Continental J. Applied Sciences 11 (2): 1 - 15, 2016 ISSN: 1597 – 9928
© Wilolud Journals, 2016 http://www.wiloludjournal.com
Printed in Nigeria doi:10.5707/cjapplsci.2016.11.2.1.15

RESEARCH ARTICLE

# A LOGISTIC REGRESSION MODEL ON ACADEMIC PERFORMANCE OF STUDENTS' IN MATHEMATICS.

M. Mustapha[1] F. W Usman[2] Suleiman Yusuf[3]
[1,3]University of Maiduguri and [2]Ramat Polytechnic, Maiduguri
Department of Mathematics and Statistics, University of Maiduguri, P. M. B 1096, Maiduguri, Borno State.

## ABSTRACT

*In recent years 100 level (part 1) students offering element of calculus (MTH101) have being faced with problems ranging from failure in the course to low grades, which may lead to low CGPA, repeat and spill over among students. This study is aimed at making a model that relates performance of students in MTH101 together with some independent variables that can affect the students' performance and also recommend suitable intervention strategies that could improve the rates of academic performance of students offering the course. The study reveals that the factors that significantly influence academic performance in MTH101 course are; lecturers does not come to lecture theaters regularly, loss of interest in Mathematics because of the course and that the students does not discuss ideas from lectures, or read from the course with people outside the class.*

**KEYWORDS**: *CGPA, Element of Calculus, Logistic Regression, MTH 101*

Received for Publication: 11/09/16 Accepted for Publication: 16/11/16
Corresponding Author: aisamiram@gmail.com

## INTRODUCTION

Element of calculus (MTH101) is a course in the Department of Mathematics/Statistics offered by one hundred level (part one ) students of Faculty of Science, Faculty of Engineering, College of Medical Sciences, Faculty of Agriculture, Faculty of Veterinary Medicine, Faculty of Education and Pharmacy as a pre-requisite course in the University of Maiduguri.

Calculus provides the foundation for understanding higher-level science, mathematics, and engineering courses (Gainen and Willemsen, 1995). It also provides the mathematical background and foundation for future science courses. The importance of succeeding in first year calculus among freshman science students has been emphasized in several studies (Gainen and Willemsen, 1995). Despite the importance of calculus, certain difficulties inhibit students from leaning it, leading to unprecedented failure. Due to poor performance in calculus among freshman students in the last ten years, the undergraduate calculus course has attracted an unprecedented level of national interest (Bonsangue and Drew 1995). Some of the calculus-learning inhibition factors can be particularly related to the characteristics of first-year calculus (Burton, (1989). According to Burton, first-year calculus course is relying on some specific mathematics skills that students are presumed to have mastered in the secondary school.

In recent sessions one hundred level (part 1) students offering MTH101 have being faced with problems ranging from failure in the course, low grades which led to low CGPA, repeat and spill over among students. This was due to the lack of fundamental knowledge of mathematics in their secondary schools, lack of concentration in the course and some due to lack of convenience in the lecture halls. Hence the study is necessary to know whether this problem is of effect in the performance of the students.

**Literature Review**

Several studies have ascertained if demographic factors, previous experiences, or background are associated with students' course performance.

However, there are many factors which have an impact on study performance and progress. Some factors are at student level, some at institutional or programme level and others at structural level (Van den Berg and Hofman, 2005). In a comprehensive study in the Netherlands they find that variance in study progress and performance is largely determined by student factors. Such factors might be study techniques, ability, age, sex, motivation, time spent on paid work and time spent on studies. Wang and Goldschmidt (1999) examined the performance of mathematics courses taken by students at the junior and high school level. This study reported that the number of mathematics courses taken at the junior and high school level plays a prominent role in higher level mathematics achievement.

A study by Johnson (2006) also revealed that students who were more likely to earn a GPA higher than 2.00 were less likely to leave college for academic reasons. Johnson further suggested that high school GPA and SAT score had a positive effect on student persistence and performance and also added that college first-year GPA was the most important factor of student success. (Tai, Sadler and Loehr, 2005) reported that the number of high school science and Mathematics courses completed by students had little influence on students' performance in college science, whereas students' expectations were better predictors of their college science performance.

