ORDINAL LOGISTIC REGRESSION MODEL ON THE FACTORS AFFECTING STUDENTS PERFOMANCE IN STATISTICS: A CASE STUDY OF MAASAI MARA UNIVERSITY.

NASAMBU CAROPHINE

BS03/012/2012

Research project submitted in partial fulfillment for the award of Bachelor of Science degree in applied Statistics with Computing, School of Science, Department of Mathematics and Physical Science, Maasai Mara University.

MAY 2016.

DECLARATION

I declare that this research project is my original work and has not been presented for a degree or any award in any examination body.

STUDENT DETAILS

DATE: _____ SIGN: _____

NASAMBU CAROPHINE

BS03/012/2012

SUPERVISOR'S DECLARATION

This research project has been submitted for examination with my approval as the university supervisor.

DATE: _____ SIGN_____

MR. KIPKOECH CHERUIYOT

DEDICATION

I dedicate this project to my beloved parents whose financial and moral support has been instrumental in the successful completion of the research, department of mathematics and physical science and my fellow classmates.

ACKNOWLEDGEMENT

Most of all, I thank the Almighty God for giving me strength, good health and wisdom to carry out the research peacefully and in union.

I also wish to acknowledge all whose inputs immensely contributed to success of this project. Particularly I wish to recognize the inputs of my university supervisor Mr. Cheruyoit for his unending assistance throughout the research work. I also wish to appreciate our Statistics lecturers who impacted statistical knowledge and their invaluable support on the information required. I cannot forget mentioning fellow students who participated actively in responding to the questionnaires.

ABSTRACT

As a result of poor performance in statistics introductory course among the students in the universities; the study wanted to find out the factors that influence the performance of students.

The purpose of the study was to conceptualize and empirically test ordinal logistic model on the factors affecting students' statistics performance in public universities. Students' performance was indicated by the grades they scored in the past statistical courses they undertook. The predictor variables to be studied were: high school mathematics performance, drug abuse, time for study, class size, class attendance, premature closure of the school to end year examination, belief and attitude, place of residence and gender. A detailed questionnaire of closed ended was used to collect data from the second year students who have had an introductory statistics course in Maasai Mara University and have done at least one unit of statistics. The data was entered in Microsoft excel and the imported to R Studio for analysis of the final project presentation. The results from the study was used as guidance in producing an effective and innovative remedy towards student's statistics performance in universities. The odds ratio of lack of interest is high indicating that it is a crucial factor that is affecting performance.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Definition of terms

Performance- includes activities which ensure that goals are consistently being met in an effective and efficient manner.

Logistic regression –is a regression model where the dependent variable is categorical.

Odds ratio- is a measure of association between an exposure and an outcome

Factors – are the subjects that influence performance.

Abbreviations

SATS: - Students Attitude towards Statistics.

Log: - Logarithm

ML: - Maximum Likelihood

OLR: - Ordinal Logistic Regression.

ATS: - Attitude Towards Statistics.

Por: - Place of residence.

CHAPTER ONE

1.1 Introduction

School, colleges and universities have no worth without students. Students are most essential asset for educational institute. The social and economic development of the country is directly linked with students' performance in different disciplines. Nowadays; the grades especially for statistics and mathematics courses are specified for one to be considered for employment. Many companies and institutions have cut off points in statistics related courses which is used as the reference for one to be employed. A big question is that, why do we have a declining graph in statistics performance? The students' performance plays an important role in producing the best quality graduates who will become great leaders and manpower for the country thus responsible for the country's economic and social development (Ali et.al 2009). Students' academic performance measurement have received considerable attention in the previous study ; it is a challenging aspects of academic literature and statistics students' performance are affected due to social, economic, environmental and personal factors. These factors strongly influence on students' performance but they vary from person to person and country to country.

High failure rate at tertiary institution result in unacceptable levels of attrition reduced graduate throughput and increased cost of training a nation's labor force. It is imperative that diagnostic studies are carried out to identify the major factors that are associated with suboptimal academic performance in statistics with a view of instituting corrective measures to improve the performance in all subjects.

Statistics related course is compulsory for all students who are not in sciences faculty in Maasai Mara University. Statistics is a methodology discipline that offers other fields of study a coherent set of ideas and tool for dealing with data. Therefore the study about its performance and factors that affect it should be studied to enhance its improvement in performance. The course is essential for many advanced courses in mathematics and statistics. Difficulties in learning and teaching statistics are well known by teachers and students' performance has been a subject of interest for many researchers during the last few years. Not only in statistics that students have poor performance but also in other subjects therefore subject of research study e.g. (Iduseri & Edokpa 2011) and (Mamta & Sandra 2014).

Consequently, there must be some inhibiting factors militating against students' performance. The interest of this study will be identifying such factors among the students in statistics department of the university in order to provide basis for improvement intervention strategy on course delivery.

1.3 Research question

What are the important factors that affect students' academic performance in the introductory statistics courses in university?

