., Jibasen, D. 2014. A comparison of the performance of the discriminant analysis and the logistic regression methods in classification of drug offenders in Kwara State. *International Journal of Advanced Research*, Vol. 2(10), 280-286.

Balogun, O.S., Oyejola, B.A. And Akingbade, T.J. 2012. Use of discriminant analysis in classification of drug peddlers and non-drug peddlers in Kwara state. *International Journal of Engineering Research and Application*, 2(5), 936-938.

Barnard, M.M., 1935. The secular variations of skull characters in four series of Egyptian skulls. *Annals of Human Genetics*, *6*(4), pp.352-371.

Bayaga, A. 2010. Multinomial logistic regression, usage and application in risk analysis. *Journal of Applied Quantitative methods*, Vol.5 (2). 288-297.

Betensky, A., Williams, P.L. 2001. A comparison of models of clustered binary outcomes: analysis of a designed immunology experiment. *Applied Statistics*, Vol. 50(1), 43-61.

Bull, S., Donner, A. 1987. The efficiency of multinomial logistic regression compared with multiple group discriminant analysis. *Journal of the American Statistical Association*, Vol. 82(400), 1118-1122.

Burns, R., Burns, R. 2008. Business Research Methods and Statistics using SPSS, Chapter 24.

Burtner, J. 2005. The use of discriminant analysis to investigate the influence of non-cognitive factors on engineering school persistence. *Journal of Engineering Education*, Vol. 94(3), 335-338.

Cizek, G.J., Fitzgerald, S.M., 1999. Methods, plainly speaking: An introduction to logistic regression. *Measurement and evaluation in counseling and development*, *31*(4), 223.

Clark, W. H., Elder, D.E., Guerry, D., Braitman, L.E., Trock, B.J., Schultz, D. 1989. Model for predicting survival in stage I melanoma based on tumor progression. *Journal of the National Cancer Institute*, Vol. 81(24), 1893-1904.

Court, P.W., Radloff, S.E. 1990. A comparison of multivariate discriminant and logistic regression analysis in the prediction of corporate failure in South Africa. *De Ratione*, Vol. 4(2), 11-15.

Cox, D.R., 1958. The regression analysis of binary sequences. Journal of the Royal Statistical Society, Series B (Methodological), 215-242.

Cox, D.R., 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.215-242.

Cramer, J.S. 2002. The origins of logistic regression. University of Amsterdam and Tinbergen Institute, Faculty of economics and econometrics, Amsterdam, 119/4.

Crask, M.R. and Perreault Jr, W.D., 1977. Validation of discriminant analysis in marketing research. *Journal of Marketing Research*, pp.60-68.

Crask, M.R., Perreault Jr, W.D., 1977. Validation of discriminant analysis in marketing research. *Journal of Marketing Research*, Vol.14 (1), 60-68.

Dey, E.L., Astin, A.W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in higher education*, Vol 34(5), 569-581.

Dohoo, I.R., Ducrot, C., Fourichon, C., Donald, A. And Hurnik, D. 1997. An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive veterinary medicine*, Vol 29(3), 221-239.

Easa, A. M., Alkarkhi, A.F.M. 2008. Comparing discriminant analysis and logistic regression model as a statistical assessment tool of arsenic and heavy metal contents in cockles. *Canadian Center of science and Education*, Vol. 1(2), 102-106.

Elgohari, H. 2017. Efficiency of discriminant analysis and multivariate logistic regression for the detection of anemic children with chronic kidney disease. *International Journal of Statistics and Applications*, Vol 7(2), 131-136.

El-Habil, A., & El-Jazzar, M. 2014. A comparative study between linear discriminant analysis and multinomial logistic regression. *An-Najah University Journal of Research, Humanities*, Vol. 28(6), 1525-1548.

Erimafa, J.T., Iduseri, A., Edkopa, I.W. 2009. Application of discriminant analysis to predict the class of degree for graduating students in a university system. *International Journal of Physical Sciences*, Vol. 4, 016-021.

Frank, R.E., Massy, W.F., Morrison, D.G., 1965. Bias in multiple discriminant analysis. *Journal of Marketing Research*, Vol.2 (3), 250-258.

