



UNIVERSITY OF NAIROBI

SCHOOL OF MATHEMATICS

**PREDICTING STUDENT' S LOAN DEFAULT IN KENYA: FISHER'S
DISCRIMINANT ANALYSIS APPROACH**

By

Mwangi Johnson Muthii

This research project is submitted to the School of Mathematics of the University of Nairobi in partial fulfillment of the requirement for the degree of Masters of Science in Social Statistics.

© JUNE 2015.

DECLARATION

I, the undersigned, declare that this project is my original work and has not been presented for an academic credit in any other university.

Signature:

Date: ____/____/2015

Mwangi Johnson Muthii

Reg. No: I56/68336/2013.

This thesis has been submitted for examination with my approval as the University supervisor

Signature:

Date: ____/____/2015

Dr. Nelson Owuor Onyango

University of Nairobi

DEDICATION

I dedicate this project to my dad (David Mwangi), my mother (Esther Karuana), and my sister (Eunice Mwangi) for their unconditional love and support. I also extend my sincere gratitude to my friends for the moral support and above all, to the Almighty God for good health and sound mind throughout my studies.

ACKNOWLEDGEMENT

I am thankful to the Almighty God for the good health and wellbeing that enabled me to complete this thesis.

I am extremely grateful and indebted to my supervisor Dr. Nelson Owuor Onyango, for his continuous guidance, patience, motivation and immense knowledge that he accorded towards me during project writing.

I also wish to extend my sincere gratitude to HELB recovery department officials, for providing me with the data for use in my project.

Special thanks goes to my family for their unceasing encouragement and support. I also thank my fellow class mates, for the simulating discussions, for the late evenings and weekends we worked together to beat deadlines and for the fun we had for the last two years of our course work.

Last but not least, I express my sense of gratitude to one and all, who directly or indirectly have extended their support in this venture.

TABLE OF CONTENT

DECLARATION	i
DEDICATION	ii
ACKNOWLEDGEMENT	iii
LIST OF TABLES	vi
LIST OF ABBREVIATIONS	vii
ABSTRACT.....	viii
CHAPTER ONE: INTRODUCTION.....	1
1.0 Background to the study.....	1
1.2 Statement of the problem	2
1.4 Objectives.....	3
1.4.1 Main Objectives.....	3
1.4.2 Specific Objectives	3
1.5 Justification	3
CHAPTER TWO: LITERATURE REVIEW	4
2.0 Introduction	4
2.1 Student Demographics	4
2.1.1 Age.....	4
2.1.2 Gender	4
2.1.3 Marital status	4
2.2 Socio-economic status.....	5
2.2.1 Number of dependents.....	5
2.2.2 Parent income	5
2.3 Academic Experience.....	5
2.3.1 Major	5
2.3.2 Loan debt	6
2.4 Post Collegiate Experience.....	6
2.4.1 Employment.....	6
2.5 Conceptual frame work	6
CHAPTER THREE: METHODOLOGY	8
3.0 Introduction	8
3.1 Sources of data	9

3.2 Variables Description	10
3.4 Selection of predictor variables	10
3.5 Analytical Technique	11
3.5.1 Discriminant function	11
3.5.2 Fisher's linear discriminant functions	12
5.4 Allocation rule	14
3.6 Assumptions	15
3.6.1 Equal variance –Covariance matrix	15
3.6.2 Multicollinearity	15
3.6.3 Linearity.....	15
3.6.4 Absence of outliers	16
3.6.5 Sample size	16
3.7 Procedures of running discriminant analysis	16
3.7.1 Evaluating the group centroids difference.....	16
3.7.2 Box's test of equality of covariance matrices.....	17
3.7.3 Eigen value	17
3.7.4 Canonical correlation.....	17
3.7.5 Wilk's lambda.....	18
3.8 Classification.....	18
3.8.1 T-test for equal size groups in dependent variable	19
3.8.2 Press's Q for Uneven sample groups.....	19
CHAPTER FOUR: SURVEY FINDINGS	20
4.1 Descriptive Statistics	20
4.2 Discriminant analysis findings	22
4.2.1 Test of equality of Covariance matrices by using Box's M	22
4.2.2 Significance of the discriminant function.....	23
4.2.3 Fisher's discriminant function	25
4.2.4 Group Centroids	26
4.2.5 Assessing the internal validity of the model.....	26
4.2.6 Press's Q test for unequal sample groups	27
CHAPTER FIVE: DISCUSSION, CONCLUSION AND RECOMMEDATIONS	28
REFERENCE.....	31

LIST OF TABLES

Table 1: Percentage description of categorical variables.....	21
Table 2: Means and standard deviations of the continuous variables.....	22
Table 3: Box's M test of equality of covariance of matrices	23
Table 4: Significance of the discriminant function.....	24
Table 5: Fisher's discriminant function summary output.....	25
Table 6: Classification matrix.....	27
Table 7: Classification accuracy as compare to chance.....	27

LIST OF ABBREVIATIONS

HELB - Higher Education Loan Board

HELF - Higher Education Loan Fund

USA - United States of America

USSR - Union of Soviet Socialist Republics

LDA - Linear Discriminant Analysis

FLDA - Fisher's Linear Discriminant analysis

OLS - Ordinary Least squares

TSC - Teacher Service Commission

RAND - Random number

ABSTRACT

The current high student enrollments in Kenyan universities has outstretched HELB in terms of loan disbursement. High default rate which stood at 43% as at 2012/2013 financial year have been the major challenge to HELB in meeting its core mandate of disbursing loans, scholarship and bursaries to needy students who have qualified to join local universities. The goal of this study was to develop a student loan default model that can predict if a new loan applicant is likely to be a defaulter or non-defaulter. This study examines characteristics of 7,354 loan borrowers from HELB between year 2009 and 2013. The study predictors were; age, gender, marital status, dependence, degree major, employment, loan awarded, family income, and bursary application, while the outcome variable was loan status (default or non-default).

The findings showed that, employment status had the greatest discriminatory power in classifying the borrowers. This was followed by age, degree major (education), bursary application and gender in that order. The predicted model explained 36 percent of the variance in the discriminant function. In addition, the developed model was able to correctly classify 77 percent of the loan borrowers as either defaulters or non- defaulters. Interventions that would focus on the success of the student after college were seen as the main actions that would curb loan default.

CHAPTER ONE: INTRODUCTION

1.0 Background to the study

The increase in dependence on student's loans is a function of several factors, partly the rising price of attending college, which has outpaced inflation rates and median family income levels for at least a decade (Hearn & Holdsworth, 2004). In addition, new growth in student volume has put upward pressure on the student loan system (Katz et al., 2012). Educational loans schemes operate in about 70 countries around the world (Shen & Ziderman, 2009). According to Ziderman (2004), loans schemes differ across countries in terms of the underlying objectives, organizational structure, and sources of initial funding, student coverage, loan allocation procedures and collection methods. However, the similarities of all schemes across the countries are that they are highly subsidized by governments (Ismail, 2011).