1.4 Research objectives

1.4.1 General objective

To explore the important factors that affect the students' academic performance in the introductory statistics course other than the one studied earlier.

1.4.2 Specific objectives

a) To identify the best predictors of statistics performance
b) Provide instructors and students with a means of remediation.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Logistic regression model

In statistics, logistic regression or logit regression is a regression model where the dependent variables – that is where it can take only two values such as pass/fail, alive/dead, healthy/diseased or win/lose. Logistic regression was developed by a statistician called David Cox in 1958. It measures the relationship between categorical dependent variable and one or more dependent variables by estimating the probabilities using a logistic function.

Logistic regression is widely applied and used in many fields including the medicine and social sciences. It is used in engineering for predicting probability of failure of a given process, system or product. In economics it can be used to predict the likelihood of a person choosing to be labor force and in business application it would predict the likelihood homeowner defaulting on mortgage.

There are several studies on factors that inhibit students' performance in courses at the undergraduate. Some of the factors identified are high school performance in mathematics (Russel and Julie 2006), belief and attitude (Kloosterman & Stage 1992)and (Kloosterman et al,.1996) and socio-economic(Akanle 2007). The approach of studying this factors is based on logistic regression method (Aromolaran et al. 2013), (Omar et al.2002), (Angela et al 2013) and (Adejumo 2012).

A number of studies have been carried out to identify and analyze the numerous factors that affect academic performance in various centers of learning. Their findings identify students effort, previous schooling (Siegfried &Fels,1979; Anderson & Benjamin, 1994), parents education, family income (Devadoss & Foltz ,1996), self-motivation , age of the student, learning preferences (Aripin, Mahmood,Rohaizad, Yeop, &Anuar, 2008),class attendance (Romer, 1993)and entry qualifications as factors that have a significant effect on the students' performance in various settings. The utility of this study lies in the need to undertake corrective measures that improve the

academic performance of students. Although there have been considerable debates about the determinants of academic performance among the educators, policymakers, academics and other stakeholders, it is generally agreed that the impact of these determinants vary with context, for example, culture, institution and course of study. Since not all factors are relevant for a particular context, it is imperative that formal studies be carried out to identify the context- specific determinant for sound decision making. The literature review provides a brief examination of some of the factors that influence academic performance.

2.2 Class attendance and academic performance

Class attendance is, in this sense a good measure of students' engagements. It has been also directly related to positive effect on exam performance. In his widely cited paper, Romer (1993) is one of the first few authors to explore the relationship between student attendance and exam performance. The major reasons given by students for non- attendance include assessment pressures, poor delivery of lectures, timing of lectures and work commitments (Newman-Ford, Lloyd & Thomas, 2009). Research on this subject seems to provide a consensus that students who miss classes perform poorly compared to those who attend the classes (Devadoss & Foltz, 1996; Durden & Ellis,1995).

2.3 Other determinants of academic performances

The influence of age and gender on academic performance has been investigated in a number of studies with widely differing conclusions. Most of the differences in reported findings are due to varying context such as subject of study, age and gender interactions. Research has shown that men perform better than women in a certain settings while women outperform men on other settings. Socioeconomic status of students and their families show moderate to strong relationship with academic performance (Sirin, 2005).

2.4 Student attitude towards statistics

Most graduate students enrolled in social and behavioral sciences program worldwide are required to take at least one statistics course and/ or a quantitative based research methodology program (Onmwuegbuzie 2003). It is widely acknowledged that statistics and quantitative methods courses cause problems(Murtonen and Lehtinen 2003a,b) especially for students in social sciences, who

generally have less interest and schooling in mathematical subjects. As a result factors related to success in statistics courses have been the subject of research.

As Lalonde and Gardner stated (1993), most of the variables that have been examined regarding the acquisition of statistical knowledge fall within three broad categories: anxiety, attitudes and ability.

Several scholars have addressed the influence of attitudes towards statistics (Bude' et al. 2007; Cashin and Elmore 2005; Chesi and Primi 2010; Gal and Ginsburg1994; Schau et al.1995). The general conclusion of these studies is that more positive attitude relate to better exam results.

A few early studies have already addressed the importance of mathematical skills for achievement in statistics courses (Chesi and Primi 2010; Harlow et al 2002) pointed out that one of the reasons that students have difficulties grasping the fundamental ideas of probability is the fact that many students have underlying difficulties with rational number concepts and basic concepts involving fractions, decimals and percentages.

2.5 Statement of the problem

Students often consider introductory courses in statistics as a difficult subject to learn. These courses are described as difficult, boring, a waste of time and not relevant by non-statisticians. This may lead to poor performance of the course. Therefore it is of great importance to study the factors that affect the performance at large and generate a model to predict on the performance level of a student. Thus the achievement in statistics will make them view themselves worthy.