The teaching method, the support of the structure of the school, the family and students' attitude towards school affect the attitudes towards mathematics. Usually, the way that mathematics is represented in the classroom and perceived by students, even when teachers believe they are presenting it in authentic and context dependent way stands to alienate many students from mathematics (Barton, 2000; Furinghetti and Pekhonen, 2002). Researchers concluded that positive attitude towards mathematics leads students towards success in mathematics. Attempt to improve attitude towards mathematics at lower level provides base for higher studies in mathematics. It also causes effect in achievement of mathematics at secondary school level (Ma and Xu, 2004).

## MATERIALS AND METHOD

### Logistic Regression

Logistic regression is part of a category of statistical models called generalized linear models. This broad class of models includes ordinary regression and ANOVA, as well as multivariate statistics such as ANCOVA and loglinear regression (Agresti, 1996). Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the dependent or response variable is dichotomous, such as presence/absence or success/failure.

The dependent variable in logistic regression is usually dichotomous, that is, the dependent variable can take the value 1 with a probability of success θ, or the value 0 with probability of failure (1-θ). This type of variable is called a Bernoulli (or binary) variable. (Tabachnick and Fidell, 1996).

As mentioned previously, the independent or predictor variables in logistic regression can take any form. That is, logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The relationship between the predictor and response variables is not a linear function in logistic regression, instead, the logistic regression function is used, which is the logit transformation of θ (Menard, 1995).

$$\theta = \frac{e^{(\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k)}}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k)}}$$

An alternative form of the logistic regression equation was presented by Kleinbaum in 1994 which is given by:

$$Logit[\theta(x)] = log_e\left[\frac{\theta(x)}{1-\theta(x)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

The goal of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model.

**Parameter Estimation**

The method used to find the parameter estimates, α, $\beta_1$, $\beta_2$,…,$\beta_k$ is the method of maximum likelihood, which entails finding the $\hat{\beta}$ that maximizes the likelihood function.

$$f(y/\beta) = \prod_{i=1}^{N} \frac{n!}{Y_i!\,(n_i - y_i)} p_i^{y_i}(1-p_i)^{n_i-y_i}$$

The maximum likelihood equation is given by

$$L(y/\beta) = \prod_{i=1}^{N} \frac{n!}{Y_i!\,(n_i - y_i)} p_i^{y_i}(1-p_i)^{n_i-y_i}$$

Where

$$P_i = \frac{e^{\sum X_{ik}\beta_k}}{1 + e^{\sum X_{ik}\beta_k}}$$

Hence

$$L(\beta/Y) = \prod_{i=1}^{N} \left(e^{\sum X_{ik}\beta_k}\right)\left(1 - \frac{e^{\sum X_{ik}\beta_k}}{1 + e^{\sum X_{ik}\beta_k}}\right)^n$$

$$= \prod_{i=1}^{N} e^{\sum X_{ik}\beta_k}(1 + e^{\sum X_{ik}\beta_k})^n$$

$$logL(\beta/y) = \sum Y_i \sum X_{ik}\beta_k - n_i \log\left(1 + e^{\sum X_{ik}\beta_k}\right)$$

To obtain the critical points of the log likelihood function, we set the first derivative with respect to each β equal to zero, and solve for β.

$$\frac{\partial}{\partial\beta}\log(\beta/y) = \sum y_i X_{ik} - n_i \frac{1}{1 + e^{\sum X_{ik}\beta_k}} X_{ik} e^{\sum X_{ik}\beta_k}$$

$$= \sum y_i X_{ik} - n_i \frac{e^{\sum X_{ik}\beta_k}}{1 + e^{\sum X_{ik}\beta_k}} X_{ik}$$

$$= \sum y_i X_{ik} - n_i P_i X_{ik}$$

The maximum likelihood estimates for β can be found by setting each of the k+1 equations in (3.9) equal to zero and solving for each $\beta_k$. This can be done by iterative procedure using SPSS statistical package.

**Goodness of Fit Test**

Several different measures exist for determining goodness-of-fit of a logistic regression. These measures include: Likelihood-ratio test, Chi-square test, Hosmer-Lemeshow statistic and so on.

