An Ordinal Logistic Regression Model to Identify Factors Influencing Students Academic Performance at Njala University

Regina Baby Sesay[*1], Mohamed Kpangay[2], Sheku Seppeh[3]

[1,2,3]Department of Mathematics and Statistics, School of Technology, Njala University, Njala, Sierra Leone

Abstract: **The academic performance of students in learning institutions like universities is a significant determinant of the future socio-economic status of the students concerned. An outstanding students' academic performance can be regarded as an instrument used to achieve rapid economic, social, political, technological and scientific growth in a developing country like Sierra Leone. It is, therefore, the desire of every university student to be on top of their classes in almost all their university courses. Despite the huge desire to excel, there are both academic and socio-economic factors influencing the academic performance of undergraduate students at Njala University. This work, therefore, aims to identify the main factors influencing students' academic performance at Njala University. For this purpose, a stratified random sampling method was employed to select 284 respondents proportionately from each university faculty. Data were collected from the selected respondents using structured questionnaires. An ordinal logistic regression modeling technique was used to identify the main factors influencing the academic performance of the undergraduate students at Njala University, Njala Campus. Several factors were initially considered as potential determinants of students' academic performance. However, the result of the empirical analysis revealed that, the number of study hours; father's income level; mother's educational level and mother's income level are the main factors influencing undergraduate students' academic performance at Njala University. An increase in the number of study hours increases students' academic performance; an increase in the father's income level increases students' academic performance; an increase in the mother's educational level increases student's academic performance while an increase in the mother's income level decreases students' academic performance.**

Keywords: **Ordinal Regression, Brant Test, Variance inflation Factor, proportional odd, academic Performance.**

I. INTRODUCTION

Students' academic performance is a major area of concern for almost all learning institutions including universities and colleges. In the school, and university system, student's academic performance is popularly defined by student's cumulative grade point average over the years. **It** is a measurement of student's achievement across various academic subjects in their years of study. University lecturers often measure academic achievement using classroom performance in the form of examinations and tests. Reference [24] mentioned in their work that academic performance consists of scores obtain by a student in assessments such as class exercise, class test, mid-semester mock examination and end of semester examination. A cumulative grade point average over the years of study, sometimes referred to as CGPA is often used as a measure of a student's academic performance for all of his or her courses. At Njala University, students' academic success is measured using both the sectional grade point average (SGPA) and the academic years' cumulative grade point average (CGPA). Students who meet the performance standards prescribed by the university are promoted to the next level (or year) of studies.

More importantly, students who do well in their courses at university are at greater advantage of coping with the technologically demanding occupations of the future and are better able to achieve occupational and economic success after graduation. A good academic performance helps the graduated student to find a good place in the society. Students who graduated with high quality degrees are more likely to have more employment opportunities than those who were graduates with low quality degrees. This is in line with the common saying among students, 'good academic performance result to a good standard of living.'

Also, increase in students' academic performance can decrease students' drop- out rate, and therefore reduce the possible crime rate in a country. This implies that an excellent academic performance is not only profitable to the individual student or to his or her parents but also to society.

Therefore, the desire of each, and every Njala University student is to academically top their classes in all the taught courses. However, despite the huge desire to excel academically, there are socioeconomic, academic, and environmental factors influencing the academic performance of most Njala University Students.

The socio-economic status of the family has a great impact on both the students' personality and his or her academic performance at all levels of the students' education. Social and economic status of student is generally determined by combining parents' qualification, occupation and income standard [14]. Students from families with good financial and educational background tend to perform much better than those from the poor and uneducated family backgrounds.

Parents of students from a well- to-do financial backgrounds can provide the latest technological facilities in a best possible way to enhance the educational capability of their children. Reference [21] clearly stated that children whose families are of high educational scales have a statistically far better chance of participating in Tertiary Education. Reference [2] also observed that in a modern society, family influence played a very important role in the academic lives of students. Reference [10] even noticed in their findings that parent's income or social status positively affects the student's test score in examination.

Again, the student's entry qualification plays a significant role in the academic performance of the student. A good student can be identified by his or her entry qualification. According to the input output system of learning, high-quality students admitted into the university system will produce high-quality university graduates. As the quality of a university is reflected in its graduates, the quality of students that a university admits into its system is a significant determinant of the quality of graduated students that the university produces. Most universities, therefore, lay emphasis on the admission of those students who could perform better in the relevant university degree programs.