2.6 Research gap

There have been many studies of the factors affecting students' performance in different subjects. The choice of the subjects and the factors affecting its performance differ from on setting to another. Some of the factor that has been investigated include: age, gender, family stress, class size, financial constraints, students learning and attitude in other countries e.g. Nigeria and U.S.A. There is a little known on these factors in Kenya since conclusions are made on what was studied in different countries. Furthermore there have been no studies on how drug abuse, closure of the school due to strikes, time used for studies and performance in mathematics in previous

examination. In conjunction to this like to investigate how this factor affect the performance in statistics using logistic regression model.

CHAPTER THREE

3.0 RESEARCH METHODOLOGY

3.1 Introduction

This section presents the study methodology used in attempt to achieve the objectives of the study. The chapter provides the data source and collection procedures, description of the target population, research design and the statistical model used for analysis development process.

3.2 Research design

A survey that was in form of questionnaires was administered. The survey design enabled indirect observation through the use of structured closed ended questions. The study allowed for large number of students to be included in the study and was appropriate to use in gathering information regarding the factors that affect the students' performance in statistics and general subjects by means of a questionnaire at a specified.

3.3 Target population

The target population were the students who have had an introductory statistic course per semester in Maasai Mara University in Kenya. The sample size was selected randomly from different disciplines that introduce the student to statistical course as compulsory and their performance in statistics. The population constituted students from different schools found in the university. It will consisted of students who were in session to enhance consistency of data and response

3.4 Study variables

3.4.1 Dependent variable

Statistics academic performance

3.4.2 Explanatory variables

Peer influence

Class attendance

Class size

Place of residence

Strikes that end the semester prematurely

Drug abuse

Interest

3.5 Data source

The source of the data was primary. This was because it was easy to collect and administer. The door to door technique was used to give out the questionnaire among the students.

3.6 Sample and sampling techniques

The research adopted simple random sampling technique. The sampling frame constituted of: the second years students who have undertaken an introductory statistics course.
The appropriate sample size for the study was determined using the following formulas from the data that will be collected

$$n = \frac{z^2 pq}{d^2} \qquad 3.1$$

Where n is the sample size

p is probability of success

q is the probability of failure

d^2 is the level of significance

$$n_1 = \frac{N}{1 + \frac{N}{n}} \qquad 3.2$$

Where n_1 the new sample is size and N is the total population.

n_1 =190.

3.7 Data collection Instrument

A closed ended questionnaire was used to collect data. The questionnaire was used to assess the performance of the students in statistics. The questionnaire was used because it was typically more efficient, economical and practical besides allowing for the use of a larger sample (fraenkell&Wallen, 2000). Standard instructions were given to all the respondents but the researcher also explained questions that were not clear to the respondents (Ary, Jacobs &Razavieh, 1979). Administration and scoring of a structured questionnaire was straightforward and the results lend themselves readily to analysis (Mugenda&Mugenda, 1999).

3.8 Pilot study

The questionnaire designed based on the research questions needed a pilot study to refine the questions before administering to the selected sample. The reason for the pre test is that it led to improvement and revealed areas of weakness about the questions to be administered. Therefore it gave advance warning about where the main research project could fail, where research protocols may not be followed, or whether proposed method are inappropriate or too complicated. The accuracy of the data to be collected largely will depend on the data collection instruments in terms of validity and reliability.

3.9 Validity

Content validity was employed to measure the degree to which the data that was to be collected using the instrument used to collect data represents a specific domain or content of a particular concept. To establish the validity of the research, was to seek opinion of experts in the field of study especially the lecturer. This helped improve the content validity of the data that was collected. It also facilitated the necessary revision and modification of the tool thereby enhancing validity

3.10 Statistical model

3.10.1 Multiple logistic regression model

Logistic regression analysis is popular and widely used analysis that is similar to linear regression model except that the outcome is dichotomous (e.g. success/failure). In essence we will examine the odds of an outcome occurring (or not) that is basing on the study will determine the odds of a factors affecting the performance in statistics, and by using the natural log of the odds of the outcome as the dependent variable the relationship can be linearized and treated much like multiple linear regression.

Multiple logistic regression analysis applies when there is a single dichotomous outcome and more than one independent variable. Outcome indicating pass/fail in statistics course and define p as the probability of y to be 1; p=prob (y=1) that the outcome is 1, the multiple regression model will be written as follows:

Assumptions of ordinal logistic regression model

The ordinal logistic regression model has the following assumptions:

a) The dependent variable assumes its distribution is within the range of exponential family distribution.
b) Proportional odds or the parallel line regression assumption: the relationship between each pair of outcome group is the same since the coefficient that describes the relationship

between the lowest versus all higher categories of the response variable, are the same as those that describe the relationship between the next lowest and all higher categories (Ombui, 2011.)

c) Logistic regression model does not assume a linear relationship between the dependent and the independent variables but a linear relationship between the logit of the response and the explanatory variables and further does not require the independent variable to be in an interval or unbounded.

d) The model assumes that the error terms are independent and that all the variables are included in the regression model. There is no assumption of a normal distribution over the error term.

