# A STUDY ON THE INFLUENCE OF SELECTED DEMOGRAPHIC CHARACTERISTICS ON ACADEMIC PERFORMANCE OF STUDENTS USING MULTIPLE LOGISTIC REGRESSION

**ODIOR, K. A.[1] and EKERIKEVWE, K. I[2]**

***Department of Statistics, Delta-State Polytechnic, Otefe- Oghara, Delta- State.**

**E-mail:** odifullness@gmail.com[1] , **kennedykerikevwe@gmail.com[2]**
**Phone number 0803 466 3466 0806 464 7455**

**ABSTRACT**

**In this communication an attempt was to empirically examine and investigate the influence of selected salient demographic characteristics on academic performance of student using the multiple logistic regression modelling. A multiple logistic regression model was fitted with cumulative grade point average (CGPA) as response variable while selected demographic characteristics as the explanatory variables. From the estimated multiple regression coefficients and test of significance of regression coefficients, it was evident that income/allowance of students and attendance to class has a significant influence on the CGPA. Also, Wald test equally revealed that income/allowance and attendance to class are statistically significant in the model. Thus, the reduced model which is parsimonious for the study were specified as**

$$P(CGPA \geq 2.50/X_2X_5) = \frac{e^{1.394(income)+1.278(attendance)}}{1+e^{1.394(income)+1.278(attendance)}}$$

**Keywords**: *CGPA, multiple logistic regression, demographic characteristics, academic performance*

## INTRODUCTION

Student's academic performance and achievement represent the extent to which a student has achieved their short or long-term educational goals in an institution of higher learning. Students' academic performance (SAP) and the graduation rate (GR) have recently become an issue of great concern to both institutional management and parents alike. Student's academic performance is naturally considered a critical aspect of a learning process as it serves as a feedback mechanism in monitoring academic advancement of individual students in an institution of learning. To this end, many learned articles examined specific factors particularly those affecting students' academic performance in tertiary institutions. The issue of academic performance has become a source of concern to government and the general public especially in the wake of declining standard of education and student's negative attitude towards academic activities. Academic performance of students is specifically measured by cumulative grade point average (CGPA). Students' academic performance are veritable means of revealing the extent of learning, acquisition of knowledge, skills, values, beliefs, norm and habits through the process of teaching, tutoring, training, preparation, inculcation, cultivation, guidance that is intended to bring about all round development, growth, improvement and bettering of the students' life both academically and character-wise.

The institution student's environment, family background/home, society are open system that are either vertically or horizontally interdependent and interconnected as there is great interface and interaction among them as regards academic performance of students. Therefore, the academic performance of students is influence by a number of factors. In the view of [2], the state of the home affects the individual since the parents are the first socializing agents in an individual's life. This is because the family background and context of a child affect his/her reaction to situations and his level of performance. [13] reported that parent constant disagreement affects children emotionally and this could lead to poor academic performance. He reported that the family financial support encouragement and following up have positive impact on students' academic performance as measured by their GPA. [17] concluded in their research that family income level, attending full-time, receiving grant/aid and completing advanced level classes in high school having statistically significant effect on college performance among first generation college students. [10] reported that student academic performance is affected by different factors such as learning abilities, gender and race. [16] conducted a prospective study to explore the psychosocial, cognitive and demographic predictors of academic performance of first year Australian University students. Results revealed that previous academic performance was identified as most significant predictor of

university performance. Integration into university, self-efficacy and employment responsibilities were also predictors of university performance. [11] conducted a study to find out the factors which affect college students' performance. In their study, the researcher mainly focuses on how to explore the factors that are associated with performance of students in intermediate examination. This study concludes that attitude towards attendance in class, time allocation for studies, parents level of income, mothers age and mother's education were main factors that affect performance of students of private colleges.

[7] carried out a study with fresh college students to evaluate the efficiency of students learning style and other university admission variable in predicting students' academic performance and retention. Act composite score, high school class rank, high school core GPA, and learning style were used as predictors. Results showed that core GPA and Act score were best predictors for predicting academic performance of first year of college.