To have good quality graduates from Njala University, the university, like most other learning institutions all over the world has set two admission criteria, the university, and the departmental admission requirement criteria. These are the two admission requirements criteria that must be fulfilled before a student can be admitted in to any degree program at Njala University. This view is supported by [22], who discovered in his findings that students with high scores in the standardized eligibility scores performed better. In addition, many other studies carried out on the relationship between entry qualification and academic performance have discovered that students with high entry qualification often perform better than those with low entry qualifications (for instance: [1] , [3] and [25]).

Again, the students' university academic performance is greatly influenced by the type of school that they attended. Reference [23] observed that students previous educational outcomes are significant indicators of students future achievement, This means that, the higher the quality of the previous academic performance the better the student's university academic performance in future endeavors.

This research work therefore, used an ordinal logistic regression approach to identify -the main factors influencing students' academic performance at Njala University.

II. MATERIALS: THEORETICAL FRAMEWORK

This section deals with theories and concepts used in the empirical analysis

A. Regression Models for Ordinal Dependent Variables

In the literature, there exists various regression models used to analyze data with an ordinal response variable (e.g. [18], [9], [7] and [8]). These include: cumulative logit model; constrained and unconstrained partial odds models; continuation ratio models; polytomous logistic model; stereotype logistic model and adjacent-category logistic model.

In the context of ordinal logistic regression, ordinal means order of the categories. The ordinal logistic regression is, therefore, a regression technique used when the dependent variable is measured at the ordinal level, given one or more explanatory variables, which could be ordinal, continuous or categorical. Therefore, when the outcome variable is polychotomous and ordinal in nature, the best choice model often used to preserve information about the ordering of the categories of the dependent variable is the ordinal logistic regression model. It is mostly considered as a generalization of the binomial logistic regression model. The ordinal logistic regression model for ordinal dependent variables was first considered by [19] and is used to predict an ordinal dependent variable given one or more independent variables.

In ordinal logistic regression, instead of modelling the probability of an individual event, as we do in logistic regression, we are considering the probability of that event and all others above it in the ordinal ranking. We are concerned with cumulative probabilities rather than probabilities for discrete categories.

Hence the model;

$$logitP\big((Y \leq j)\big) = \beta_{j=0} - \beta_{j=1}x_1 + + \cdots + \beta_{j=p}x_p for\ j = 1, \dots, j-1 .$$

With P predictors is called the ordinal logistic regression Model

This model uses cumulative probabilities up to a threshold, thereby making the whole range of ordinal categories binary at that threshold. For instance, considering the level of the dependent variable used in the ordinal regression analysis let the response *be* $\mathrm{Y} = 1, \dots, \mathrm{J}$ with a natural ordering.

Also, let $\{\mathrm{p}_0, \mathrm{p},_1 \dots, , \mathrm{p}_{\mathrm{j}-1}\}$ be the associated probabilities. The cumulative probability of a response less than and equal to *j* is given as:

$$\mathrm{P(Y \leq j)} = \frac{\exp(\alpha_\mathrm{j} + \beta\mathrm{X})}{\big(1 + \exp(\alpha_\mathrm{j} + \beta\mathrm{X})\big)}$$

Where $log\left[\frac{P(Y \leq j)}{P(Y > j)}\right] = \alpha_j - \beta X, \quad j \,\epsilon\, [1, J-1]$

α**j** is the intercept and is the log-odds of falling into category ***j*** or below.

β_k is the parameter that describes the effect of the independent variable X_i on the dependent variable Y

The cumulative logit is given as:

$$log\left(\frac{P(Y \leq j)}{P(Y > j}\right) = log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right)$$

$$= log\frac{p_1 + \cdots + p_j}{p_{j+1} + \cdots + p_j}$$

The cumulative logit measures how likely the response is to be in category *j* or below versus in a category higher than *j*.