3.10.2 Ordinal logistic regression model

Ordinal logistic regression model was used to model the ordered categorical response variables and only applies to the data that meet the proportional odds assumption. The model fits multiple logistic regressions on a multi-category ordered response variable that has been dummy coded. The response variable y will be a categorical response variable with k+1 categories

$$y = \begin{cases} 0, \textit{poor} \\ 1, \textit{fair} \\ 2, \textit{good} \\ 3, \textit{excellent} \end{cases} \quad ..$$

The proportional odds assumption states that the number added to each of the set logarithms to get the next is the same in every case to form an arithmetic sequence has the linear regression function with the regression parameters given as:

$$\beta_j = \left(\beta_{1j}, \beta_{2j}, ..., \beta_{pj}\right) \quad \text{where j=0, 1, 2… k}$$

And a set of predictor variable vector with p predictor variables

$$X_i' = \left(X_{1p}, X_{2p}, ..., X_{ip}\right) \quad \text{Where i=1, 2, 3… n}$$

The link function was a transformation of the probabilities that allowed for estimation of the model, in OLR model the link transformation is logit. The logit equation will form a comparison the log odds, and is given as below:

$$\log\left[\frac{\pi_1 + \pi_2 + \pi_3}{\pi_{j+1} + \pi_{j+2} + ... + \pi_j}\right] = \log\left(\frac{P\left(Y = j / X_i\right)}{P\left(Y = 0 / X_i\right)}\right) = X_i' \tag{3.3}$$

The response probability will then be uniquely defined and thus their summation is equal to one.

$$P\left(Y = 0 / X_i\right), P\left(Y = 1 / X_i\right), ..., P\left(Y = j / X_i\right)$$

$$\sum_{j=1}^{k} P\left(Y / X_i\right) = 1 \tag{3.4}$$

$$\pi_{ij} = \frac{\ell^{\sum_{k=1}^{j} X_i \beta_k}}{1 + \ell^{\sum_{k=1}^{j} X_i \beta_k}} \tag{3.3}$$

$$\pi_{1j} = \frac{1}{1+\ell^{\sum_{k=1}^{j} X_i \beta_k}} \qquad (3.5)$$

$$\log\left[\frac{P\left(x_1, x_2, x_3 \ldots x_p\right)}{1+P\left(x_1, x_2, x_3 \ldots x_p\right)}\right] = \beta_{0j} + \beta_{1j}X_{1j} + \beta_{2j}X_{2j} + \ldots + \beta_{pj}X_{pj} \quad (3.6)$$

The outcome of the logistic regression analysis were coded 0 or 1 where 1indicates that the outcome of interest is present, and 0 that the outcome of interest is absent. Let y be the binary outcome indicating pass/fail in statistics course and define p as the probability of y to be 1 ; p=prob (y=1) that the outcome is 1, the multiple regression model will be written as follows:

$$\hat{p} = \frac{\ell^{\left(\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6\right)}}{1+\ell^{\left(\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6\right)}} \qquad (3.7)$$

$\hat{p}$ Is the expected probability that the outcome is present.

x_i are distinct independent variables to be studied where $_i$ =(0,1,2,3,4,5,6)

β_i Are the regression coefficients where $_i$ = (0, 1, 2, 3, 4, 5, 6).

3.10.3 The likelihood function

The likelihood function for a given model is interpreted as the joint probability of the observed outcomes expressed as a function of the chosen model (Dietz 2005). It was useful when investigating the contribution of more than one predictor. The likelihood function is given as

$$L(\beta) = \prod_{i=1}^{n} \pi_{i1}^{yi1} \pi_{i2}^{yi2} \ldots \pi_{ij}^{yij} \left(1 - \pi_{i1} - \pi_{i2} - \ldots - \pi_{ij}\right)^{1-yi1-yi2-\ldots-yij} \qquad (3.8)$$

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\ell^{\sum_{k=1}^{j} X_i \beta_k}}{1 + \ell^{\sum_{k=1}^{j} X_i \beta_k}} \right)^{yij} \qquad (3.9)$$

$L(\beta)$ Is the likelihood function

The maximization process to estimate the coefficients will be accomplished by getting the log of the likelihood function

3.10.4 Log odds

We have to transform probability to log odds because it is usually difficult to model a variable which has restricted range, such as probability. This transformation maps the probability ranging

between 0 and1 to log odds ranging from negative infinity to positive infinity. The other reason is that the log of odds is the easiest to understand and interpret. The transformation is called logit transformation.

$$\ln\left(\frac{\hat{p}}{(1-\hat{p})}\right)$$

$$\ln\left(\frac{\hat{p}}{(1-\hat{p})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 \quad (3.10)$$

Where $\hat{p}$ the probability of performing well and 1- $\hat{p}$ is the probability of not performing well

3.10.5 Goodness of fit

The likelihood ratio statistic computed as twice the difference between log-likelihoods from the two models, and are referred to the chi-square distribution for significance testing. Likelihood ratios used to test for the contribution of the variables in an ordinal logistic regression model to the overall relationship between the dependent variable and individual variable. Because the likelihood for the full or larger model must be larger than the likelihood for the nested model the difference will be always positive. Therefore using the logit model the likelihood ratio R squared value given as:

R-squared=2*[log-likelihood (full model)/log-likelihood (nested model)]* (3.11)

Deviance statistics

Deviance statistics involves the comparison of the log of the maximum of the full model and a null model- a model with the intercept only.