Student's loan history in Kenya dates back to 1952 when the British colonial government would award loans under the then Higher Education Loans Fund (HELFB) to Kenyans pursuing university education in universities outside East Africa and most specifically Britain, the USA, the former Union of Soviet Socialist Republics (USSR), India and South Africa. After independence in 1963, the government suspended the scheme and opted to directly meet the costs of higher education. However, in 1974 the number of students seeking university education had grown coupled with the dismal economic performance occasioned by the oil shocks of 1970s, it became increasingly difficult for the government to fully finance university education by provision of full scholarships and grants (<http://www.helb.co.ke>).

In response, the government introduced the University Students Loans Scheme in 1970-1974 financial year, which was managed by the Ministry of Education. Unfortunately, the government did not have articulated policies to guide recovery of mature loans from loanees (Otieno, 2010). In July 1995, the government through an act of parliament set up the Higher Education Loans Board (HELBB) with the mandate of not only administering the Student Loans Scheme, but also recovering all outstanding loans disbursed by HELFB since 1952 with a goal of establishing a revolving Fund from which funds can be drawn to loan needy Kenyan students pursuing higher education (<http://www.helb.co.ke>).

Without recovered loans, HELB would not be in a position to support the number of students it currently supports. When it was step up, the board inherited a large portfolio of unpaid debts, with rate of recovery being as low at only 3.3% (Otieno, 2010). The board has however made great strides in loan recoveries, by achieving 57% recovery rate in 2012 financial year (HELB, 2012) compared to 18% recovery rate between 2000 – 2001, (Otieno, 2010) . The Board's lending capacity has also increased to current standing at Kenya shillings 4.5 billion for the financial year 2012/2013 with private sponsored students also benefiting (HELB, 2012)

Despite this success the Board is continually facing a number of challenges in its endeavor to increase loan recoveries from past beneficiaries due to high unemployment rates, and increased demand from rising student population and swelling costs of education (HELB, 2012). Default by previous beneficiaries of the loans scheme leads to redundancy of the established revolving fund, thus affecting the running of the scheme and access to university education by qualified Kenyans, who cannot afford to meet the ever increasing cost of education (Kipkech, 2011). Identifying the major cause of loan default rates in student loan schemes and developing pro-active interventions might be the key to easing the burden of HELB towards realizing its core mandate of providing financing in the form of affordable loans and scholarships to millions of Kenyan students.

1.2 Statement of the problem

The effects associated with loan defaults are far reaching. For instance, the government incurs losses from the funds it provides to HELB, while the institutions incur indirect loss of tuition due to drop out. However, the real loss is felt by the future financially needy students who have qualified to join higher education (Thobe, 1997). According to Hillman (2014) financing a college education on credit is not necessarily perceived as a public policy problem, rather as an opportunity to increase educational opportunities for millions of students. However, when a huge number of borrowers cannot repay their education debts, serious questions on the efficacy disbursement and recovery of student's loan in any society arises. The revolving nature of HELB fund has failed to meet all the yearly demand of the new applicants due to increased default by previous beneficiaries. Student loan default might be associated with numerous systematic patterns, which perhaps if pinpointed may aid design public policy interventions, which might reduce the odds of defaulting.

At the time of this survey there was no published research on factors influencing student loan default using the national dataset from HELB. With this context in mind, there is dire need of intensive research to examine the factors associated with student loan default in Kenya.

1.4 Objectives

1.4.1 Main Objectives

To examine factors associated with defaulting on HELB loans in Kenya.

1.4.2 Specific Objectives

1. To develop a model to classify future HELB loan applicants as either defaulters or non-defaulters.
2. To determine the accuracy ratio (hit ratio) of the predicted model.

1.5 Justification

Fisher's linear discriminant analysis (FLDA) does not make assumptions of normality for the predictors unlike linear discriminant analysis which assumes that the predictors comes from a normally distributed population. Given that our predictor variables have both continuous and categorical variables, FLDA will provide unbiased results by taking into account any non-normality effects from the data. The predicted FLDA model will identify future students most likely to default based on their personal characteristics. This will enable HELB to come up with pro-active interventions that will reduce the current high default rates.

CHAPTER TWO: LITERATURE REVIEW

2.0 Introduction

This chapter reviews the literature into the factors which are believed to contribute to default of higher educational loans. The review is organized in four subsections: student demographic, socioeconomic factors, academic experience, and post college experience.

2.1 Student Demographics

2.1.1 Age

There is mixed evidence of the influence of age on student loan default. While a number of studies associate older students with default risk, some have found the opposite. Herr & Burt (2005) and Woo (2000) found age to be positively associated with default i.e. as age increases so does the probability of defaulting. On the other hand Hillman (2014), established a non-significant relationship between age and default, while Steiner and Teszler (2003) found this pattern only among the students older than 34 years. Herr and Burt (2005) pointed that older students are likely to have greater, say family obligations that may hinder loan servicing.

2.1.2 Gender

The relationship between gender and loan repayment is not very clearly outlined in the literature. A number of previous studies have established that gender influences loan repayments.(Woo, 2002; Steiner & Teszler, 2003; Herr & Burt, 2005 & Hillman, 2014) found that men are more likely to default than women. A study by Choy & Li (2006) suggests that women take longer to repay loans. However, others studies have failed to find any significant relationship between gender and default (Harrast, 2004).

2.1.3 Marital status

Family structure can affect a numbers of ways the likelihood of defaulting on loans. Being single, divorced or widowed was found to increase the probability of defaulting by more than 7 percent (Volkwein & Szelest, 1995) and up to approximately 40% (Volkwein et al., 1998). In addition,

being married lowers dramatically the default rate of some groups of students (Volkwein et al., 1998; Ismail & Serguieva, 2009).

2.2 Socio-economic status

2.2.1 Number of dependents

Studies have found that students who have children or other dependents (brothers and sisters) are expected to have more financial obligations compared to those who do not have dependents, hence resulting to greater chances of defaulting (Woo, 2002 & Hillman, 2014). In addition, more children require the sharing of limited financial resources, thereby decreasing the ability to repay loans (Herr & Burt, 2005).

2.2.2 Parent income

As it may be expected, borrowers with high family earnings after they leave college are less likely to default than those with low earnings (Herr & Burt, 2005 & Steiner & Tesler, 2005). Similarly, according to Hillman (2014), the odds of defaulting steadily decline as family income levels rises. Families with higher incomes are able to provide a financial safety net unavailable to students from lower income families. This safety net also helps student to meet their loan obligations through fluctuations in personal income. In general the higher the family income the lower the likelihood the student will default (Woo, 2002).

2.3 Academic Experience

2.3.1 Major

Researchers have established that academic majors are linked to the probability of defaulting. Given that some majors appear to be more robust to the job market and may require students to accumulate less debt (Harrast, 2004), there is a probability that students choosing some majors are less likelihood to default. A college major in a scientific, engineering, or agricultural discipline lowers the default probability by over 4 percent among the borrowers (Volkwein & Szelest, 1995). Studying special education, sociology, art history, or risk management and insurance was associated with high level of debt relative to other fields (Harrast, 2004). Flint (1997) argues that

the greater the mismatch between a student's undergraduate major and his or her current employment, the higher default risk.

2.3.2 Loan debt

Although the opposite would be taken to make more sense, borrowers with high indebtedness are less likely to default than borrowers with low indebtedness (Woo, 2002). Other scholars have argued that the more a student borrows the greater the chance of default (Choy & Li, 2006). Coupled with this mixed results Hillman (2014) found out that the mixed results from Choy and Li (2006) and Woo (2000) were due to a nonlinear relationship between debt and default. Students who drop out of college before graduating are less likely to accumulate debt, while those who graduate are likely to accumulate more debt due to their longer enrollment period (Hillman, 2014).