[13] attempted to determine the academic success factors of business students focusing on factors potentially influencing their performance. They targeted gender, age, ethnicity and other demographic and academic variables. They found that students' performance is strongly correlated with demographic characteristics. More specifically and importantly, their study confirmed that demographic variables can be influential, as trait such as age, gender, or ethnicity has been shown to be correlated to academic performance.



[5], [18] and [8] conducted studies about the link between the students age and student academic performance. They found that mature students achieved higher grades than youthful students do. This result demonstrates a positive relationship between student's age and their CGPA. However, this is inconsistent with some other studies, which revealed that grades earned by younger students are higher than higher than mature students [15]. The fact that mature students have other responsibilities than study that create distractions may lead to poor performance comparing with young students who are concerned about their study only.

In addition, other researchers have continued to either elaborate on known relationship and association or attempt to discover new influencing variables. [4] confirmed that student's personal effort was strongly correlated with students' academic performance, concluding their effort in academic performance regardless of course content.

In summary, the factors that influence student's academic performance is debatable. Despite this, they still recommended that future research efforts should attempt to examine these variables in greater depth and coverage.

RESEARCH METHODOLOGY

The data for this study was collected using a designed questionnaire that was distributed to randomly selected students from different institutions of higher learning.

## LOGISTIC REGRESSION MODEL SPECIFICATION AND DEVELOPMENT

The term logistic regression analysis comes from logit transformation, which is applied to the dependent variable. Logistic regression is a generalized linear model used to model a binary categorical variable using numerical and categorical predictors. This study addresses the influence of selected demographic characteristics on academic performance of students as measured by cumulative grade point average (CGPA) using the multiple logistic regression analysis. The statistical tool provides a flexible general-purpose modeling strategy with straightforward interpretation and its application in research permeates social and biomedical sciences involving the need to predict the probability that an event will occur or not. [19] in his work employed the logistic regression to model staff performance using age, job knowledge and skills, educational qualification length of service etc. as influencing variables. The statistical model seeks to explain the relationship among multiple variables influencing job performance. [1] utilized the statistical tool in modelling accident fatality on Akure-Owo high way. The study reported that drivers are a major influence factor in road traffic crashes. It is a unique technique for modeling dichotomous (two category) Y variables but with polychotomous explanatory variables [9]. Multiple logistic regression analysis is intended to obtain the best fitting equation and most parsimonious model to describe the relationship between a categorical out-come variable and the one or more categorical or continuous predictors (independent) variables [9]. Multiple logistic regression analysis is instead a designation of one of two possible out-comes: alive or dead, success or failure, yes or no, defect or defect free by [2]. Over the years, logistic regression modeling has become applicable in many fields the standard method of analysis in this situation. Generally, in logistic regression, statistical theory as well as in practice the relationship between E(Y) and $X_1$, $X_2$…$X_p$ can be better specified by a non-linear equation.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 ... = \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 ... = \beta_p x_p}} \quad (1)$$

Given that the two values of the dependent variable $\pi(x)$ are coded as 0 or 1, the $\pi(x)$ in equation (1) provides the probability that $Y = 1$ given a particular set of value for the

independent variables $X_1, X_2, X_3, \ldots, X_n$. The transformation of $\pi(x)$ is pivotal in order to express $\pi(x)$ as linear functions of the regression parameters. The logit transformation can be defined as:

$$g(x) = \frac{\pi(x)}{1+\pi(x)} \tag{2}$$

$$g(x) = \ln\left[\frac{\frac{e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}}}{1+\left(\frac{e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}}\right)}\right] \tag{3}$$

$$g(x) = \ln\left[\frac{e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}} \times \frac{e^{\beta_0+\beta_1x_1+\beta_2x_2.}}{1}\right]$$

$$g(x) = \ln\left(e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}\right)$$