The likelihood ratio test is given by:

$$Y = \frac{\log of\ the\ \ null\ model}{likelihood\ of\ the\ given\ model}$$

Chi-square test can be based on the residuals, (Peng and So (2002). $y_i - \hat{y}_i$
A standardized residuals can be defined as

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}(1 + \hat{y}_i)}$$
$$\hat{y}_i(1 - \hat{y}_i)$$

Where the standard deviation of the residuals is $\hat{y}_i(1 - \hat{y}_i)$. One can then form a $\chi^2$ statistic as

$$\chi^2 = \sum_{i=1}^{n} r_i^2$$

This equation follows a $\chi^2$ distribution with n – (k + 1) degrees of freedom, so p-values can be calculated.

The Hosmer-Lemeshow statistic is given by

$$H - L = \sum_{k=1}^{g} \frac{(O_k - n_k^1 \bar{P}_k)^2}{n_k^1 \bar{P}_{k(1-\bar{P}k)}}$$

Where g denotes the numbers of groups, $n_k^1$ is the number of observations in the $k^{th}$ group, $O_k$ is the sum of the Y values for the $k^{th}$ group, $\bar{P}_k$ is the average of the $\hat{P}_k$ for the $k^{th}$ group.

**Statistical Significance of Individual Regression Coefficient**

If the overall works well, the next question is how important each of the independent variables is. The logistic regression coefficient for the $i^{th}$ independent variable shows the change in the predicted log odds of having an outcome for one unit change in the ith independent variable, all other things being equal. That is if the ith independent variable is changed 1 unit while all of the other predictors are held constant, log odds of outcome is expected to change $b_i$ units. There are a couple of different tests designed to assess the significance of an independent variable in logistic regression, the likelihood ratio test and the Wald statistic.

**Likelihood Ratio test**
The test statistic is calculated as follows:

$$G = -2In\frac{L_o}{L_1} = -2(InL_o - InL_1)$$

The statistics is compared with a $\chi^2$ distribution with 1 degree of freedom. To access the contribution of individual predictors one can enter the predictors hierarchically, and then compare each new model with the previous model to determine the contribution of each predictors.

**Wald Test**
The Wald statistic is asymptotically distribution as Chi-square distribution. Bewick *et al.,* (2005)

$$\left(\frac{\beta_j}{SE\beta_j}\right)^2$$

Each Wald statistic is compared with a Chi-square with 1 degree of freedom. Wald statistic are easy to calculate.

**Odds ratios with 95% CI**
An appropriate confidence interval for the population log odds ratio is

$$95\% CI\ for\ the\ In(OR) \pm 1.96X\{SEIn(OR)\}$$

Where In (OR) is the sample log odds ratio, and SEIn(OR) is the standard error of the log ratio (Morris & Gardner, 1998). Taking the antilog, we get 95% confidence interval for the odds ratio:

$$95\% CI\ for\ OR = e^{In(OR)\pm 1.96X\{SE(OR)\}}$$

**Prediction Accuracy**
The classification table is a method to evaluate the predictive accuracy of the logistic regression model (Peng and So, 2002). In this table the observed values for the dependent outcome and the predicted values (at a user defined cut-off value) are cross-classified. For example, if a cut-off value is 0.5, all predicted values above 0.5 can be classified as predicting an event, and all below 0.5 are not predicting the event.

**Method of data collection**
The data used are the grades of the (200-400) level students in MTH101 course which was obtained by using the questionnaire administered to 1000 students. The questionnaire covered largely the independent variables that were investigated. In the questionnaire, they were categorical data, multiple choice questions and likert scale.

## DISCUSSION OF RESULTS

### Bar Charts

A bar chart was used to display some of the demographic variables, which are Gender, age, place of residence as shown in Figures 1 to 3 below:

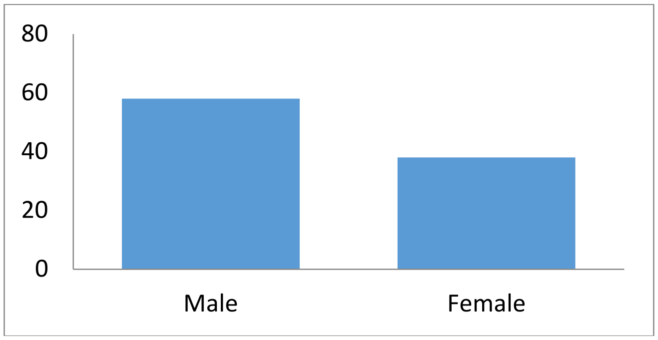


Figure 1: Gender Respondent

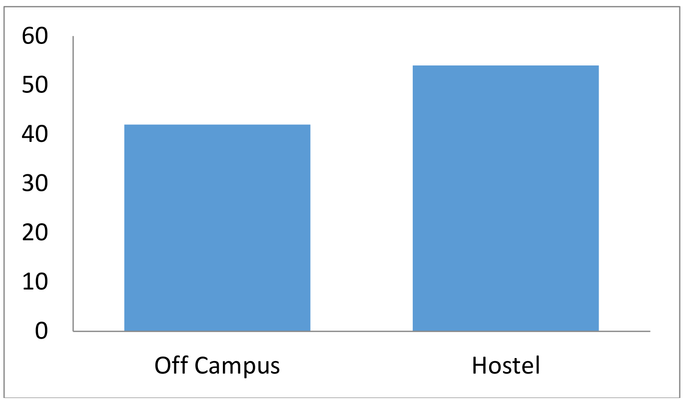


Figure 2: Respondents' Place of Residence

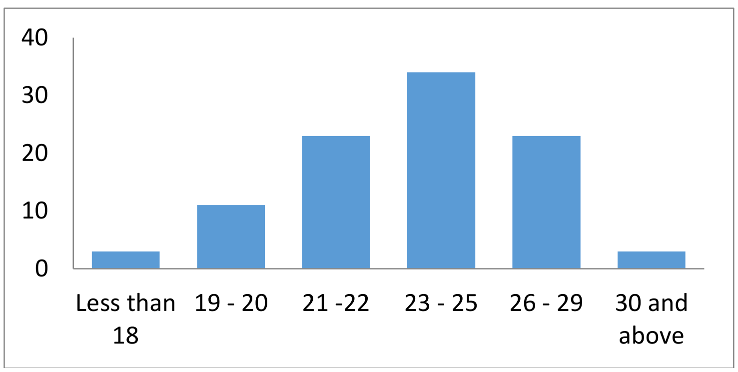


Figure 3: Age of Respondent

**Doughnut Chart**

A doughnut chart is used to represent Parent level of interest, categorization of grades in MATH 101, and JAMB/UME score which is shown in Figures 4 to 6 below:

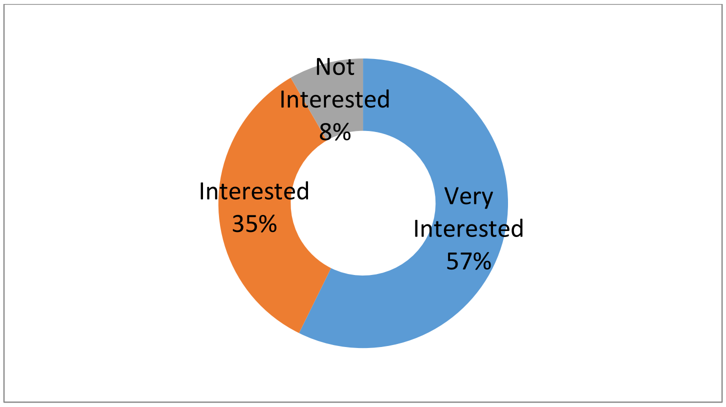


Figure 4: Respondent's Parents' Level of Interest

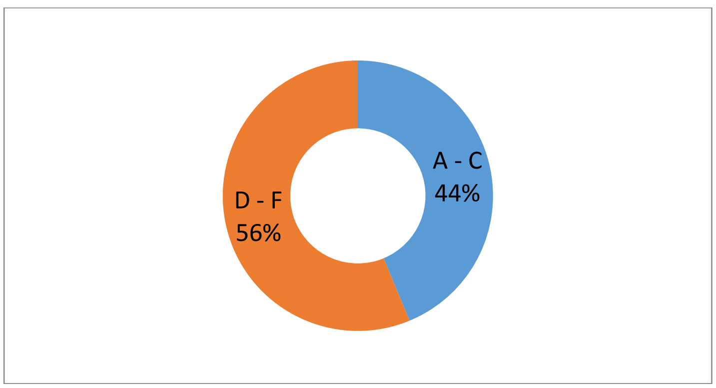


Figure 5: Categorization of grades of MATH101 of Respondent

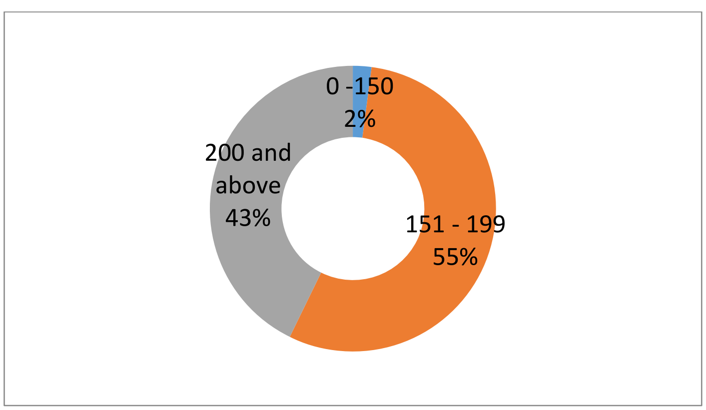


Figure 6: Respondents' JAMB Score

The result displayed in Table 1 gives information about the usefulness of the model. The Cox and Snell R Square and the Nagelkerke R Square values provide an indication of the amount of variation in the dependent variable explained by the model (from minimum value 0 to maximum value of approximately 1). These are described as pseudo R square statistic, rather than the true R square values in the multiple regression output. In Table 1 the two values are 0.179 and 0.238, suggesting that between 17.9% and 23.8% of the variability in the dependent variable is explained by the independent variable.

**Table 1 MODEL SUMMARY**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 125.557[a] | .075 | .101 |
| 2 | 119.060[b] | .136 | .181 |
| 3 | 114.178[b] | .179 | .238 |

The result displayed in Table 2 supports the model as being very useful. This test is a model fit test. For the Hosmer-Lemeshow Goodness of Fit Test, poor fit is indicated by a significance value less than 0.05, to support the model, value greater than 0.05 is actually required. In the above table, the chi-square value for the Hosmer-Lemeshow Test is 6.686 with a significance level of 0.571. Since the significance value is greater than 0.05, therefore it indicates that the model is good fit to the data.

The results displayed in Table 3 shows that the model correctly classified 63.5 percent of the overall cases, sometimes referred to as percentage accuracy in classification (PAC).

**Table 2 HOSMER AND LEMESHOW TEST**

| Step | Chi-square | df | Sig. |
|---|---|---|---|
| 1 | .325 | 1 | .569 |
| 2 | 3.131 | 6 | .792 |
| 3 | 6.689 | 8 | .571 |

**Table 3 CLASSIFICATION TABLE**

| Observed | | | Predicted | | |
|---|---|---|---|---|---|
| | | | MTH101 Grade | | Percentage Correct |
| | | | D-F | A-C | |
| Step 1 | MTH101 Grade | D-F | 35 | 13 | 72.9 |
| | | A-C | 25 | 23 | 47.9 |
| | Overall Percentage | | | | 60.4 |
| Step 2 | MTH101 Grade | D-F | 34 | 14 | 70.8 |
| | | A-C | 19 | 29 | 60.4 |
| | Overall Percentage | | | | 65.6 |
| Step 3 | MTH101 Grade | D-F | 30 | 18 | 62.5 |
| | | A-C | 17 | 31 | 64.6 |
| | Overall Percentage | | | | 63.5 |

The result displayed in Table 4 gives information about the contribution or importance of each of the predictor variables. The test that is used here is known as the Wald Test, and the values of the test statistic for each predictor is given in the column labeled Wald. The values less than 0.05 are labeled Sig. These are the variables that contributed significantly to the predictive ability of the model. In this case there are three significant variables (Lecturers not coming to class regularly, p=0.020, How MTH 101 affected their interest in Math, p=0.025, how often they discuss MTH101 outside the class, p=0.031). The table further shows that the major factors influencing whether a student will be successful in the course are Lecturer's not coming to class regularly, how MTH101 affected their interest in Math and how often they discuss MTH101 outside the class. The other factors did not contribute significantly to the model, since the stepwise logistic regression method was used, these factors were removed.