2.4 Post Collegiate Experience

2.4.1 Employment

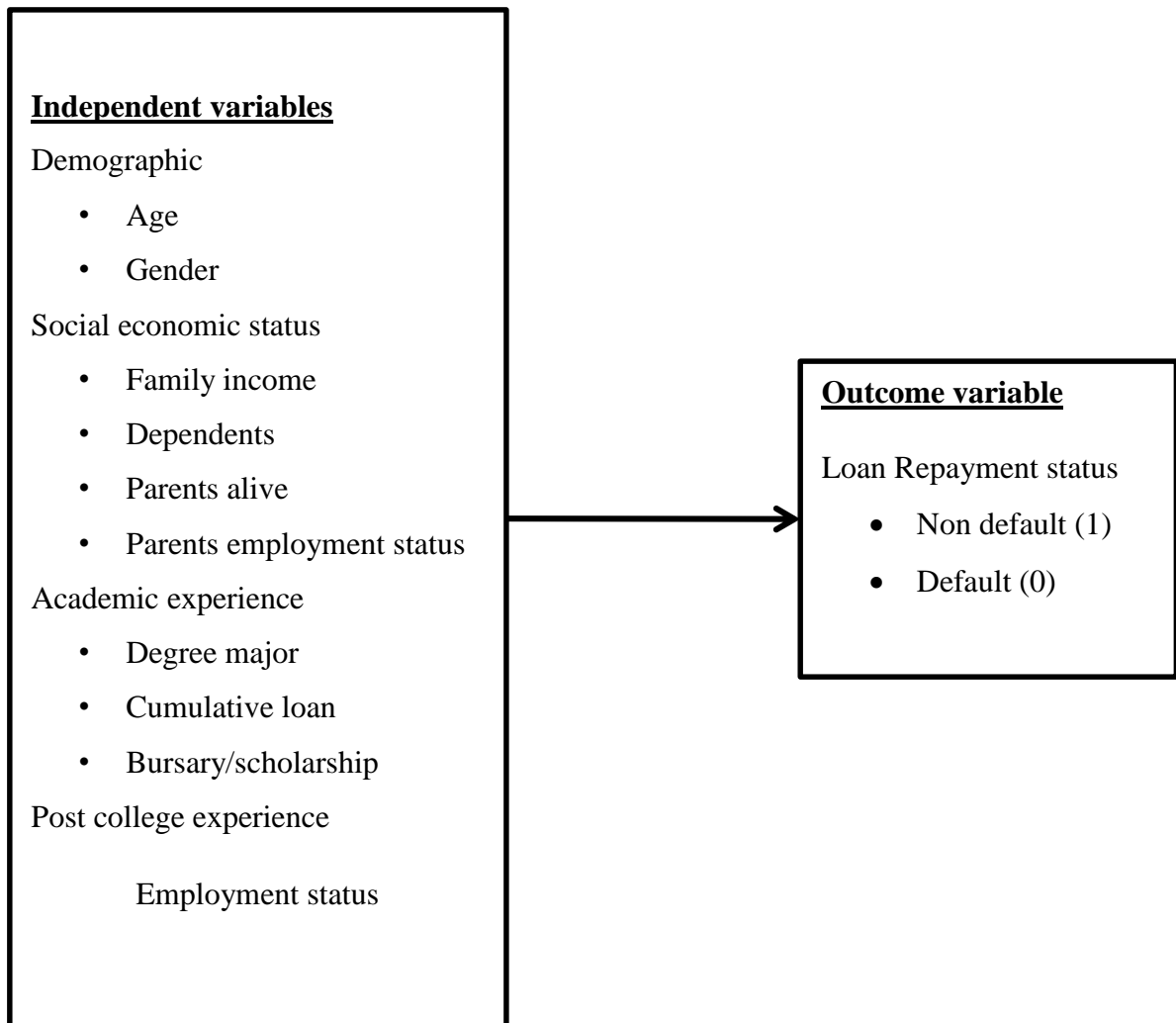
If students cannot get employment upon graduation or lose their job at some point during repayment, then they may have higher probability of defaulting. Borrowers who experienced unemployment showed an 83 percent increase in their probability of default over their original probability (Woo, 2002). This finding has been consistent in other default studies since job loss results in fewer financial muscles with which to clear the student's loan debts (Monteverde, 2000). One most recent study found that borrowers who are unemployed have nearly two times greater odds of defaulting than those who are employed (Hillman, 2014).

2.5 Conceptual frame work

In an effort to examine the relative effects of individual characteristics on loan repayment, a conceptual framework (Figure 1) was developed to guide the variable development and analysis. This model draws heavily upon the literature on multilevel analysis of student loan default (Hillman, 2014). A conceptual frame work indicates the effect of the independent variables on the outcome (Kombo & Donald, 2006).The conceptual framework incorporates four perspectives of student characteristics (demographic, social status, academic experience, and post college experience). From the research literature, it's evident that they contribute to student's loan

repayment outcomes. The conceptual frame work is as shown in figure 1 below. This study was geared towards pointing out how the student characteristics can be used to discriminate/classify between defaulters and non-defaulters of student loans.

Figure 1: Theoretical frame work of factors associated with student loan default



CHAPTER THREE: METHODOLOGY

3.0 Introduction

The goal of this study was to identify the major factors which best explain what causes student HELB loans defaults. Choosing the best analytical technique to implement ensures robust estimates for prediction purposes. However, the task to choose the best analytical technique is paramount, because there is need to apply an approach that can predict an outcome and at the same time classify new cases to their respective groups. Logistic regression and discriminant analyses are two multivariate statistical methods which are majorly used for the evaluation of the association between various predictors and categorical outcomes (Antonogeorgos et al., 2010). While both are suitable for the development of linear classification or prediction models, the choice between the two is heavily dependent on the model assumptions. Linear discriminant analysis (LDA) assumes that the sample comes from a normally distributed population whereas logistic regression is called a distribution free test. On the contrary, Fishers linear discriminant analysis (FLDA) which is an earlier form of linear discriminant analysis does not assume normal population (Johnson & Wichern, 2007). Despite these differences in assumptions, Pohar et al., (2006) and Antonogeorgos et al., (2009) argues that, the differences between these two methods become negligible if the sample size is large enough, say 50 observation or more. In addition, a study by Cleary & Angel (1984) proved that the results of OLS, discriminant analysis, and logistic regression are often similar, while Antonogeorgos (2010) had the same conclusion of convergence of results from logistic and discriminant analysis. It's with the above knowledge from literature review that this survey settles on the use of Fisher's linear discriminant analysis which according to Ramayal et al., (2006 & 2010) is a parametric test that is more powerful than the non-parametric test-logistic regression.

3.1 Sources of data

This study uses a nationally representative student loan survey data, sourced from Kenya Higher Education loan board (HELB) database. The study sample consists of Kenyan students both in private and public universities who benefited from HELB loans between the year 2009 and 2013. A student only qualifies for HELB loan if he or she supplied complete and accurate information on the loan application form. After graduating, the borrowers are given a grace period of one financial year, after which they are supposed to start servicing their loans by filling in a repayment form which contains recovery details. The background information from application data base was merged with recovery data using the national identification number as the unique identifier for the two databases. The merged data formed the analysis basis for this study.