Gabor, M.R., Maniu, A.I. 2011. Analysis of the discriminating, applications identifying the preference of endowment with goods. *Romanian statistical review 10*.

Garson, D. 2012. Testing statistical assumptions, School of Public and International Affairs. North Carolina State University.

Garton, B.L., Ball, A.L., Dyer, J.E. 2002. The academic performance and retention of college of agriculture students. *Journal of Agricultural Education*, Vol 43(1), 46-56.

Geller, A. C., Johnson, M, D., Miller, D. R., Brooks, K.R., Layton, C. J., Susan, M. 2009. Factors associated with physical discovery early melanoma in middle-aged and older men. *Arch Dermatol*, Vol 145, 409-414.

Ghasemi, A., Zahediasl, S. 2012. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians International. *Journal of Endocrinology and Metabolism*, Vol 10(2), 486-489.

Gordon, D. J., Probstfield, J.L., Rubenstein, C., Bremner, W. F., Leon, A.S., Karon, J.M. 1984. Coronary risk factors and exercise test performance in asymptomatic hypercholesterolemic men: Application of Proportional Hazards Analysis. *American Journal of Epidemiology*, Vol. 120(2), 210-224.

Gwary, M., Gwary, T., Mustapha, B. 2012. Discriminant analysis of the influence of Farmers socio-economic characteristics on their participation in research and extension activities in Borno State, Nigeria. *International Research Journal of Social Sciences*, Vol. 1(4), 1-6.

Hair, J, Jr., Black, W., Babin, B., Anderson, R., Tatham, R. 2010. *Multivariate Data Analysis*. Pearson Education, Inc., Upper Saddle River, New Jersey.

Haq, M.A., Dar, I.S., Qura-Tul-Ain. 2015. Performance comparison of classification techniques artificial neural network, discriminant analysis and logistic regression, application establish more private academies or not. *Science International (Lahore)*, Vol. 27(3), 1803-1807.

Hosmer, D. W., Lemeshow, S. 2000. *Applied Logistic Regression, 2nd Edition*. Wiley and sons Inc. New York.

http://ncss.wpengine.netdna.cdn.com/wpcontent/themes/ncss/pdf/Procedures/NCSS/Logistic_Regression.pdf, chapter 320.

Huberty, C., Olejnik, S. 2006. Applied Manova and Discriminant Analysis, 2nd Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

Ito, T., Nishimura, S., Saito, M., Omori, Y. 1997. The level of erythrocyte aldose reductase: A risk factor for diabetic neuropathy. *Diabetes research and clinical practice*, Vol. 36(3), 161-167.

Izenman, A.J., 2013. Multivariate regression. In Modern Multivariate Statistical Techniques, 159-194. Springer, New York, NY.

Jihad A. A. Al Shamali. 2015. Using linear discriminant analysis and multinomial logistic regression in classification and prediction. A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Applied Statistics Al-Azhar University – Gaza, Palestine.

Johnson, R.A., Wichern, D.W 2007. Applied multivariate statistical analysis. Prentice Hall International. INC., New Jersey.

Kabakchieva, D. 2013. Predicting student performance by using data mining methods for classification. Bulgarian Academy of Sciences. *Cybernetics and Information Technologies*, Vol. 13(1), 61-72.

Kasper, E. 2006. Internal research and development markets. Physica-Verlag, a Springer Company, Germany.

Kennedy, J.W., Kaiser, G.C., Fisher, L.D., Maynard, C., Fritz, J. K., Myers, W. 1980. Multivariate discriminant analysis of the clinical and angiographic predictors of operative mortality from the Collaborative study in coronary artery surgery. *The Journal of Thoracic and Cardiovascular surgery*, Vol. 80, 876-887.

Kim, K.H., Bentler, P.M. 1999. Tests of Homogeneity of means and covariance matrices for multivariate incomplete data. *Paper presented at the Western Psychological Association convention, Irvin, USA.*

Kirschenbaum, M., Oigenblick, K., Goldberg, S. 2000. Analysis of the predictors or work accident proneness. *Chest*, 118-125.

Kisaka-Lwayo, M. 2007. A discriminant analysis of factors associated with the adoption of certified organic farming by small holder farmers in Kwazulu-Natal, South Africa. *African association for agricultural economists' conference proceedings*, 411-416.