**Table 4 VARIABLES IN THE EQUATION**

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1[a] | How MTH101 affected interest in MATH | .814 | .309 | 6.944 | 1 | .008 | 2.257 | 1.232 | 4.136 |
| | Constant | -1.794 | .716 | 6.274 | 1 | .012 | .166 | | |
| Step 2[b] | Lecturer coming to class regularly | -.877 | .367 | 5.713 | 1 | .017 | .416 | .203 | .854 |
| | How MTH101 affected interest in MATH | .794 | .315 | 6.338 | 1 | .012 | 2.212 | 1.192 | 4.105 |
| | Constant | -.151 | .967 | .024 | 1 | .876 | .860 | | |
| Step 3[c] | Lecturer coming to class regularly | -.884 | .380 | 5.407 | 1 | .020 | .413 | .196 | .870 |
| | MTH101 affected interest in MATH | .739 | .329 | 5.049 | 1 | .025 | 2.093 | 1.099 | 3.986 |
| | How often you discuss MTH 101 outside class | .450 | .208 | 4.663 | 1 | .031 | 1.568 | 1.042 | 2.359 |
| | Constant | -1.142 | 1.103 | 1.071 | 1 | .301 | .319 | | |

The B values provided in the second column in Table 4 are equivalent to the B values obtained in multiple regression analysis. It is necessary to check if the B values are positive or negative. This will give the direction of the relationship (which factors increase the likelihood of the success and which factors decrease it). Negative B values indicate that an increase in the independent variable score will result in decreased probability of a student been successful in the course. Since two of the significant variables (how MTH101 affected their interest in Math, how often they discuss MTH101 outside the class) have a positive B values, it indicates that students with more interest in MTH 101 and they discuss MTH 101 with their colleagues outside the class are more likely to be successful in the course. And the other variable (Lecturer's not coming to class regularly) with negative B value, indicate that the less the lecturer comes to class the higher the failure and the more the lecturer comes to class the less the failure.

Thus, the B values will be used to developed the model,

$$= \frac{e^{(-1.142 \quad .884 lecturer\ not\ coming\ to\ cla \quad .739 interest\ in\ maths + 0.45 how\ often\ you\ discuss\ math\ outside\ class)}}{1 + e^{(-1.142 - \quad .884 lecturer\ not\ coming\ to\ class + 0.739 interest\ in\ maths \quad .45 how\ often\ you\ discuss\ math\ outside\ class)}}$$

Exp (B) and confidence interval results given in Table 4 indicate that the odds that a student with higher interest in MTH101 will be successful in the course is 2.093 times than students with less interest in MATH101 with a 95% confidence interval 1.099 to 3.986, the odds that a student who discuss MTH 101 with his colleagues outside the class will be successful in this course is 1.568 times than students who does not discuss MTH 101 outside the class with a 95%confidence interval of 1.042 to 2.359, and the odds that the lecturer's not coming to the class regularly will lead to students failure is 0.413 times than the lecturer is coming to class regularly with a 95% confidence interval of 0.196 to 0.870. In addition, the confidence interval (CI) of the odds ratio for the three significant variables did not contain 1, indicating that student with interest in MTH101, who discuss MTH 101 regularly with his colleagues outside the class and if the lecturer comes to class regularly will be successful in the course.

## CONCLUSION

The study reveal that the factors that influence the academic performance of students in MTH101 are; lecturer not coming to class regularly, students does discuss ideas form lectures, or reading from this course with people outside the class and students lack interest towards the course. The results obtained in this research supports the deduction Pisa, (2003) which examine that the standard tests and evaluations reveal that students do not perform to the expected level due to lack of interest in what they are studying. It is also in line with the comprehensive study in Maldives, Malaysia by Lawsha and Hussain, (2011), they find that students with negative attitude towards Mathematics spent less time in studying it, which influences their academic performance. In general, the students that receive low grades in MTH101 are likely to have; less interest in MTH 101, they are not discussing ideas from lectures, or reading from this course with people outside the class, and the lecturer does not come to class regularly.