- Considering student academic performance as the , dependent variable with ordered categories: poor performance, good performance, very good performance and excellent performance. Let the proportions of students of the selected sample who would perform "poor", "good", "very good", and "excellent" be p_1, p_2, p_3, p_4 respectively . Then the logarithms of the odds of performing in certain ways are:

$$Poor, \quad log\frac{p_1}{p_2+p_3+p_4}; \quad 1$$

$$poor\ or\ good, \quad log\frac{p_1+p_2}{p_3+p_4}, \quad 2$$

$$poor, good\ or\ very\ good, \quad log\frac{p_1+p_2+p_3}{p_4}, \quad 3$$

Generally, incorporating the independent variables in to the model gives

$$poor \quad L_1 = \beta_{11} + \beta_{12}X_1 + \beta_{13}X_2 + \beta_{14}X_4 + \cdots + \beta_{j=p}X_p,$$

$$good \quad L_2 = \beta_{21} + \beta_{22}X_2 + \beta_{23}X_3 + \beta_{24}X_4 + \ldots + \beta_{j=p}x_p$$

Very good $L_3 = \beta_{31} + \beta_{32}X_2 + \beta_{33}X_3 + \beta_{34} + \cdots + \beta_{j=p}X_p$

$$excellent \quad L_4 = \beta_{41} + \beta_{42}X_2 + \beta_{43}X_3 + \beta_{44}X_4 + \cdots + \beta_{j=p}X_p, \quad p = 1,2,3,4$$

It should be noted that, the ordinal regression model is only valid or good to use when the independent variables have identical effects at each level of the dependent variable. (The proportional odds assumption) and when the independent variables are not highly correlated with each other (the multicollinearity assumption). This leads us to the main assumptions of using the ordinal regression model.

B. Assumptions of Ordinal Logistic Regression

The following are the main assumptions that the ordinal logistic regression makes about the underlying data:

- The dependent variable is ordinal.
- One or more of the explanatory variables are either continuous, categorical or ordinal..
- No multi-collinearity.
- The odds are proportional (proportional odd, or parallel lines assumption): This means that each independent variable has an identical effect at each cumulative split of the ordinal dependent variable. The parallel line assumption implies that, there is one regression equation for each category except the last category.

C. Link functions

Link functions are transformations of the cumulative probabilities that allow estimation of the model. Link functions are used to link the (cumulative) response to the set of predictor variables. The common default and most commonly used link function when fitting ordinal regression model is the logit link function. The logit link function allows for the effects associated with specific predictor variables to be expressed as odds-ratios. Depending on the distribution of the **ordinal** outcomes, there are other link functions that can be used in building ordinal regression. Some of these link functions are presented in Table I

Table I : Link Functions and Their Forms In Ordinal Regression

Link Function	Form	Common Application
Logit	$log\left(\frac{F_k(x_i)}{1-F_k(x_i)}\right)$	when categories are evenly distributed
Probit	$\Phi^{-1}(F_k(x_i))$	Normally distributed latent variable
Cauchit	$tan(\pi(F_k(x_i) - 0.5))$	Outcome with many extreme values used when the extreme values are present in the data
Complementary log-log	$log[(-log(1 - F_k(x_i)))]$	Higher categories are more probable. Recommended when the probability of higher category is high.
Negative log-log	$-log(F_k(x_i))$	This link function is recommended when the probability of the lower category is high.

III . METHODOLOGY

A. Study Area

The study was carried out at Njala University, Njala Campus. The University is located on the banks of the Taia River in the Kori Chiefdom, Moyamba district in the southern part of Sierra Leone.

1) Population and Sample: The target population consisted of all final year students for the 2019/2020 academic year. A stratified random sampling method was employed to select 284 respondents proportionately from each university faculty.

Data were collected from the selected respondents using structured questionnaires.

2) *Dependent Variable*: The dependent variable used in this study is the student's academic performance. Students' Academic performance is an outcome of education because it measures the extent to which a student, teacher or an establishment has achieved their educational goals [6] Students' academic performance is mostly measured by the Cumulative Grade Point Average (CGPA) [12]. The CGPA shows the overall students' academic performance as it shows the overall average of all examinations' grade for all semesters during the students' years of study in the university. It is even believed that a higher CGPA is an indication of better learning [4]. This study used CGPA as a measure of students' academic performance.

3) Description of Dependent Variable: The dependent variable used in this analysis is students' academic performance originally measured on the continuous scale using students' cumulative grade point average (CGPA) which was later categorized in order of merit ranging from poor performance to excellent performance. The CGPA ranges from one (1) point to five(5) points. Below three (3) point is considered a poor performance for which a student can repeat a program. Three point to three point five (3 to 3.5) is considered a good performance for which a student can be awarded a third class degree. Three point six (3.6) to four point two (4.2) is considered a very good performance for which a student can be awarded a second class degree and four point three to five point (4.3 to 5) is considered an excellent performance for which a student can be awarded a first class degree.