For validation purposes, the sample data set was further split into two portions. This was achieved by splitting the sample into analysis subsample, for developing the discriminant function and a holdout subsample for validating the predictive accuracy of the model (Ramayah et al., 2006). According to previous studies, there is no hard and fast rule that has been established for splitting the sample data, and some researchers prefer a 60 – 40 or 75 – 25 split between the analysis and the holdout groups, depending on the overall sample size (Hair et al., 2010). In this study we split the sample by computing a variable (RAND) using the function below:

$$\text{RAND} = \text{UNIFORM}(1) < 0.65$$

The value 0.65 represents 65% analysis subsample while the remainder (35%) represents the holdout sample with each subsample being proportional to the respective original group sample.

3.2 Variables Description

The number of predictor variables groups can be two or more, but they must be mutually exclusive and exhaustive, distinct and unique on the set of outcome variables chosen (Hair et al., 2010). To achieve the objective of the study in predicting student loan default, the status of the student loan repayment which is a binary outcome set to “1” for those in repayment and “0” for defaulters was chosen as the outcome variable. The predictor variables were namely;-background characteristics (age, gender), social economic status (family income, dependents, marital status, parent employment status, parents alive), academic experience (degree major, loan awarded, bursary award), and post college experience (employment status).

Given that the majority of our predictor variables are categorical, it was necessary to come up with a way to incorporate them in our model as continuous variables. Hair et al., (2010) argues that categorical variable can be represented as a dummy variable and included in the analyses requiring only continuous variables. This approach was adopted in this study where all categorical variables were converted into dummy variables. In addition, Uddin (2013) suggests that the predictor variables in both discriminant analysis and regression can either be continuous or categorical.

3.4 Selection of predictor variables

The linear combination of predictors which best explains loan default was selected using stepwise method approach. The stepwise approach involves entering the predictors into the discriminant function on the basis of their discriminating power (Hair et al., 2010). The stepwise procedure begins by considering the best discriminating variable. The initial variable is then paired with each of the other independent variables one at a time, and the variable that is best able to improve the discriminating power of the function in combination with the first variable is included in the function. Essentially, either all predictors will have been included in the function or the excluded variables will have insignificant classification power.

3.5 Analytical Technique

Fisher's linear discriminant analysis is a technique for classifying and grouping observations into different groups based on a set of random predictors (Hezlin, 2009). This method involves deriving a linear combination of two or more random predictors that will discriminate best between given groups. The terms Fisher's linear discriminant and linear discriminant analysis are often used interchangeably; Fisher's (1936) describes a method which does not make some of the assumptions of linear discriminant analysis, such as normally distributed classes (Wichern & Johnson, 2007). Hence, as stated earlier this study adopts Fisher's linear discriminant analysis.

3.5.1 Discriminant function

The linear combination for Fisher's linear discriminant analysis, also known as the discriminant function, is developed as in (1).

$$Z_{jk} = \alpha + W_1 X_{1K} + W_2 X_{2K} + \dots + W_n X_{nK} \quad (1)$$

Where,

Z_{jk} = discriminant Z score of discriminant function j for object k

α = intercept

W_i = discriminant weight for independent variable i

X_{iK} = independent variable i for object k

The discriminant score Z_{jk} is the summation of the values obtained by multiplying each predictor variable by its discriminant weight. The number of discriminant functions to be estimated from a discriminant analysis is less than or equal to the categories in the outcome variable minus one or the number of predictor variables whichever is less (Hair et al., 2010).

3.5.2 Fisher's linear discriminant functions

In this study, Fisher's linear discriminant function is used to develop the predictive model based on the discriminant function. Fishers linear discriminant analysis is a classification method originally developed with an idea to transform the multivariate observations x 's to univariate observation y such that y 's derived from population one and two are separated as much as possible (Wishern & Johnson, 2007).

Suppose a fixed linear combination of the x 's takes the values; $y_{11}, y_{12}, y_{13} \dots y_{1n_1}$ and $y_{21}, y_{22}, y_{23} \dots y_{2n_2}$ for the observation from the population one and two respectively. The separation of the two set of univariate y 's is assessed in terms of the difference between \bar{y}_1 & \bar{y}_2 , expressed in standard deviation units. This separation is defined as below;

$$\text{Separation} = \frac{|\bar{y}_1 - \bar{y}_2|}{S_y^2}$$

$$\text{Where } S_y^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

The goal is to choose the linear combination of the x 's to achieve maximum separation of the sample means \bar{y}_1 & \bar{y}_2 .

$$\hat{Y} = W_0 + \underline{W}^T \underline{X} \quad (2)$$

Given a linear model (2) we are supposed to find a vector \underline{W}^T , so that when we project data along it, the two sample means are separated as far as possible from each other while at the same time variance remaining as close as possible. Working with two univariate set of data, we describe their means and variances as follows;

$$E[\hat{Y} / x \in c_i] = W_i + W^T \mu_i$$

$$\text{Var}[\hat{Y} / x \in c_i] = W^T \mu_i W$$

We find W that maximizes the ratio (3) below;

$$J(W) = \frac{(\mu_0^T W - \mu_1^T W)^2}{W^T S_0 W + W^T S_1 W} \quad (3)$$

$$= \frac{[(\mu_0^T - \mu_1^T)W]^2}{W^T (S_0 + S_1)W} \quad (4)$$

Let $M = \mu_0^T - \mu_1^T$ and $S_p = S_0 + S_1$, after replacement (4) simplifies to;

$$J(W) = \frac{[M^T W]^2}{W^T S_p W} \quad (5)$$

We define S_p in form of two matrices;

$$S_p = R^T R \quad \text{Where, } R \text{ is the square root of } S_p$$

$$J(W) = \frac{[M^T W]^2}{W^T R^T R W} \quad (6)$$

We project W through R and create a vector V , such that

$$V = R W \quad \text{hence, } W = R^{-1} V \quad (7)$$

Replacing (7) into (6), we get;

$$J(W) = \left([(R^{-1})^T M]^T \frac{V}{|V|} \right)^2 \quad (8)$$

To maximize (9), we find a vector V that will project (9) by ensuring that the two vectors (i.e. $[(R^{-1})^T M]^T$ and V) are in the same direction.

$$J(W) = \left(a [(R^{-1})^T M]^T \frac{V}{|V|} \right)^2 \quad (9)$$

The vector V that maximizes (9), where a is a constant is given as follows;

$$V = a [(R^{-1})^T M]^T \quad (10)$$

After replacing the values of M and W in (10) we get;

$$\begin{aligned} V &= a (R^{-1})^T (\mu_0 - \mu_1) \\ W &= R^{-1}V = a R^{-1}(R^{-1})^T (\mu_0 - \mu_1) \\ &= a (R^T R)^{-1} (\mu_0 - \mu_1) \end{aligned}$$

Therefore;

$$W = S^{-1}(\mu_0 - \mu_1) \quad (11)$$

Hence, (11) gives vector W , which is the linear coefficients of the Fisher's discriminant function that maximizes ratio (3) by projecting the means of the two groups as far as possible from each other, and the variances as close as possible.