D= -2[log (nested model)-log (full model)] (3.12)

This gives the deviance statistics based on a comparison of full and nested models of the ordinal logistic regression model equation. Whenever the log likelihood of the null model is small relative to the log likelihood of the full model, then the value of Dis large, an indicator that the current model is not good hence a small value of D gives a good model for use.

3.11 Statistical analysis

The data collected in the research was entered in the Microsoft excel and the imported to R studio for analysis. Also SPSS software was used in the analysis of descriptive statistics. The data was coded according to ordinal logistic regression. The data that was collected was manipulated to fit the student's statistics performance in Kenya by the process of bootstrapping. Multiple ordinal logistic regression analysis will be done to enable us to determine the specific factors that affect students' performance in statistics through the interpretation of the odds ratio and probability.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the results, analysis and discussion of the data that was collected during the study. The purpose of the study was to evaluate the factors affecting students' performance in the introductory statistics course by use of ordinal logistic regression model in Maasai Mara University.

The study aimed at achieving the following objectives:

a) Identify the core factors that affect students' performance in statistics course by use of ordinal logistic regression model.
b) To identify the best predictors of statistics performance
c) Provide instructors and students with a means of remediati

4.2 DEMOGRAPHIC CHARACTERISTICS

4.2.1 Distribution by gender

Table 4.1: Gender Distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	97	51.1	51.1	51.1
	Male	93	48.9	48.9	100.0

	Total	190	100.0	100.0

The background characteristics from the study showed that the response was distributed diversely among the gender as shown below in table 1

The sample size n=190 respondents. This contributed (n=97, 51.1%) female and (n=93, 48.9%) male implying that the number of women was slightly higher than male by n=4.

Age of the respondents

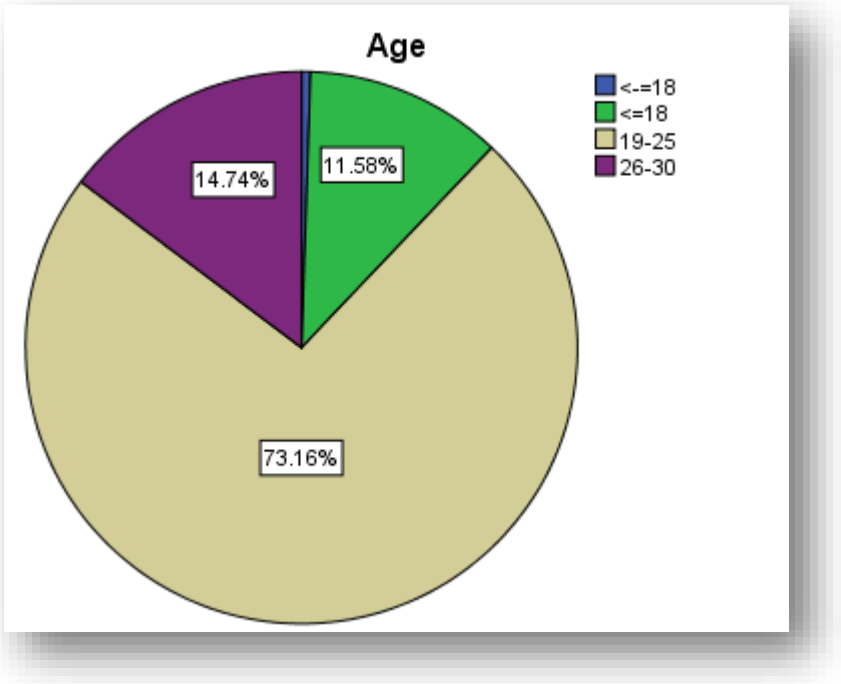

Figure 4.1: Age of the respondents

The population comprised of the age bracket of <=18,19-25, 26-30 and above 30. The distribution by the age showed that those in the age bracket 19-25 (n=139,73.2%), this indicates that the university students are in the age bracket of 19-25.this was followed by 26-30(n=28, 14.7%) and finally greater than 30 contributed (n=22,11.9%) as indicated in the pie chart below

4.2.2 Place of residence

Table 4.2: place of Residence

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Hostel	62	32.6	32.6	32.6
	Off campus	128	67.4	67.4	100.0
	Total	190	100.0	100.0	

The students indicated where they currently live and it was one of the factor that was examined to affect performance. As indicated in table3 (n=128, 67.4%) students live off campus while (n=62, 32.6%) got accommodation in the hostels. Therefore most students live in rental house.

4.2.3 Entrance grades in mathematics

This was another factor that contributed to the performance in the introductory statistics course in campus.