In short, let the academic performance of the final year undergraduate students be denoted by Y. Also, let Y be a categorical response variable with k+1 (k=3) ordered categories, coded as 1, 2, 3, 4, with higher values indicating higher level of academic performance. Then:

$$Y = Students' \, accademic \, performance$$

$$= \begin{cases} 1 = poor, & if \; CGPA \; is \; less \; than \; 3 \; points \\ 2 = good, & if \; CGPA \; lies \; between \; 3.0 \; to \; 3.5 \\ 3 = very \; good, & if \; CGPA \; lies \; between \; 3.6 \; to \; 4.2 \\ 4 = excellent, & if \; CGPA \; lies \; between \; 4.3 \; to \; 5 \end{cases}$$

1) Independent variables: The independent variables together with their descriptions are presented in Table II

IV. EMPIRICAL ANALYSIS

The main purpose of these analyses was to find out which variable(s) has significant effect on the academic performance of students at Njala University. Below is the descriptive and exploratory analysis.

Descriptive and Exploratory Data Analysis

A. Table II: Dependent (DV) and Independent (IV) Variables to Be Modeled

Variable Name	IV/DV	Valid Range	Variable Type
Student's academic Performance	DV	Excellent Very Good Good Poor	Categorical (Ordinal)
Father's Educational Level	IV	No Formal Education Primary Education Secondary Education Tertiary Education	Categorical (Ordinal)
Mother's Educational Level	IV	No Formal Education Primary Education Secondary Education Tertiary Education	Categorical (Ordinal)
Mother's Income Level	IV	Low Medium High	Categorical (Ordinal)
Father's Income Level	IV	Low Medium High	Categorical (Ordinal)
Study Hours		1-4 hours	Continuous
Student's Previous School	IV	Public Private	Categorical (Dichotomous)
Student's Admission Point	IV	Excellent Very Good Good Poor	Categorical (Ordinal)
Parents Marital Status	IV	Married Single	Categorical (Dichotomous)
Student's Residence		On Campus Out of Campus	Categorical (Dichotomous)

The bar plot presented in fig 1 displays the level of the academic performance for each ordered category of the dependent variable (Students' academic performance). From fig I, it is clear that the category of good academic performance is higher than all the other categories. However, the category of poor academic performance is higher than that of the category of excellent academic performance.

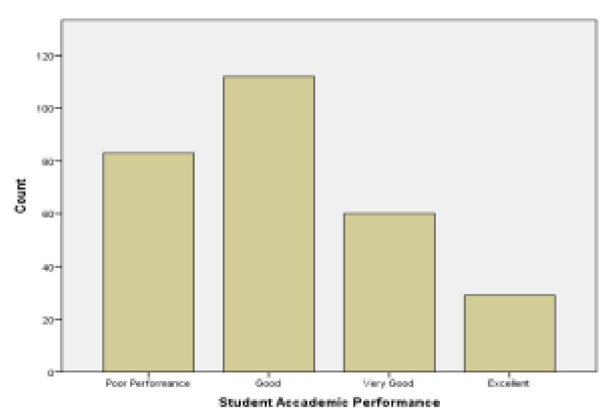

Fig1: Bar plot of Levels of the Outcome Variable

Table Ш: The Distribution of Categorical Variables Use

Case Processing Summary

		N	Marginal Percentage
Student Accademic Performance	Poor Performance	83	29.2%
	Good	112	39.4%
	Very Good	60	21.1%
	Excellent	29	10.2%
Previous School	Private	87	30.6%
	Public	197	69.4%
Mother's Income	very low	97	34.2%
	LOW	164	57.7%
	MIDDLE	19	6.7%
	HIGH	4	1.4%
Mother's Education	No Formal Education	89	31.3%
	Primary	40	14.1%
	Secondary	60	21.1%
	Tertiary Education	95	33.5%
Residence	On Campus	166	58.5%
	Out of Campus	118	41.5%
Father's Income	Very Low	87	30.6%
	Low	173	60.9%
	Middle	24	8.5%
Valid		284	100.0%
Missing		0	
Total		284	

Table **Ш** presents the case processing summary of the categorical variables. From Table **Ш**, the marginal percentage of poor academic performance is high and therefore needs immediate attention regarding identifying the main factors contributing to the poor academic performance of the students.