5.4 Allocation rule

Finally, a new case is classified by projecting it onto the maximally separating direction and classifying it using an allocation rule based on Fisher's discriminant function defined below;

We allocate X_0 to group one if

$$\hat{y}_0 \geq \hat{m} \text{ or } \hat{y}_0 - \hat{m} \geq 0$$

Otherwise, we allocate X_0 to group two if

$$\hat{y}_0 < \hat{m} \text{ or } \hat{y}_0 - \hat{m} < 0$$

Where,

$$\hat{y}_0 = (\bar{\mu}_1 - \bar{\mu}_2)^T S_{\text{pooled}}^{-1} X_0$$

$$\hat{m} = 1/2 \{ (\bar{\mu}_1 - \bar{\mu}_2)^T S_{\text{pooled}}^{-1} (\bar{\mu}_1 + \bar{\mu}_2) \}$$

3.6 Assumptions

3.6.1 Equal variance –Covariance matrix

The main assumption in conducting the discriminant analysis is that the groups have equal variance-covariance matrices despite their means being considerably different. This assumption is tested by using a transformed value of Box's M , which assesses the significance differences in the matrices between groups (Hair et al., 2010). In the test, the null hypothesis (H_0) is that the variance-covariance matrices of the groups are the same in the population. The aim is to have non-significant probability level to indicate that there is no difference between the group covariance matrices. Lanchernbruch (1975) argues that discriminant analysis is a robust technique which can sustain some deviation from this assumption of equal variance-covariance matrix. The violations of the equal group covariance matrices can be tolerated (Thobe, 1997). In addition, according to Tabachnick and Fidell (1996), violation of this test is not a problem if it's due to skewness compared to outliers. Further, if the sample size is large, then the violation cannot be a big problem and the validity of estimating the discriminant function can be checked by hit ratio of the holdout sample (Uddin, 2013)

3.6.2 Multicollinearity

Multicollinearity, measured by tolerance ($1 - R^2$), denotes that two or more predictor variables are highly correlated, which means that variables with high correlations can be explained by other variable(s) and thus it adds little or no explanatory power in the model (Hair et al., 2010). The correlation matrix can be used to check multicollinearity of the variables or can also be solved by using stepwise discriminant analysis.

3.6.3 Linearity

In discriminant analysis the outcome variable is categorical and therefore there is no linear relationship between the outcome and predictor variables. The assumption is checked by the degree of relationship between one predictor variable and another predictor variable, and the degree of relationship between one predictor variable with the rest of the predictor variables. If

one variable is consistently found to be nonlinear with the other variables then only that variable should be considered too. Linearity can be assessed using Pearson's correlation test (Filed, 2010; Hair et al., 2010)

3.6.4 Absence of outliers

Discriminant analysis is very sensitive to outliers. Outliers refer to observations of characteristics identifiable as distinctly different or distinct from other observations (Hair et al., 2010). Multivariate outliers can be identified by the value of the standardized score for each continuous observation and if the standardized score is more than 2.5, then there is presence of outliers.

3.6.5 Sample size

Discriminant analysis is quite sensitive to the ratio between sample size and number of predictors. Some researchers suggest 20 sample sizes per predictor. Burns and Burns (2008) argue that this ratio should be at least 5 times higher than the number of predictor variables. It should be noted that the results become unstable as the sample size decreases comparative to the number of predictors.

3.7 Procedures of running discriminant analysis

3.7.1 Evaluating the group centroids difference

A measure of success of the discriminant analysis is its ability to define the discriminant functions that results in significantly different group centroids. The transformed chi-square statistic $\{\chi^2 = -[(n - 1) - 0.5(m + p + 1)] \ln \Lambda, df. = k - 1, m = \text{number of discriminant function, \& } p = \text{number of predictors}\}$ tests the hypothesis that the means (centroids) of the functions listed are equal across groups as well as to the overall centroid (Uddin et al., 2013)

3.7.2 Box's test of equality of covariance matrices

In order to estimate a valid discriminant function, an important assumption is that the variance-covariance matrices of the groups should be the same. The covariance matrices are the same if the log determinants for the two groups are the same. Box's M tests the null hypothesis that the covariance matrices do not differ between groups formed by the dependent variable. By use of this test, we are looking for a non-significant M to show similarity and lack of significance difference (Wishern & Johnson, 2007).

3.7.3 Eigen value

The Eigen values assess relative importance by showing the percentage of variance explained by the predictor variable. Theoretically, Eigen value (λ) is a ratio of between group sum of squares to within group sum of squares and ranges between 0 and 1. It is calculated as shown below.

$$\lambda = \frac{BSS}{WSS} = \frac{\sum (\bar{z}_j - \bar{z})^2}{\sum (z_{ij} - \bar{z}_j)^2}$$

Zero λ is equivalent to zero discriminatory power. As the value of λ increases so does the discriminatory power of discriminant function. To determine if the Eigen value is significant, two statistical indicators are derived from the Eigen value which are: canonical correlation eta (η) and Wilks' lambda (Λ) for the model.

3.7.4 Canonical correlation

The canonical correlation (η) is the measure of association between the discriminant function and the predictor variables. The square of canonical correlation coefficient (η^2) gives the percentage of variance explained in the dependent variable. Hence η is defined as;

$$\eta = \sqrt{BSS/TSS} = \sqrt{\frac{\sum (\bar{z}_j - \bar{z})^2}{\sum (z_i - \bar{z})^2}}$$

3.7.5 Wilk's lambda

The Wilk's lambda (Λ) criterion was used for choosing variables for analysis. The smaller the lambda for an independent variable, the more that variable contributes to the discriminant function. Λ is a chi-square distributed with degrees of freedom = $(k - 1)$, where k equal to the number of parameters estimated. Wilks' lambda (Λ) is defined as;

$$\Lambda = \frac{WSS}{TSS} = \frac{\sum (z_{ij} - \bar{z}_j)^2}{\sum (z_i - \bar{z})^2} \approx \chi^2(k - 1)$$

3.8 Classification

Finally, the statistical tests for assessing the significant of the discriminant function only assess the degree of difference between the groups based on the value of the discriminant function (Z scores), but do not indicate how well the functions predicts (Hair et al., 2010). To determine the predictive ability of a discriminant function, a classification matrix should be constructed. When prediction is perfect, all cases will lie on the main matrix diagonal giving the percentage of correct classification know as hit ratio. A cutting score representing the dividing point used to classify the observations into groups is calculated based on the two group centroids and relative size of the two groups as follows;

Cutting score for an unequal group sizes (unequal prior probabilities)

$$Z_{ij} = \frac{N_i \bar{Z}_j + N_j \bar{Z}_i}{N_i + N_j}$$

Cutting score for equal group sizes

$$Z_{ij} = \frac{\bar{Z}_j + \bar{Z}_i}{2}$$

Where Z_{ij} = cutting score between groups i and j , \bar{Z}_i & \bar{Z}_j are centroids of groups i & j while N_i & N_j are number of observations.

3.8.1 T-test for equal size groups in dependent variable

If both groups have equal observations, a t-test can be conducted to test if the hit ratio is higher than chance (Hair et al., 2010), where the null hypothesis is that the model's hit ratio is not higher than the chance ratio. P is the proportion correctly classified observations and the degree of freedom (df) is total sample size (N) minus two.