## RECOMMENDATIONS

Based on the results obtained in the survey, the following recommendations are proposed:
1. One of the ways to solve the issues of lecturer's not coming to class regularly is; the department should make sure that lectures commence at the beginning of the semester so as to cover up the course outline at the stipulated time

2. The low performance of students due to lack of interest in mathematics can be remedied by counseling/orienting students about the importance of mathematics so as to make them develop interest in the course as soon as they are matriculated into the university.

3. The low performance of students due to lack of discussion of ideas from lectures, or reading from this course with people outside the class can be tackled by organizing the students a tutorial class to revised.

4. And lastly the students who are at risk of receiving low grades in this course can be reliably identified; this identification can be both from the classroom or tutorial class. It can also be identified through the issuance of continuous assessment reports, and also the use of quiz assessment in the tutorial, and implementation of proactive counseling by peers or instructors to make awareness on the implications of low grades and to offer them advice and assistance on improving their grades.

## REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley and Sons, Inc.

Barton, A. C. (2000), Crafting multicultural science education with pre-service teachers through service-learning. *Journal of Curriculum Studies,* Volume 32(6), pp. 799-820.

Bewick, V., Check, L and Ball, J. (2005). *Statistics Review 7:* Correlation and Regression, Grit care, 7:451-459.

Bonsangue, V. M., and Drew, E. E. (1995). Increasing minority students' success in calculus. In Gainen, J. and Willemsen, W. E. (Eds.), *New Directions for Teaching and Learning: Fostering student success in quantitative gateway courses* (Vol. 61, pp. 23-34). San Francisco: Jossey Bass.

Burn, N. and Grove, S. K. (2003). *Understanding nursing research 3rd Edition.* Philadelphia: W. B. Saunders Company.

Burton, M. B. (1989). The effect of Prior calculus Experience on "Introductory" College. *The American Mathematical Monthly, 96(4), 350-354.*

Furinghetti, F. and E. Pehkonen (2002). Rethinking characterizations of beliefs. In: G. Leder, E. Pehkonen, and G. Toerner (eds.), Beliefs: A Hidden Variable in Mathematics Education? *Kluwer Academic Publishers,* pp. 39-58.

Gainen, J., and Willemsen, E. W. (Eds.). (1995). *Fostering student success in quantitative gateway courses.* New Directions for Teaching and Learning (Vol.61). San Francisco: Jossey Bass.

Johnson V. E. (2006). Grade inflation:A crisis in college education *Springer Science and Business media.*

Kloosterman (1988). Self-confidence and motivation in mathematics. *Journal of Educational Psychology,* 80(3), 345-351.

Lawsha M. and Hussain W. (2011). Secondary Students' Attitude towards Mathematics in Selected Schools in Maldives. *International Journal of Humanities and Social Science Vol. 1 No. 15 (Special Issue – October 2011).*

Ma, X. and J. Xu (2004). Assessing the relationship between attitude towards mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education, Volume* 28(1), pp. 26-47.

Menard S. (1995). Applied Logistic Regression Analysis. Sage Publications. Series: *Quantitative Applications in the Social Sciences*, No. 106.

Nonis, S. A., and Hudson, G.I. (2010). Performance of College Students: Impact of Study Time and Study Habits. *Journal of Education for Business* 85: 229-238.

Parahoo K. (1997). *Nursing Research: Principle process, Issues.* London: Macmillan.

Peng, C. J. and So, T. H. (2002). *Logistic regression analysis and reporting.* Aprimer Understanding Statistics, 1(1), 31-70.

Pisa (2003). OECD Programme for International Student Assessment (PISA). Retrieved on 29 June 2011 from http://www.pisa.oecd.org/

Tabachnick , B. and Linda F. (1996). *Using Multivariate Statistics, Third edition.* Harper Collins.

Tai, R. H., Sadler, M. S., and Loehr, J. F. (2005).Factors Influencing Success in Introductory College Chemistry. *Journal of Research in Science Teaching,* 42, 987-1012.

Van den Berg, M.N., and Hofman, W.H.A. (2005). Student Success in University Education: *A Multi-measurement Study of the Impact of Student and Faculty Factors on Study Progress.* Higher Education 50: 413-446.

Wang, J., and Goldschmidt, P. (1999). Opportunity to learn, language proficiency and immigrant status on mathematics achievement, *The Journal of Educational Research,* 93 (2), 101 -111.