Table IV presents the Chi-square test of independence. The Chi-square test of independence was used to test whether there was a relationship between each categorical independent variable and the categorical ordinal dependent variable. The test works by comparing the observed frequencies to the expected frequencies. The null hypothesis of the Chi-square test is that, there is no relationship between the two categorical variables. This implies that, knowing the value of one variable does not help to predict the value of the other variable. The alternative hypothesis is that, the variables are dependent. This implies that, there is a relationship between the two categorical variables. That is, knowing the value of one variable helps to predict the value of the other variable.

From the output presented in Table IV, the p-values for each pair of dependent and independent variables are each less than the significance level of 5%. This leads to the rejection of the null hypothesis in favor of the alternative hypothesis for each pair of variables. Rejecting the null hypothesis of the Chi-square test of independence for each pair of dependent and independent variable means there exist significant relationships between the students' academic performance and each of the categorical independent variables considered in the ordinal regression analysis.

Table IV: Chi-Square Tests of Independence

Factor	Pearson $\chi 2$ Value	df	P value	Decision
Acc_per and Prev_Sch	31.948	4	.0000	Academic performance is dependent on the previous school of the student
Acc_per and Fa_Incom	78.022	8	.0000	Academic performance is dependent on the fathers' income level
Acc_per and Edu_Mo	117.73	12	.0000	Academic performance is dependent on the mothers' educational level
Acc_per and Mo_Incom	117.96	12	.0000	Academic performance is dependent on the mothers' income level

B. Ordinal Logistic Regression Analysis

Table V provides the model fitting information that contains the likelihood ratio chi-square test used to compare the full model against a null (intercept-only) model. The test result in Table V was, therefore, used to check if the final (present) model with explanatory variables included is an improvement over the intercept only model. At the 5% significance level, the test result presented in Table V is statistically significant. The significant chi-square statistic ($p<.0005$) indicates that the final model gives a significant improvement over the baseline or intercept-only model. This showed that a model including independent variables, student' previous school; mothers income; mothers education; residence and sturdy hour fits significantly better than the null (or intercept only) model. This further showed that at least one of the regression coefficients in the model is not equal to zero.

Table V: The Model Fitting Information

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	601.753			
Final	544.420	57.333	11	.000

Link function: Logit.

The R-square presented in Table VI gives information about how much variance is explained by the independent variable. The Cox and Snell, Nagelkerke's, and McFadden's pseudo-R^2 statistics were used to estimate the variance explained by the independent variables used in the ordinal logistic regression model. The pseudo R^2 values (e.g.Negelkerke.=0.841=84%) presented in Table VI indicate that the ordinal logistic regression model with its independent variables explains a relatively large proportion of the variation between students in their academic performances at Njala University. The present values (Negelkerke=.0.841, Cox and Snell= 0.790, McFadden=0.557) of R^2 indicate that a model containing students' admission point; father's income; mother's income; mother's educational level and study hours is most likely to be a very good predictor of the academic performance for any *particular* student.

Table VI: Pseudo R-Square Statistics

Pseudo R-Square

Cox and Snell	.790
Nagelkerke	.841
McFadden	.557

Link function: Logit.

Table VII presents the values of the regression coefficients and intercepts, together with the corresponding standard errors, t values and their p-values. The last four rows are the intercepts, sometimes referred to as the cut-points that represent the threshold of the ordinal logistic regression. For the ordinal logistic regression model, there are n-1 intercepts, where n is the number of categories in the dependent variable (Student performance). The intercept can be interpreted as the expected odds of identifying the listed categories when other variables in the model assume a value of zero. For example, for the first of the last four rows of Table VII, the intercept, 1|2 (i.e., poor performance\Good) takes the value of 1.8305, indicating that the expected odds of identifying a poor performance category, when other variables assume a value of zero is 1.8305. However, the intercept are not usually included in the interpretation of the ordinal logistic regression analysis result.

In Table VII, the variable with the largest coefficient with p-value less than the chosen significant level of 0.05 is considered the most significant influential factor.

Therefore, with the exception of the independence variables, residence and previous school, with p-values greater that the chosen significant value of 0.05, the p-value for all the other independence variables are each less than 0.05 (<0.05). This shows that all the independent variables, except residence and previous school are statistically significant at the 5% level of significance.