$$g(x) = \log_e e^{\beta_0+\beta_1x_1+\beta_2x_2\ldots=\beta_px_p}$$

$$g(x) = \beta_0 + \beta_1X_1 + \ldots + \beta_pX_p \tag{4}$$

The transformed g(x) has many of the desirable properties of a linear regression model. The logit g(x) is linear in its parameters may be continuous and may range from - ∞ to + ∞ depending on the range of x [12]. The regression parameters are estimated using the maximum likelihood approach. This can be achieved through the construction of likelihood function. Suppose, Y is coded as 0 or 1, $\pi(x)$ provides the set of parameters. Given that $P(y_i = 0/x)$, thus it follows that the quantity 1- $\pi(x)$gives the conditional probability that Y is equal to zero given x, $P(y_i = 0/x)$where $y_i = 1$, the contribution to the likelihood function is $\pi(x)$ where $y_i = 0$, the contribution to the likelihood function is 1- $\pi(x)$

The pair of $x_i$ can be expressed as

$$\pi(x_i)^{y_{ii}}(1-\pi(x_i)^{1-y_i} \tag{5}$$

In general, odds ratio enables researchers to compare the odds for two different events, since the observations are assumed dependent, the likelihood function is obtained as the product of the terms in equation (5).

$$\beta = \overset{n}{\pi}\pi(X_i)^y\left[(X_i)^{1-y}\right] \tag{6}$$

The likelihood expression of equation (5) yields.

$$L(\beta) = \ln(\beta) = \sum_{i=1}^{n} [y_i \ln(\pi(X_i)] n_i = 1 + (1 - y_i) \ln(1 - \pi(X_i)$$

To obtain the value of β that maximizes L(β), we differentiate L(β) with respect to the parameters partially and set the resulting expression equal to zero. Odd ratios are often used comparatively to describe the strength of an effect. Odds ratio is the ratio of the odds at two different values of X. They provide another excellent way to interpret logic coefficients. The odds in favor of an event occurring is viewed as the probability that the event will occur divided by the probability the event will not occur. In logic regression, the odds in favor of Y=1 can be specified as:

$$odds = \frac{p(y = 1/ X_{1,} X_{2,}, ..., X_p)}{p(y = 0/ X_1, X_2, ..., X_p)}$$

$$odds = \frac{p(y = 1/ X_1, X_2, ..., X_p}{1 - p(y = 1/ X_1, X_2, ..., X_p)_1}$$

Thus Odds ratio is estimated by $odss\ ratio = e^{\hat{Y}}$. Odds ratio measures the impact in the odds of a one- unit increase in only one of the independent variables. In general, odd ratio enables researchers to compare the odds for two different events.



DATA ANALYSIS

In this study academic performance of students was classified into two as define: 0 = CGPA < 2.50, 1 = CGPA ≥ 2.50). The following independent variables were considered in this study: age, gender, home environment, income/allowances, attendance to class, post UTME score, Jamb score and time allocated for study.

**The Wald Statistic**

The univariate Wald test compares the difference between a maximum likelihood point estimate of a single parameter and a hypothesized value to its standard error which is compared to chi-squared table value:

$$Wald = \frac{(\hat{\beta}_j - \theta)^2}{S^2_{\hat{\beta}_j}}$$

Where;

$\hat{\beta}_j$ is the estimate

$\theta$ is hypothesized value

$S_{\hat{\beta}_j}$ variance of estimate

The test statistic is compared against a chi-square distribution with one degree of freedom. The basis of this univariate procedure is that the maximum likelihood estimates are assumed to be asymptotically normal.

**Model Specification of the study**

DATA ANALYSIS

In this study academic performance of students was classified into two as define: 0 = CGPA < 2.50, 1 = CGPA ≥ 2.50). The following independent variables were considered in this study: age, gender, home environment, income/allowances, attendance to class, post UTME score, Jamb score and time allocated for study.

**Table 1: Variable specification**

| *Variable* | *Data type* | *Data coding* |
|---|---|---|
| **Age** | Quantitative | - |
| **Gender** | Qualitative | Male = 1, female = 0 |
| **Marital status of parents** | Qualitative | Married = 1, separated/divorced = 0 |
| **Income/allowances (₦)** | Quantitative | - |
| **Attendance in class (%)** | Quantitative | - |
| **Post UTME score** | Quantitative | - |
| **Jamb score** | Quantitative | - |
| **Time allocated for study** | Quantitative | - |
| **CGPA** | Qualitative | 0.99 – 2.49 = 0, 2.50 – 4.00 = 1 |



**Discussion of result**

**Table 2: Omnibus Tests of Model Coefficients**

| | | Chi-square | Df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 15.960 | 8 | .043 |

From the table of goodness of fit test on the fitted model, it stated that the chi-square test is statistically significant since the p-value is less than 0.05, implying that both the intercept and all other coefficients are nonzero. Hence, the conclusions that will be reached in this model will be efficient.