T-test is given below;

$$t = \frac{p - 0.5}{\sqrt{\frac{0.5(1.0 - 0.5)}{N}}} \approx (N - 2)$$

To establish the standard of comparison for the Hit ratio we determine the percentage that could be classified correctly by chance as follows:

$$C_{\text{equal groups}} = 1 \div \text{Numbers of groups}$$

3.8.2 Press's Q for Uneven sample groups

A statistical test for the discriminatory power of the classification matrix when compared with a chance model is Press's Q statistic, which compares the number of correct classification with the total sample size and the number of groups. The calculated press's Q value is compared to a chi-square value of 1 degree of freedom. If Q statistic value exceeds the critical value, then the classification matrix will be deemed statistically better than chance. The press Q value is calculated as below;

$$\text{Press's Q} = \frac{[N - (nK)]^2}{N(K - 1)} \approx \chi^2_{(1)}$$

Where, N = total sample size, n = number of observations correctly classified and K = number of groups. The percentage that could be classified correctly by chance is as follows;

$$C_{\text{unequal groups}} = P^2 + (1 - P)^2$$

Where, p = proportion of observations in group 1 and 1 - p = proportion of observations in group two.

CHAPTER FOUR: SURVEY FINDINGS

4.1 Descriptive Statistics

Table 1, shows the frequency for each binary variable available in the study sample data. Out of the overall sample, 51% of borrowers had defaulted on their student loans. The largest proportion (87%) of the employed borrowers were servicing their loans while 70% defaulters were not employed. In terms of gender, more males were defaulting compared to females at 56% and 44% respectively. A sizeable portion of the defaulters (48%) had their parent employed while 51% of those in active payment had non-working parents. More of the students who had applied for bursary had defaulted compared to those who had no bursary application. A considerable number of students who had both parents alive in default and non-default subsample were actively servicing their loans, 96% and 95% respectively. In terms of degree major in education 48% were non-defaulters compared 33% defaulters. Those majoring in business had equal representation (22%) as default and non-default. Finally, 27% of students majoring in STEM had defaulted as compared to 15% who were non-default.

Table 1: Percentage description of categorical variables

Variable	Overall (n=7354)		Defaulter (n=3733)		Non-Defaulter (n = 3621)	
	0	1	0	1	0	1
Whether or not default occurred (0 = default 1 = non-default)	51	49				
Employment (0 = No 1 = Yes)	42	58	70	30	13	87
Gender (0 = male 1 = female)	56	44	58	42	53	47
Marital status (0 = married 1 = Single)	2	98	1	99	2	98
Father education (0 = non-tertiary 1 = Tertiary)	55	45	54	46	55	45
Parent employed (0 = No 1 = Yes)	52	48	52	48	51	49
Orphaned (0 = No 1 = Yes)	90	10	90	10	90	10
Bursary application (0 = No 1 = Yes)	4	96	4	96	5	95
Dependents (0 = No 1 = Yes)	20	80	21	79	19	81
Arts major (0 = Others 1 = Arts)	94	6	93	7	95	5
Business major (0 = Others 1 = business)	78	22	78	22	78	22
Education major (0 = Others 1 = education)	60	40	67	33	52	48
Health major (0 = Others 1 = health)	95	5	95	5	96	4
Humanities major (0 = Others 1 = humanities)	95	5	95	5	95	5
STEM major (0 = Others 1 = STEM)	78	22	73	27	85	15

The figures in the table represent the percentage of the sample in each respective category

Table 2, depicts the group means and standard deviations for each continuous variable in the overall sample. The means give an overview of whether the means of the two groups are different. Both defaulters and non-defaulters averaged almost the same age at 22 years. Average student's loan awarded was higher for the non-defaulters compared to defaulters. On the other hand, the average family income was much higher for defaulters. In overall, there are notable differences for the age, family income, and amount of HELB loan awarded.

Table 2: Means and standard deviations of the continuous variables

	Overall (n=7354) mean (s.d.)	non-default (n=3733) mean (s.d.)	default (n=3621) mean (s.d.)	Range	
				Min	Max
Age of the student	22.47 (1.09)	22.67 (1.27)	22.26 (.85)	20	26
Family income in Ksh.	469359.55 (373392.44)	467164.45 (374601.297)	471468.12 (372291.48)	0	4081176
Total loan awarded	41537 (5305.54)	41372.67 (5325.65)	41304.35 (5287.01)	35000	60000

s.d = standard deviation

4.2 Discriminant analysis findings

4.2.1 Test of equality of Covariance matrices by using Box's M

In discriminant analysis the basic assumption is that the covariance matrices are equivalent. Box's M tests the null hypothesis that covariances matrices do not differ between the groups formed by the outcome variable. The covariance matrices are equal if the log determinants for the default and non-default covariance matrices are almost equivalent plus the larger the log determinants the

more the group's covariance matrix differs. When tested using Box's M, we are looking for a non-significant M to show similarity and non-significance difference. From table 3, the log determinants appear to be equal, Box's M is 822.65 with F = 54.78 which is significant at $p < 0.001$. This means that the null hypothesis of equal variance covariance is rejected; consequently violating the assumption. Despite the violation of the assumption, quite often, the discriminant analysis can pass the reliable test during the time of validity check (Uddin, 2013). It's also important to note that, with a large sample size, like in this survey, small deviations from homogeneity will be significant; hence, Box's M should be interpreted alongside the log determinants (Agresti, 1996)

Table 3: Box's M test of equality of covariance of matrices

Account status	Rank	Log		Approx.F	Sig.
		Determinant	Box's M		
Default	5	8.17	822.65		
Non-default	5	7.54		54.78	0.001
Pooled within-groups	5	7.69			

Tests the null hypothesis of equal population covariance

4.2.2 Significance of the discriminant function

From table 4, function 1 means the estimation of one function from a two-group discriminant analysis. The larger the Eigen value, the more the variance in the outcome variable is explained by the function. The canonical correlation (η) for the estimated function is 0.602 and it measures the association between the discriminant function and the dependent variable. This means that 36% of variation in the discriminant function is explained by the two groups (defaulters and non-defaulters). The wilk's lambda value is 0.634, measuring how well the estimated function separates cases into different groups. The Wilks' lambda ranges from zero to one; the closer the value is to zero the greater its discriminatory ability (Thobe, 1997). The associated chi-square statistic is 2170.10 with 5 degrees of freedom and significance at the 0.001 level. This χ^2 tests the hypothesis that the means (centroids) of the two functions are equal to the overall mean. The small significance χ^2 value indicates that the discriminant function does better than chance at separating the groups.

Table 4: Significance of the discriminant function

Function	Eigenvalue	Canonical				
		Correlation	Wilk's λ	χ^2	df	Sig.
1	.569 ^a	.602	.638	2170.10	5	0.001

First I canonical discriminant was used in the analysis

After applying stepwise discriminant analysis to predict loan default, five of the fifteen independent variables were found to significantly discriminate defaulters from non-defaulters. These five were; employment status, age, gender, bursary application and degree major (education).