The first significant independent variable presented in Table VII is Fa_Incom (Father's income), with coefficient, 1.3989 and p-value = 0.0000< 0.05. This implies that, for a unit increase in father's income (e.g. from low level income to middle level income), we expect a 1.3989 increase in the ordered log odds of being in a higher level of academic performance given that, all of the other variables in the model are held constant.

In the case of Mo_Incom (mother's income) with negative coefficient -1.3062, and p-value = 0.00< 0.05, implies that for a unit increase in the level of mother's income (i.e., from low level income to middle or higher level income), we expect a 1.3062 decrease in the ordered log odds of being in a higher level of academic performance given that all the other variables in the model are held constant.

For Edu_Mo (mother's education) with coefficient, 0.4146 and p-value = 0.00< 0.05, implies that, for a unit increase in the mother's educational level (e.g. from no formal education level to primary, secondary or tertiary education level), we expect a 0.4146 increase in the ordered log odds of being in a higher level of academic performance given all of the other variables in the model are held constant.

Table VII Parameter Estimates Of the Ordinal Logistic Regression Model New

	Value	Std. Error	t-value	p-value
Prev_Sch	-0.2198	0.26191	0.8393	0.401
Fa_Incom	1.3989	0.33404	4.1879	0.000
Mo_Incom	1.3062	0.27916	4.6790	0.000
Edu_Mo	0.4146	0.10571	3.9223	0.000
Residence	0.0581	0.23647	0.2458	0.805
Stu_hrs	1.3063	0.26894	4.8575	0.000
1|2	1.8304	0.53941	3.3934	0.000
2|3	3.7874	0.58188	6.5088	0.000
3|4	5.3102	0.61365	8.6534	0.000

Table VIII: presents the confidence intervals for the parameter estimates.

The parameter estimate under consideration is statistically significant if the confidence interval (CI) does not include or cross zero (0), From Table VIII, apart from the CI of Residence and Pevious School (that include 0), the confidence interval for each of the other independent variables did not include or crossed 0. This is just a confirmation of the result presented in Table VIII.(the parameter estimate table),

Table VIII: Confidence Intervals for Parameter Estimates

	2.5 %	97.5 %
Prev_Sch	-0.7361482	0.2921077
Fa_Incom	0.7493274	2.0617608
Mo_Incom	-1.8600412	-0.7628337
Edu_Mo	0.2087891	0.6238368
Residence	-0.4045761	0.5234792
Stu_hrs	0.7858293	1.8417077

Table IX presents the coefficient parameters converted to proportional odds ratios and their 95% confidence intervals. The odds ratios were derived by exponentiating the coefficients parameters.

These odd ratios are strictly in line with the proportional odds assumption. This assumption states that, the odds ratios are the same across categories of the dependent variable. That is, for ordinal logistic regression, there are separate intercept terms at each threshold, but a single odds ratio. Thus the odd ratios are interpreted as follows:

The independent variable, previous school (Prev_Sch) was found to be an insignificant influential factor of students' academic performance. However, the odd ratio for Prev_Sch is 0.8026. This implies that:

For students from public schools, the odds of moving from "good" performance to "very good" or from "very good" to "excellent" performance (or from the lower and middle categories to the high category) is 19.74% lower [i.e., (1 - 0.8026) x 100%] than students from private school, holding constant all other variables.

For students from private school, the odds of moving from "good" performance to "very good" or from "very good" to "excellent" performance (or from the lower and middle categories to the high category) is 1.246 [i.e., 1/0.8026] times that of students from public school, holding constant all other variables (odds ratio>0).

Similarly, the odd ratio, for father's income is 4.05. This again implies that for one unit increase in the level of father's income that is going from low level of income to high level of income, the odds of moving from , "poor" performance to " good" performance; from "good" to "very good" performance or from "very good" to "excellent" performance (or from the lower and middle categories to the high category) are 3.486 times greater, given that all of the other variables in the model are held constant.``

Also, the odd ratio for mother's income is 0.271 and is less than one (<1). This is not surprising as the coefficient for mother's income (in Table VII) is negative. This implies that for a one unit increase in the level of mother's income that is going from low level of income to middle or high level of income, the odds of moving from , "poor" performance to " good" performance; from "good" to "very good" performance or from "very good" to "excellent" performance (or from the lower and middle categories to the high category) are 0.33 times lower, given that all of the other variables in the model are held constant.