Table 4.3: Entry marks in Mathematics

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	45	23.7	23.7	25.3
	B	50	26.3	26.3	51.6
	C	60	31.6	31.6	83.2
	D	32	16.8	16.8	100.0
	Total	190	100.0	100.0	

The entry mark were measured in terms of grade attained in KCSE mathematics because it is related to the statistics course that is offered in the campus to non-statisticians. Most of the students scored C (n=60, 31.6%) and those with grades A and B were (n=45, 23.7%) and (n=50, 26.3%) respectively. This means that students join with above average grades. The big question is that why do we have decrease in long run? We obtained the answer by examining other factors that influence performance in statistics apart from performance in mathematics.

4.2.4 Gender distribution and the final performance in statistics

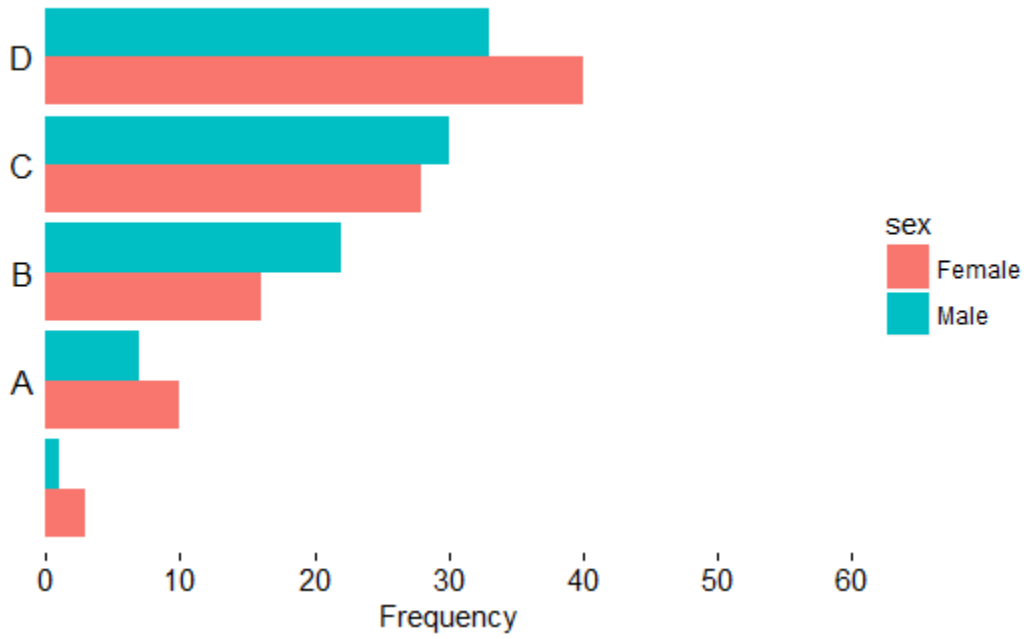

Figure 4.2: Gender Distribution and the final performance in statistics

The graph below shows the gender distribution of performance after a statistical course. From the graph women are to have performed poorly with the grade D. Also most female got grade A in statistics.

4.3 Ordinal logistic regression

Table 4.1: Showing the coefficients of the factors affecting Statistics

Coefficients:

	Value	Std. Error	t value
Por	0.54184	0.3434	1.5779
Attend class	0.09907	0.4716	0.2101
Extra time	0.05070	0.4391	0.1155
Friends	0.38478	0.5948	0.6469
Class size	-2.77485	0.8825	-3.1444
Strike	-0.25440	0.5066	-0.5021
Drug abuse	1.05977	0.4287	2.4721
Interest	3.19589	0.5793	5.5164

The ordinal logistic regression has four intercepts one for each cut point of the outcome as shown in table 5. The values of the intercepts are not so meaningful. The coefficients of the independent variables (factors affecting introductory statistical courses) are shared by the four cutoff points of the dependent variable that is (A, B, C and D). The coefficients of the place of residence (0.542), class attendance (0.099), extra time (0.051), peer pressure (0.385), drug abuse (1.060) and interest (3.200) are all positive indicating that the risk of performing poorly increases with these factors. Class size (-2.775) and strike (-0.254) have negative coefficients indicating that it protect both levels of performance. The estimated coefficients in the output are given in units of ordered log odds. One unit increase in place of residence, we expect 0.54184 increase in the expected value on the performance on the log odds scale given all the other variables in the model are constant.

One unit increase in class attendance, we expect 0.09907 increase in the expected value on the pe rformance in statistics on the log odds scale given all the other variables in the model are constan

t. Also we expect one unit increase in extra time spent in studying statistics and association with t he friends who share the same subject, we expect 0.05070 and 0.38478 increase in the expected v alue on the performance in statistics on the log odds scale respectively given all the other variab les in the model are constant.