Kislaya, I. 2012. Statistical multivariate approaches for identification of predictors of academic failure for first year students in medical school. Dissertation submitted to the University of Minho.

Klecka, W.R., 1980. Discriminant Analysis. Beverly Hills, CA: Sage.

Kleinbaum, D., Kupper, L., Muller, K., Nizam, A. 1998. Applied Regression Analysis and other Multivariable methods, 3rd Edition. Duxbury Press.

Kong-Lai, Z., Jing-Jing, L. 2010. Studies of discriminant analysis and logistic regression model application in credit risk for Chinas listed companies. *Management Science and Engineering*, Vol. 4(4), 24-32.

Lattin, J.M., Carroll, J.D. And Green, P.E., 2003. Analysing multivariate data, 206-63. Pacific Grove, CA: Thomson Brooks/Cole.

Lei, P.W. And Koehly, L.M. 2003. Linear discriminant analysis versus logistic regression: A comparison of classification errors in the two-group case. *The Journal of Experimental Education*, Vol. 72(1), 25-49.

Lewicki, P., Hill, T. 2006. Statistics: Methods and Applications Comprehensive References for Science, Industry and Data Mining. StaSoft, Inc.

Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, Vol. 62(318), 399-402.

Lin, J.J.J., Imbrie, P.K., Reid, K. 2009. Student retention modeling, an evaluation of different methods and their impact on prediction results. *Proceedings of the research in Engineering education symposium, Palm Cove, QLD*, 1-6.

Liong, C., Foo, S. 2013. Comparison of Linear Discriminant Analysis and Logistic Regression for Data Classification. *Proceedings of the 20th National Symposium on Mathematical Sciences*, *AIP Conf. Proc.* 1522, 1159-1165.

Luna, J. 2000. Predicting student retention and academic success at New Mexico Tech. Doctoral dissertation, New Mexico Institute of Mining and Technology.

Mahalanobis, P.C., 1930. A statistical study of certain anthropometric measurements from Sweden. *Biometrika*, pp.94-108.

Mahfouz, M.S. 2016. Logistic Regression Models to Predict Factors Associated with Khat Abuse among Students in Jazan Region, Saudi Arabia. *Electronic Journal of Biology*, Vol.12 (1), 1-7.

Maiprasert, D., Kitbumrungat, K. 2012. Comparison Multinomial Logistic Regression and Discriminant Analysis in predicting the stage of Breast Cancer. *International Journal of Computer Science and Network Security*, Vol. 12(4), 44-52.

Mansour, R., Zandkarimi, E., Amirhossein, H. 2013. Comparison of artificial neural network, logistic regression and discriminant analysis efficiency in determining risk factors of Type 2 diabetics. *World Applied Sciences Journal*, Vol. 23(11), 1522-1529.

Marshall, C., Gretchen B. R. 1999. Designing qualitative research. 3rd ed. London, Sage Publications.

Memic, D. 2015. Assessing credit default using logistic regression and multiple discriminant analysis, empirical evidence from Bosnia and Herzegovina. *Interdisciplinary description of complex systems*, Vol. 13(1), 128-153.

Mertler, C., Vannatta, R. 2002. Advanced and Multivariate Statistical Methods. Practical Application and Interpretation, 2nd Edition. Pyrczak Publishing, Los Angeles, CA.

Mertler, C., Vannatta, R. 2013. Advanced and Multivariate Statistical Methods. Practical Application and Interpretation, 5nd Edition. Pyrczak Publishing, Los Angeles, CA.

Metcalfe, Y. 2012. A logistic regression and discriminant function analysis of enrollment characteristics of student veterans with and without disabilities. VCU Theses and Dissertations, paper 2755.

Mihalovic, M. 2016. Performance Comparison of Multiple Discriminant Analysis and Logit Models in Bankruptcy Prediction. *Economics & Sociology*, Vol. 9(4), 101-118.

Montgomery, D.B., 1975. New product distribution: An analysis of supermarket buyer decisions. *Journal of Marketing Research*, Vol. 12(3), 255-264.

Montgomery, M.E., White, M.E., Martin, S.W. 1987. A comparison of discriminant analysis and logistic regression for the prediction of coliform mastitis in dairy cows. *Canadian Journal of Veterinary research*, Vol. 51(4), 495-498.