**Table 3: Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 25.495[a] | .413 | .551 |

a. Estimation terminated at iteration number 19 because parameter estimates changed by less than .001.

From the above table, the Nagelkerke R Square implies that the model is good (55.1%) and that 41.3% probability of an event of having a CGPA ≥ 2.50 is explained by the logistic regression model fitted as explained by the Cox and Snell R Square statistic.

**Table 4: Classification Table[a]**

| Observed | | | Predicted | | |
|---|---|---|---|---|---|
| | | | CGPA | | Percentage Correct |
| | | | 0.99 – 2.49 (0) | 2.50 – 4.00 (1) | |
| Step 1 | CGPA | 0.99 – 2.49 (0) | 11 | 3 | 78.6 |
| | | 2.50 – 4.00 (1) | 3 | 13 | 81.3 |
| | Overall Percentage | | | | 80.0 |

a. The cut value is .500

From the classification table, we found that 11% of the students' performance are correctly classified, and 3% are incorrectly classified hence, 78.6% of the students' performance are correctly classified in the CGPA of 0.99 – 2.49. Again, 3% are incorrectly classified whereas in the total percentage, 81.3% of the CGPA classification in 2.50 – 4.00 is correct. The overall correct percentage was 80% which implies that the model overall explanatory strength is efficient.

**Table 5: Model parameter estimates**

| | | B | S.E. | Wald | df | Sig. |
|---|---|---|---|---|---|---|
| Step 1[a] | Constant | 25675.344 | 40618.095 | .400 | 1 | .527 |
| | Age of students | .003 | .227 | .000 | 1 | .989 |
| | Income/Allowance of students | 1.394 | 0.564 | 6.111 | 1 | .013 |
| | Home environment of students | -18.993 | 29.186 | .400 | 1 | .527 |
| | Gender of students | 13.383 | 29.547 | .400 | 1 | .527 |
| | Attendance in class of students (%) | 1.278 | .380 | 11.310 | 1 | .001 |
| | Hours students allocated for study | .398 | .579 | .474 | 1 | .491 |
| | Jamb score of students | .022 | .015 | 2.132 | 1 | .144 |
| | Post UTME score of students | .079 | .044 | 3.278 | 1 | .070 |

***Assessing parameter efficiency***

***Using the P-value***

This study used a 0.05 level of confidence and the criterion for using the p-value was that a variable coefficient with p-value lesser than level of α is considered significant, insignificant otherwise. It is evident that the indicators of student performance (income/allowances and attendance to class) are statistically significant since their p-values are less than 0.05 level of significance. While the model intercept and the other indicators (age, gender, home environment, hours of study, jamb score and post UTME score) are statistically not significant, implying that they don't contribute to the academic performance students.

***Using the Wald Chi-square Statistic***

In using this approach to examine the model parameter contribution, the Wald test examines the parameters individually, causing the degree of freedom to be 1 for each parameter and the chi-square distribution value of level of significance ($\chi^2_{0.05,1}$) is 5.024. The Wald valued will be compared with the Chi-square distribution value (5.024). Based on the comparison, only two indicators of student performance (income/allowances and attendance in class) were significant. This is in support of the p-value approach and hence, the model can be reduced.