4.5 Odds ratio and the confidence interval

Table 4. 5: Showing the odds ratio and the confidence intervals

	OR	2.5 %	97.5 %
Por	1.71916774	0.880097529	3.3923882
Attend class	1.10414456	0.434886240	2.7834275
Extra time	1.05200642	0.442629188	2.4929461
Friends	1.46929539	0.460460596	4.7840397
Class size	0.06235877	0.008244509	0.3035505
Strike	0.77538501	0.285152867	2.0984666
Drug abuse	2.88569444	1.256826423	6.7889742
Interest	24.43181601	8.053104473	78.4774864

The upper and the lower confidence interval of the independent variables does not cross 0, thus they are statistically significant and therefore are required in the model. The coefficients of the odds ratio are called proportional of the odds. For place of residence, one unit increase in performance of statistics that is going from D (Low) to A (High) is 1.72 times greater given that all other factors in the model are held constant. The variable interest has the highest odds ratio of 24.43.

4.6 Test of parallel lines

Table 4.6: Test of Parallel lines

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	Degrees of freedom	Sig.
Null Hypothesis	268.724			
General	153.849[b]	114.875[c]	36	.230

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

b. The log-likelihood value cannot be further increased after maximum number of step-halving.

c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model.

Validity of the test is uncertain.

The significance value of the parallel line test is greater than alpha at 0.05 hence we have to accept the null hypothesis that the location parameters are the same across the response categories and hence the link function used is appropriate

4.7 Pseudo R-Squared

Table 4.7: showing the Pseudo R-Squared

Pseudo R-Square

Cox and Snell	.557
Negelkerke	.597
McFadden	.302

Table 12 shows that the maximum values of Cox and Snell and Negelkerke R-squared attained is less than 1.These values indicate that the ordinal logistic regression model is useful in predicting the effects of place of residence, class attendance, extra time, friends, strike, drug abuse and interest in predicting introduction to statistics performance in the campus

4.8 Goodness of fit of the model

Table 4.8: Showing Goodness of fit of the model

Goodness-of-Fit

	Chi-Square	Degrees of freedom	Sig.
Pearson	697.549	216	.000
Deviance	228.988	216	.026

Link function: Logit.

The deviance statistic is a measure of the lack of fit between the data and the model, hence large values of deviance statistic indicate a poor fit in the model of the data. The p value of the deviance is equal to 0.026, a value less than 0.05, thus the observed data is consistent with the values in the fitted ordinal logistic analysis.

4.9 The model fitting information

Table 4.9: Showing model fitting information

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	Degrees of freedom	Sig.
Intercept Only	423.405			
Final	268.724	154.681	12	.000

The -2 Log Likelihood of the model with only intercept is 423.405 while that of the model with the intercept and independent variables is 268.724. The difference (Chi-square statistic) is

423.405-268.724=154.681 that is significant at alpha=0.05and the p value is less than 0.001 hence there exist an association between the students' performance in the introductory statistics course and the independent variables used in the study.

CHAPTER FIVE

CONCLUSION

Place of residence, class attendance, extra time, friends, drug abuse, strike attitude and interest were significant in determining the students' performance in the introductory of statistics course in the university. Majorly the interest of the student towards the subject had the greatest odds ratio indicating that it has greater influence in the performance of the subject. Gender and age were not directly related to the performance. Generally, there is a great difference between the entry marks in mathematics and the end results in statistics. Most students scored Bs and Cs while the final performance most scored Ds. This shows that despite the entry mark there are other factors that affect performance. For the model fit all the values were significant with p value less than 0.001. Also to the gender distribution female are performing poorly compared to male. Most of the female scored grade D in the final paper.in summary the factors affecting performance at large are many and therefore should be determined so as to provide long lasting solution to the universities.

Recommendation

A policy on the student records and follow up in the performance is necessary in order to enable different schools offering introductory statistical courses to effectively analyze the individual performance level and if possible offer assistance whenever necessary.

The ministry of education and relevant bodies that carry out the selection of courses to students to consider students' performance in the key subjects should test the students before undertaking the course because most of the students fill it is a waste of time.

There should be an improvement in the courses being taught so that students can see the importance of statistics course being offered in the campus.

The panel that is concerned with the courses should have an educative forum that organizes talks to students about their performance so that they can realize its importance in the nearby future

Limitation of the study

It was not easy to get data especially to students since they had busy schedule.

It was expensive to print the questionnaires in terms of cost.

Areas of further study

Other factors other than stated in the study should be studied to determine how they affect performance in different subjects other than statistics.

References

Adejumo, A.O. (2013). Application of Ordinal Logistic Regression in the study of students performance. *Mathematics Theory and Modeling.* http://www.iiste.org/Journals/index

Ahlgren, A., Garfield, J., & (1998). Difficulties in learning basic concepts in probability and statistics implications for research. http://www.amstat.org/publications/jse/v9n3/keeler.html

Akanle, O.B. (2007). Socio-economic factors influencing students' academic performance in Nigeria. http://www.medwelljournals.com/abstract.