Morant, R., 1920. A Criticism of the Report of the Committee of the Council of the Section of Obstetrics and Gynecology of the Royal Society of Medicine upon the Teaching of Obstetrics and Gynecology in London. *Proceedings of the Royal Society of Medicine*, *13*(Obstet_Gynaecol), pp.59-61.

Mordkoff, J.T., 2011. The assumption (s) of normality.

Morrison, D.G., 1969. On the interpretation of discriminant analysis. *Journal of marketing research*, Vol. 6(2), 156-163.

Muchabaiwa, H. 2013. Logistic Regression to determine significant factors associated with share price change. Thesis submitted to University of South Africa (UNISA).

Normadiah, M.R., Yap, B.W. 2011. Power comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling test. *Journal of statistical modelling and analysis*, Vol. 2(1), 21-23.

Originlab corporation, O. (2016). Interpreting Results of Discriminant Analysis. Available at: http://www.originlab.com/doc/Origin-Help/DiscAnalysis-Result, Accessed 6 Dec 2016.

Osita, A. E. 2010. On the use of Logistic Regression and Linear Discriminant Analysis in classification problems (A Comparative study). A project submitted to the NnamdiAzikiwe University, Awka, Nigeria.

Panagiotakos, D.B. 2006. A comparison between logistic regression and linear discriminant analysis for the prediction of categorical health outcomes. *International Journal of Statistical Sciences*, Vol. 5, 73-84.

Park, Hyeoun-Ae. 2013. An Introduction to logistic Regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, Vol. 43(2), 154-164.

Pearson, K., 1920. The fundamental problem of practical statistics. *Biometrika*, *13*(1), pp.1-16.

Peng, C., Lee, K., Ingersoll, G. 2002. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, Vol. 96(1), 3-14.

Pohar, M., Mateja, B., Turk, S. 2004. Comparison of logistic regression and linear discriminant analysis: A simulation study. *Metodoloski zvezki*, Vol 1(1), 143-161.

Prempeh, E.A. 2009. Comparative study of the logistic regression analysis and the discriminant. Thesis submitted to the University of Cape Coast.

Press, S. J. And Wilson, S. 1978. Choosing between logistic regression and discriminant analysis. *Journal of the American statistical Association*, Vol. 73, 699-705.

Quetelet, A. (1848). De syst`eme social et des lois qui le r´egissent. Paris: Guillaume.

Quetelet, A. (1850). Notice sur Pierre-Fran¸cois Verhulst. Annuaire de l'Acad´emie Royale des Sciences, Lettres et des Beaux-arts 16, 97–124.

Rao, C.R., 1948. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), pp.159-203.

Rencher Alvin C. 2002. Methods of multivariate analysis, 2nd edition, Brigham Young University, John Wiley & sons, Inc. publication.

Sanabila, H. R., Fanany, M.I., Jatmiko, W., Murni, A. 2010. Bootstrapped Multinomial Logistic Regression on Apnea detection using ECG data. *Advanced Computer Science and Information Systems*, 181-186.

Schreiner, L.A. 2009. Linking student satisfaction and retention. Noel-Levitz, Coralville, IA.

Şen, A.B., 2010. Factors that Discriminate Between Domestic and Foreign Banks Operating in Turkey. *İktisadi ve İdari Bilimler Dergisi*, Vol. 28(1), 445-462.

Smith, G.C.S. 2005. A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. *American Journal of Epidemiology*, Vol. 161, 100-102.

Stevens, J. 2001. Applied multivariate statistics for the social sciences, 4^{th} ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sule, B.O., Saporu, F.W.O. 2015. A Logistic Regression Model of Students' Academic Performance in University of Maiduguri, Maiduguri, Nigeria. *Mathematical theory and modelling*, Vol. 5(10)124-136.

Suleiman, S., Suleiman, I., Usman, U., Salami, Y.O. 2014. Predicting an applicant status using principal component, discriminant and logistic regression analysis. *International Journal of Mathematics and Statistics Invention*, Vol. 2(10), 5-15.

Tabachnick, B.G., Fidell, L.S. 1996. Using multivariate statistics, 3rd edition. Pearson Education, Inc.