The reduced model is given thus as;

$$P(Y = 1/X_2X_5) = \frac{e^{1.394X_2+1.278X_5}}{1 + e^{1.394X_2+1.278X_5}}$$

$$P(CGPA \geq 2.50/X_2X_5) = \frac{e^{1.394(income)+1.278(attendance)}}{1 + e^{1.394(income)+1.278(attendance)}}$$

## Conclusion

This paper study the influence of selected demographic characteristics on cumulative grade point average (CGPA) of students in tertiary institution. It is evident from the analysis that income/allowance of students and attendance to class by students were statistically significant in the model. The findings of this study is in line the earlier results of [13], [17], and [4]. From the results of the analysis it is proven that age of students, gender, marital status, post UTME score Jamb score were weak predictors of CGPA of student in tertiary institution.

## References

1. **Adogoke, S. A. (2014). A Study on the Fatality on Akure-Owo high way using Logistic Regression. Journal of the Nigerian Statistical Association, Vol.26.1-10.**
2. **Ajila, C., and Olutola, A. (2007). Impact of Parents` Socio-Economic Status on University Students` Academic Performance. Ife Journal of Educational Studies, 7(1),31-39.**
3. **Bayaga, A. (2010). Multinomial Logistic Regression: Usage and Application in Risk Analysis. Journal of Applied Quantitative Methods, 5(2),288-297.**
4. **Clark, S., and Latshaw, C. (2012), Peeling the Onion: Called Students Performance; An Investigation into the Factors Affecting Student Performance in an Introductory Accounting Class. Review of Business, 33(1), 19-27.**
5. **Cohen, E. (1972). Students` characteristics and Performance in Economic Statistics. Journal of Economic Education, 3(2), 106-111.**
6. **Erum, S. and Zahoor, A. (2011). A Study on Academic Performance of University Students. International Conferenceon Recent Advances in Statistics, 255-268.**
7. **Garton, B. L.,Dyer, J. E. and King, B. O. (2000). The Use of Learning Style and Admission Criteria in Predicting Academic Performance and Retention of College Freshmen. Journal of Agriculture Education, 41(2),46-53.**
8. **Guney, Y. (2009). Exogenous and Endogenous Factors Influencing Students` Performance in Undergraduate Accounting Modules. Accounting Education, 18(1), 51-73.**
9. **Hamilton, L. C. (1991). Regression with Graphics: A Second Course in Applied Statistics. Books/Cole Publishing Company Pacific Grove, California.**
10. **Hanson, J. B. (2000). Student Performance and Student Growth as a Measure of Success: A Evaluator`s Perspective. Paper presented at the Annual Meeting of the American Educational Research Association New Orleans. Louisiana.25th April.**
11. **Hijazi, S. T. and Naqvi, S. M. (2006). Factors Affecting Students` Performance: A case of Private Colleges. Bangladesh e-Journal of Sociology, 3(1).**
12. **Hosmer, D. W. and Lemeshow, S. (2000). Applied Logistic Regression, 2nd Ed. John Wiley, New-York.**
13. **Ichado, S. M. (1998). Impact of Broken Home on Academic Performance of Secondary School Students in English Language. Journal of Research in Counseling Psychology, 4(1), 84-87.**

14. **Kaighobadi, M., and Allen, M. (2008). Investigating Academic Success Factors for Undergraduate Business Students. Decision Sciences of Innovative Education, 6(2), 427-436.**
15. **Lane, A. and Porch, M. (2002). The Impact of Background Factors on the Performance of Nonspecialist Undergraduate Students on Accounting Modules: A Longitudinal Study; Accounting Education International Journal, 11(1),109-118.**
16. **Mckenzie, K. and Schweitzer, R. (2001). Who Succeeds at University? Factors Predicting Academic Performance in First Australian University Students. Higher Education Research and Development, 20(1), 21-33.**
17. **Simmons, A. B., and Musoba, G. D. and Choong, C. G. (2005). Persistence among First Generation Students in Indiana: The Impact of Precollege preparation, College Experiences, and Financial Aid. Indiana Project on Academic Success (IPAS) Research Report.**
18. **Simpson, W. and Surnrall, B. (1979). The Determinant of Objective Test Scores by Finance Students. Journal of Financial Education, Fall, 58-62.**
19. **Runyi, E. F. (2017). The Application of Binary Logistic Regression Analysis on Staff Performance Appraisal. Science Journal of Applied Mathematics and Statistics, 5(4), 164-168.**