Table 5, depicts the summary of the predicted discriminant function. The discriminant coefficients are available in unstandardized and standardized forms. The unstandardized form (plus constant) is used to calculate the discriminant score, while the standardized are more appropriate for interpretation purposes, since they are not affected by the scale of the predictors (Hair et al., 2010). The standardized coefficients provide an index of the importance of each covariate to the function by looking at the magnitude of these coefficients while ignoring the sign (Thobe, 1997). The larger the coefficient, the greater that predictor's contribution to the discriminant function. The coefficients' sign only shows whether the variable is making a positive or negative input. In this study the negative coefficients is associated with default while the positive coefficients are associated with non-default.

As shown in table 5, four predictors displayed positive relationships with loan repayment. Borrower's employment status provides the greatest contribution in determining loan repayment status i.e. being employed increases the likelihood of loan repayment. Gender (female=1), age, and education major also predicted the probability of loan repayment. On the other hand, bursary application had a negative coefficient hence, best predicts loan default.

4.2.3 Fisher's discriminant function

One of the main objectives of this survey was to estimate a discriminant model that can be used by HELB to classify new loan applicants as either defaulters or non-defaulters. Fisher's linear discriminant function was developed using the unstandardized coefficients which are the multiplier of the independent variables in their original units of measurements. Based on the summary of the discriminant function presented in table 5, five variables were significant for discriminating between loan defaulters and non-defaulters. By using these five variables and their coefficients, the required Fisher's linear discriminant equation, also known as a discriminator was developed as below;

$$Z = - 4.629 + 2.386 \text{ Employment} + 0.289 \text{ education} - 0.262 \text{ bursary} + 0.167 \text{ age} + 0.098 \text{ gender} \quad (12)$$

The discriminator coefficient in (12) indicates the partial contribution of each variable to the discriminant function by holding other variables constant.

Table 5: Fisher's discriminant function summary output

Independent variables	Standardize	Unstandardized	Wilks'	
			Lambda	Sig.
Student Employment	0.951	2.386	0.651	0.000
Student age	0.180	0.167	0.968	0.000
Degree major(education)	0.140	0.289	0.978	0.000
Student gender	0.049	0.098	0.998	0.002
Bursary application	-0.055	-0.262	0.999	0.012
Constant		-4.629		
Centroid (Default)			-0.739	
Centroid (Non-default)			0.769	
Wilk's Labda			0.638	
Canonical correlation			0.602	

By use of the model (12), HELB officials can be in a position to predict the probability of a new loan applicant being a defaulter or a non-defaulter. A Z score for each applicant is calculated by substituting the values of the predictors in (12). The calculated Z score is there after compared to the groups' centroids shown in table 5, to determine which group the applicant is likely to belong to.

4.2.4 Group Centroids

The centroid represents the average value of the discriminator scores for each specific group. As depicted in the fourth column of table 5, the centroid for non-defaulters is 0.769 while that of defaulters is -0.739 . The two group centroids are used to classify a new applicant of student loan by comparing his or her discriminant Z score to the centroids. If the applicant Z score is negative then he or she is likely to be classified as a defaulter, else as a non-defaulter.

4.2.5 Assessing the internal validity of the model

The validity of the discriminant function is determined using a classification matrices developed using the holdout subsample. The best accuracy rate is given by the holdout sample since it's not utilized to drive the discriminant function. The model is valid if the classification accuracy (hit ratio) is at least one-fourth greater than what is achieved by chance (Ramayah et al., 2010).

The classification results from tables 6 show a hit ratio of 77.3%, which shows the percentage of students correctly classified as either default or non-default groups. This classification accuracy (hit ratio) shows that non-defaulters were classified with better accuracy (86%) compared to defaulters (69%). As illustrated by table 7, the hit ratio (77.3%) is substantially higher than both the proportional chance and the maximum criterions of 50% respectively. In addition, the hit ratio (77.3%) exceeds the threshold (at least one-fourth greater than what is achieved by chance) which in our case is $1.25 * 50 = 62.5\%$.

Table 6: Classification matrix

		Predicted Group Membership			
		Default	no-default	Total	
Estimation Sample	Count	Default	1745	716	2461
		no-default	296	2068	2364
		Total	2041	2784	4825
	%	Default	70.9	29.1	100.0
		no-default	12.5	87.5	100.0
	Holdout Sample	Count	Default	873	399
no-default			175	1082	1257
Total			1048	1481	2529
%		Default	68.6	31.4	100.0
		no-default	13.9	86.1	100.0

77.3% of unselected original grouped cases (holdout sample) correctly classified

4.2.6 Press's Q test for unequal sample groups

The final validation measure is the Press's Q test, which is a statistical measure for comparing the classification accuracy to a random process (Uddin, 2013 and Ramayah, 2010). From table 7, the calculated press Q statistics is 754.12 which exceed the critical value of 6.63. Hence, the hit ratio for the holdout sample exceeds at a statistically significant level, the hit ratio expected by chance. Thus the estimation and use of Fisher's linear discriminant model in classification of borrowers of student loans is justified using Hit ratio, chance criterion and Press's Q test.

Table 7: Classification accuracy as compare to chance

Measure	Value	Hit ratio (holdout sample)
Maximum chance	0.50	77.3
Proportional chance	0.50	77.3
1.25 times higher than proportional chance	62.5	
Press Q table value	6.635	
Press Q calculated value	754.12**	

** P < 0.05.

CHAPTER FIVE: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

The survey examined the role of personal characteristics in the student loan default. Many prior studies have shown that characteristics associated with individual students are strongly related with one's ability to repay the loan borrowed (Hillman, 2014; Steiner & Teszler, 2005; Herr & Burt, 2004; Podgursky, 2002). In this study student employment is one of the strongest predictors of loan repayment. This agrees with the findings from previous researchers who have found that uncertainty and unemployment affects loan repayment directly (Choy & Li, 2006; Lochner & Monge-Naranjo, 2004). Findings in this survey showed that 70% of the defaulters were unemployed while 83% of the non-defaulters were employed. These findings agree with Woo (2002), who found that borrowers who experience unemployment showed 83% increase in their probability of default. In a more recent study by Hillman (2014), borrowers who are unemployed have nearly two times greater odds of defaulting compared to the employed. In addition, according to Dynaski (1994), 83 percent of borrowers agreed that unemployment and lack of income were very or somewhat reasons for their having defaulted. Hence, if students cannot find gainful employment upon graduating or are unemployed at some point during loan servicing, then they are almost assured of entering into default. Among other reasons behind student default is: lack of proper equipping of students for gainful employment especially with for-profit institutions hence difficult in timely payment of large debt accumulated by attending these institutions, (Hillman, 2014).

Just like the findings from Harrast (2004) and Herr & Burt (2004), we find age to have a positive relationship with default; as age increases so does the likelihood of default. According to Woo (2000), older students are likely to default than younger students, perhaps due to a weakening of ties to parents and family who might assist the student while experiencing financial difficulties. Default by older students can also be explained by greater financial obligations that come with aging, for example, family support which may inhibit loan repayment (Herr & Burt, 2004). Additionally, older students may be more likely to default because they owe more than their younger counterparts and because they may have relatively less in available resources to repay the loans.