Tabachnick, B.G., Fidell, L.S. 2007. Using multivariate statistics, fifth edition. Pearson Education, Inc.

Takahashi, Y., Tachikawa, T., Ito, T., Takayama, S., Omori, Y., Iwamoto, Y. 1998. Erythrocyte aldose reductase protein: a clue to elucidate risk factors for diabetic neuropathies independent of glycaemic control. *Diabetics Research and Clinical Practice*, Vol. 42(2), 101-107.

Tangen, R. 2008. Listening to children's voices in educational research: Some theoretical and methodological problems. *European Journal of Special Needs Education*, Vol. 23(2), 157–166.

Tektas, N. 2014. Classification performance comparison of discriminant analysis and logistic regression analysis. *International periodical for the languages, literature and history of Turkish or Turkic*, Vol. 9(2), 1517-1527.

Teshnizi, S.H., Ayatollahi, S.M.T. 2015. A Comparison of Logistic Regression Model and Artificial Neural Networks in Predicting of Student's Academic Failure. *Acta Informatica Medica*, Vol. 23(5), 296-300.

Thammasiri, D., Delen, D., Meesad, P., Kasap, N. 2014. A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, Vol. 41(2), 321-330.

Thomas, E.W., Marr, M.J., Thomas, A., Hume, R.M. and Walker, N., 1996. Using discriminant analysis to identify students at risk. In *Frontiers in Education Conference, 26th Annual Conference Proceedings. Institute of Electrical and Electronics Engineers*, Vol. (1), 185-188.

Tu, Y.K., Kellett, M., Clerehugh, V., Gilthorpe, M.S. 2005. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. *British dental journal*, Vol. 199(7), 457-461.

Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. Correspondance Math´ematique et Physique, publi´ee par A. Quetelet 10, 113.

Vuran, B. 2009. Prediction of business failure, a comparison of discriminant and logistic regression analysis. *Istanbul University Journal of the school of business administration*, Vol. 38(1), 47-65.

Wu, S., 2010. Goodness-of-fit tests for logistic regression. A dissertation submitted to the department of statistics in partial fulfillment of the requirements for the degree of Ph.D. The Florida state university, college of arts and sciences.

Yarnold, P.R., Hart, L.A., Soltysik, R.C. 1994. Optimizing the classification performance of logistic regression and Fisher's discriminant analyses. *Educational and Psychological Measurement*, Vol. 54(1), 73-85.

Yeboah, B.E. 2012. Predicting microfinance credit default. A Dissertation submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology.

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q.P., Lillard Jr, J.W. 2014. A study of effects of multicollinearity in the multivariable analysis. *International journal of applied science and technology*, Vol. 4(5), 9.

Zeger, S, L., Diggle, P. J., Liang, K. 2004. A cox model for biostatistics of the future. Johns Hopkins University, Department of biostatistics working papers, paper 32.