Ali 2009. The students' performance plays an important role in producing the best quality graduates who will develop the country. http://www.academia.edu/6252608

Anderson,G., & Benjamin, D. (1994). The determinants of success in university introductory economic courses. http://www.alliedacademies.org/articles/what-factors-influence-firstyear-students

Angela, C., Maria, P. & Belen, M. (2013). Predicting academic performance. http://www.iiste.org/Journals/index.php/MTM/article/viewFile/25766/26332

Aromolaran, A.D., Oyeyinka, I.K., Benjamin, E.(2013). Binary logistic regression of students academic performance in Nigeria by Socio-Demographic and Economic Factors. http://www.iiste.org/Journals/index.php/MTM/article/viewFile/25766/26332

Aripin, R., Mahmood, Z., Rohaizad, R., Yeop, U., & Anuar, M. (2003). Students learning styles and academic performance.

Bude', L., Van De Wiel, M., Imbos, T., Candel, M., Broers, N., & Berger, M. (2007). Students' achievement in the statistical courses in relation to motivational aspects and study behavior. *Statistics Educational Research Journal, 6(*1*),5-21.*

Chasin, S., & Elmore, P.(2005). The survey of attitude towards statistics scale: construct validity study. *Statistics Educational Research Journal,* 9(1), 6-26.

Chesi, F., & Primi, C. (2010). Cognitive and non-cognitive related o students' statistics achievement

Devadosss, S., &Foltz, J.(1996). Evaluation of factors influencing students attendance and performance. *American Journal of Agricultural Economics, 78(3), 499-507.*

Durden, G., C., & Ellis, L.V. (1995). The effects of attendance on student learning in principles of economics.

Gal, I., & Ginsburg, L. (1994).the role of beliefs and attitude in learning statistics :towards assessment framework. *Journal of Statistics Education,2*(2) http://www.amstat.org/publications/jse/v2n2/**gal**.html

Galli, S., Chesi, F., & Primi, C. (2011). Measuring mathematical ability needed for "non-mathematical" majors.

Iduseri, A., & Edokpa, I.W. (2011). Dicriptive Discriminant Analysis of English language.

Kloosterman, P., & Stage, F.K. (1992). Measuring the belief about mathematical problem solving. *School of Science and Mathematics.* http://www.iiste.org/Journals/index.php/MTM/article/viewFile

Mamta, S. & Sandra W., (2014). Effect of High School Performance American. http://pubs.sciepub.com/education/2/4/1

Murtonen and Lehtinen 2003). Statistics and quantitative methods course cause problem. *Interest and schooling mathematics subjects.* http://users.ugent.be/~wduyck/articles

Murtonen and Lehtinen 2003aNewman-Ford, L., Lloyd, S., & Thomas, S. (2009). An investigation in the effects of gender, prior academic achievement, place of residence, age and attendance of first-year undergraduate attainment.

Romer, D. (1993). Do students go to class? Should they? http://sites.arbor.edu/studentsuccess/files/2011/11/**Class-**attendance-article1

Shakhawat Hossian, S. Ejaz Ahmed and Hatem A.Howlader, (2012): model selection and parameter estimationof a multinomiallogistic regression model

Sirin, S. R. (2005). Socioecomomic status and academic achievement.

Wasanga P., LewinK.M Wanderi E., Somerset A.,(2011): participation and performance in education in the Sub-Saharan Africa with special reference to Kenya.

APPENDIX

Questionnaire

The purpose of this questionnaire is to secure information from students. All the information supplied will be treated confidentially.

Please tick where appropriate

SECTION A

1. What is your gender?

Male [ ]

Female [ ]

2. What is your age?

18 or younger [ ]

19-25 [ ]

26-30 [ ]

Above 30 [ ]

3. Year of study

First [ ]

Second [ ]

Third [ ]

Fourth [ ]

4. In which school do you belong?

Science [ ]

Business [ ]

Education [ ]

SECTION B

5. Place of residence

Hostel [ ]

Off campus [ ]

6. What grade did you get in Mathematics?

A [ ] B [ ] C [ ] D [ ]

7. Are you/ did you do any introductory statistics course in this university?

Yes [ ] No [ ]

8. If yes, do you attend class regularly during the semester?

Yes [ ] No [ ]

9. Apart from other involving activities in campus, do you have extra time for the statistics course revision?

Yes [ ] No [ ]

10. Do you share the course with your friends?

Yes [ ] No [ ]

11. Approximately how many students attend the class?

[ ]

12. Does the number stated above affect your performance in class and examination of the unit?

Yes [ ] No [ ]

13. We have been affected by the strikes and demonstration in our university which have led to the indefinite closure of the campus, does this also affect your performance in statistics

Yes [ ] No [ ]

14. Do you take alcohol?

Yes [ ] No [ ]

15 Do you like and have interest to continue studying other statistics units offered in the University?

Yes [ ] No [ ]

16. How did you perform in the previous statistics unit?

A [ ] B [ ] C [ ] D [ ]

Thank you.