Results show that the student choice of a degree major has a positive relationship with loan repayment. Students majoring in education are less likely to default compared to others majors. The findings agree with the recovery data from HELB annual report (2012) which rates TSC as the leading organization in remitting the employees HELB deductions. Harrast (2004) argues that some majors tend to be more resilient to labor market conditions and may require students' accumulating less debt. According to Flint (1997), the greater the incompatibility between a student's undergraduate major and his or her employment, the higher the risk factor for default.

From our findings there is exist a positive relationship between gender (1 = female) and loan repayment. Female borrowers are slightly less likely to default compared to male borrowers. These findings agree with a numbers of recent studies that have shown men are more likely than women to default on student loans (Andruska, 2014; Steiner & Teszler, 2005; Woo, 2002; Podgursky et al., 2002).

The relationship between bursary application and default of students' loan has not been studied much as per the available literature. The results of this study showed that bursary applications have a negative relationship with loan repayment implying that bursary applications are at best suited to predict loan default. A previous study by Greene (1989) found that grants and scholarships reduced the probability of default.

Finally, the discriminant model correctly classified 77.3% of the loan applications (i.e. 1082 that repaid as agreed and 873 that defaulted out of 2529). Essentially, if this discriminant model was used to make the loan decisions, 1481 of the loans would have been granted and 399 (27%) would have defaulted.

In conclusion, the empirical evidence suggests that 36 percentage of the variation in the dependent function is accounted for by the difference between the two groups. On the other hand, the model (12) correctly classifies 77.3 per cent of the loan applications. Five factors emerged as the main reasons associated with student loan default, namely; employment status, age, gender, degree major, and bursary application. The economy performance clearly appears to play the greatest role in determining default by affecting the employment status of the students. Other aspects like schooling process and administration of the loan program also affects loan default, namely; choice of major and bursary application. This means that the institutions involved directly or indirectly in loan provision should work in harmony to ensure efficacy of the whole process of disbursement and recovery of the student loans. For effective use of the prediction model (12) in discriminating between default and non-defaulters, a whole lot of other factors beyond the borrower's characteristics should be considered e.g. institutional characteristics.

REFERENCE

- Agresti, A. (1996). *An introduction to categorical data analysis*. New Jersey: John Wiley and sons.
- Antonogeorgos, G., Panagiotakos, D. B., Priftis, K. N., & Tzonous, A. (2010). 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*, 1-9.
- Bruce, S. P., & Volkwein, F. J. (1995). Individual and campus characteristics associated with student loan default. *Research in Higher Education*, 36(1), 41-72.
- Choy, S. P., & Li, X. (2006). *Dealing with debt: 1992 - 93 bachelor's degree Recipients 10 years later*. US Department of Education, National Center for Education Statistics, Washington DC.
- Dynarski, M. (1994). Who defaults on student loans? Findings from the National postsecondary student aid study. *Economic of Education Review*, 13(1), 55 - 68.
- Flint, T. (1997). Predicting student loan defaults. *Journal of Higher Education*, 68(3), 322-354.
- Greene, L. L. (1989). An economic analysis of student loan default. *Educational and policy analysis*, 11(1), 61-68.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Prentice Hall.
- Harrast, S. A. (2004). Undergraduate borrowing: A study of debtor students and their ability to retire undergraduate loans. *Journal of student financial aid*, 34(2), 21 - 37.
- HELB. (2012). *2012 annual report of the higher education loan board*. Retrieved from Higher Education Loan Board: <http://www.helb.co.ke>.
- Herr, E., & Burt, L. (2005). Predicting student loan default for the University of Texas at Austin. *Journal of Student Financial Aid*, 35(2), 27 - 49.
- Hillman, N. W. (2014). A Multilevel analysis of student loan default. *The review of Higher Education*, 37(2), 169 - 195.
- Ionescu, S. A., Murgoci, C. S., & Gheorghe, C. M. (2009). Towards profitability on the financial markets; A discriminant analysis approach. *Transactions on Business and Economics*, 6(3), 99-111.
- Ismail, S. (2011). Students' attitude to educational loan repayments: A structural modelling approach. (*Doctoral dissertation*).

- Ismail, S., & Serguieva, A. (2009). Investigating possible benefits of student loan-backed securitization in the context of the Malaysian Higher Education. *Research in Higher Education Journal*, 5, 1-9.
- Kipkech, R. (2011). Determinants of students' loan default : a case study of higher education loans board.
- Kombo, D. K., & Tromp, D. (2006). *Proposal and thesis writing*. Paulines Publications africa.
- Lochner, L., & Monge, A. N. (2003). Education and default incentives with government student loan programs. *Presented at the Society of Labor Economist (SOLE) meeting*. San Antonio, Texas, April 30-May 1, 2004.
- Monteverde, K. (2000). Managing student loan default risk: Evidence from a privately guaranteed portfolio. *Research in Higher education*, 41(3), 331 - 351.
- Otieno, W. (2010). Student loans in Kenya: Past experiences, current hurdles, and opportunities for the future. *Journal of Higher Education in Africa*, 2(2), 75-99.
- Podgursky, M., Ehlert, M., Monroe, R., Watson, D., & Wittstruck, J. (2002). Student loan defaults and enrollment persistence. *Journal of Student Financial Aid*, 21(5), 497-512.
- Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: A simulation study. *Metodološki Zvezki*, 143-161.
- Rahman, H. A., Malek, H. A., & Fadzil, M. M. (2009). Discriminant analysis of animal species odors response. *Advances in Natural Science*, 2(2), 21-27.
- Rogers, S. L., & Webber, K. L. (2014). Student loan default: Do characteristics of four year institutions contribute to the puzzle? *Journal of Student Financial Aid*, 44(2), 99-124.
- Steiner, M., & Teszler, N. (2003). *Multivariate analysis of students loan defaulters at Texas A&M university*. Produced by Texas Guaranteed in Association with Texas A&M University.
- Steiner, M., & Teszler, N. (2005). *Multivariate analysis of student loan defaulters at Texas A&M University*.
- Tabachnick, B., & Fidell, L. (1996). In *Using Multivariate Statistics* (pp. 56 -110). New York.: Harper Collins Publishers.
- Thobe, T. M., & Deluca, B. M. (1997). A model for predicting perkins loan defaulters. *Journal of Student Financial Aid*, 27(1), 31-43.
- Uddin , N., Meah, S. M., & Hossain, R. (2013). Discriminant analysis as an aid to human resources selection and human resources turnover minimization decisions. *International Journal of Business and Management*, 8(17), 153-169.
- Volkwein , F., & Szelest, B. (1995). Individual and campus characteristics associated with student loan default. *Research in Higher Education*, 36(1), 41 - 72.

- Volkwein, F. J., Alberto, C. F., & Napierski, P. M. (1998). Factors associated with student loan default among different racial and ethnic groups. *The Journal of Higher Education*, 69(2), 206-237.
- Wishern, R. W., & Johnson, R. A. (2007). *Applied multivariate statistical analysis*. Pearson Education Inc.
- Woo, J. H. (2002). Factors affecting the probability of default : Student loans in California. *Journal of student finanacial aid*, 32(2), 5 - 23.
- Ziderman, A. (2004). Policy options for student loan schemes: lessons from five Asian. *International Institute for Educational Planning*, 1(6).
- Ziderman, A., & Shen, H. (2009). Student loans repayment and recovery: international comparisons. *Higher Education*, 315-333.