Appendices

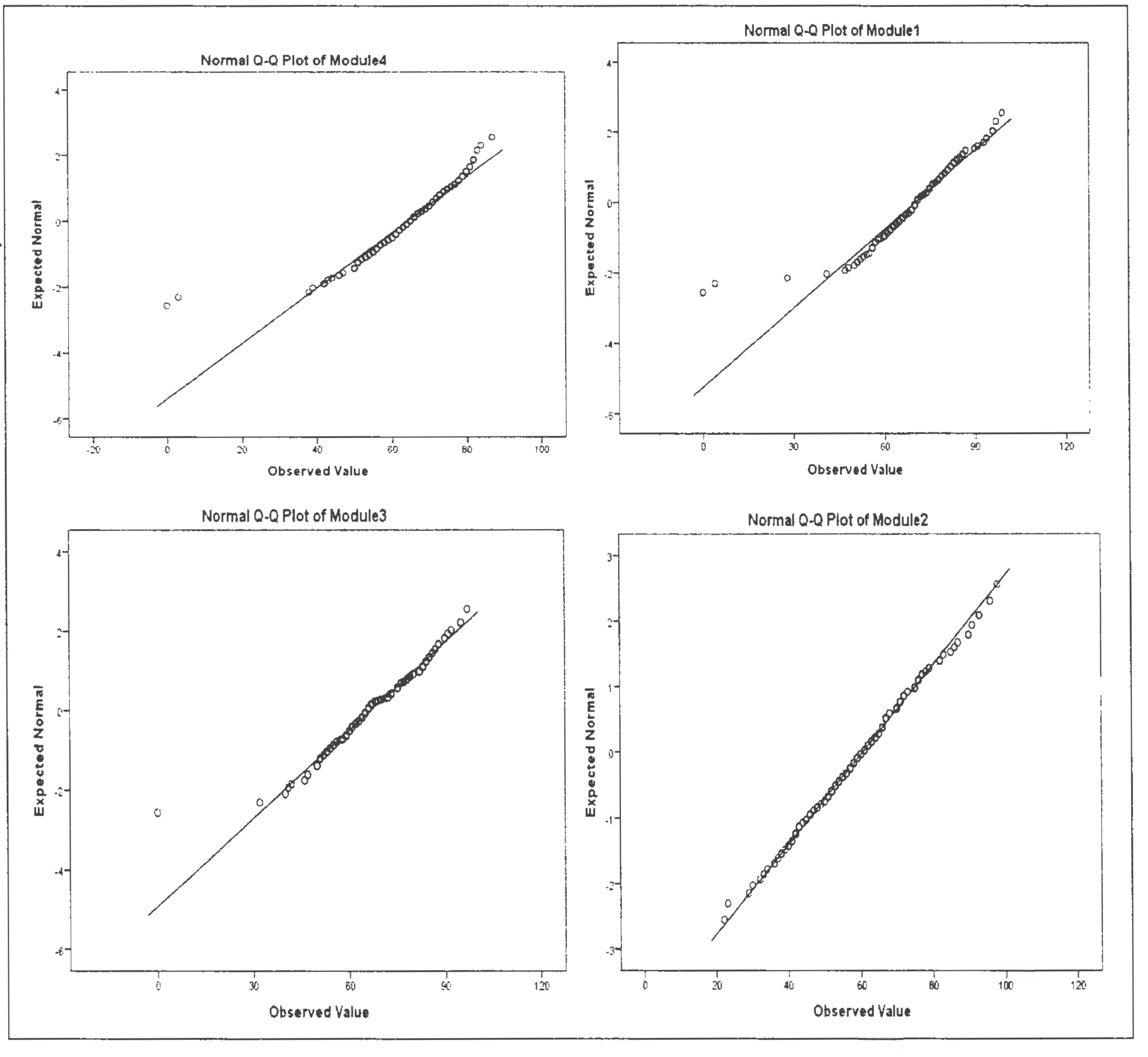

Figure 7.1: Normal Q-Q plots

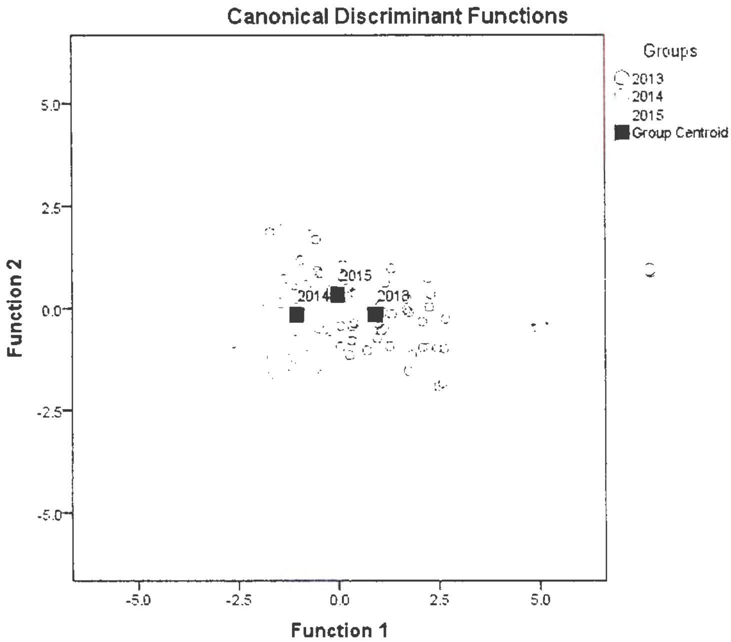

Figure 7.2: Territorial Map