Again, the odd ratio for mother's educational level is 1.39. This implies that for students whose mothers attained higher level of education, (secondary or tertiary education) the odds of good, very good or excellent academic performance (i.e., versus poor or very poor academic performance) are 1.39 times greater (i.e., increases 39%)than students whose mothers only had primary or no formal education holding all other variables constant.

Lastly, the odd ratio for students' sturdy hours is 3.69. This implies that, for every one unit increase in student's study hours, the odds of high academic performance (good, *very* good, or excellent versus poor) are 3.69 times greater, holding constant all other variables.

In other words, for every one unit decrease in student's study hours, the odds of poor or very poor academic performance are 3.69 times greater, holding constant all other variables.

The Variable, Residence, is not a significant determinant (p value > 0.05) of students' academic performance at Njala University and therefore, not included in the interpretation of odds ratios.

Table IX: Odd Ratios and Confidence Intervals new

Variable	Odd ratio	Confidence Interval	
		2.5%	97.5%
Prev_Sch	0.8026482	0.4789552	1.3392472
Fa_Incom	4.0510031	2.1155767	7.8597972
Mo_Incom	0.2708420	0.1556662	0.4663431
Edu_Mo	1.5138640	1.2321852	1.8660740
Residence	1.0598649	0.6672596	1.6878900
Stu_hrs	3.6928341	2.1942258	6.3073000

C. Checking Assumptions for Ordinal Logistic Regression Model

Like every statistical analysis, there are certain assumptions that needed to be met if the result of the fitted ordinal logistic regression model must be useful and generalized. Therefore, to ensure the validity of the fitted model, the following verifications and tests were carried out to make sure that the data did not fail those assumptions.

1. *The dependent variables are ordered:* The first and major assumption of the ordinal logistic regression is that he dependent variable is ordered. The dataset used in this research work satisfied this assumption as the dependent variable, students' academic performance measured using the students' cumulative grade point average (CGPA) (though initially measured on the continuous scale) was categorized to depict the natural categorical ordering that reflects students ranking as excellent, very good, good and poor.

2. *One or more of the independent variables are continuous, categorical or ordinal*: The second assumption is that one or more of the independent variables are either continuous, categorical or ordinal. This assumption was satisfied as there is one continuous independent variable called sturdy hours (Stu_hrs), two categorical independence variables called residence and school type and three ordinal independence variables, mother's income, fathers income and mothers educational level.

3. *No multi-collinearity*: Multicollinearity occurs when the model includes multiple independent variables that are correlated with each other. It is a type of disturbance that may be present in the data. If this disturbance is not eliminated from the data, any statistical inference made about the data may not be reliable. The variance inflation factor (VIF) test is normally used to check for multicollinearity that may exist in the set of independent variables used in an analysis. This research work therefore used the VIF test to test for the presence of multicollinearity in the set of independent variables used in the analysis.

However, the categorical nature of the dependent variable used in ordinal logistic regression analysis makes the validity of the value of the VIF calculated from ordinal logistic regression questionable. Nevertheless, since multicollinearity statistics in regression concern the relationships among the independent variables, ignoring the dependent variable, linear regression with the same list of independent variables can be used to calculate the VIF. However, categorical predictors need to be transform to sets of dummy variables in order to run collinearity analysis in regression.

- *Test for Multicollinearity*

Despite the categorical nature of the dependent variable used for the ordinal logistic regression model, the CGPA used as the dependent variable was numeric as it was originally measured on the continuous scale before it was categorized in order of magnitude (or merit). Therefore, a multi-linear regression was fitted to the data, from which the VIF was directly calculated.

In testing for multicollinearity using the variance inflation factor (VIF), a VIF value greater than 10, implies there is multi-collinearity. In the result presented in Table X, none of the VIF values is greater than 10. Hence multi-collinearity is not a problem here. Therefore, the assumption of no multicollinearity is satisfied.

Table X: Variance Inflation Factor (Vif) Test Output to Check for Multi-Collinearity

Prev_Sch	Mo_Incom	Fa_Incom	Edu_Mo	Residence	Stu_hrs
1.080	3.231	3.327	1.256	1.170	1.12

4) The odds are proportional (proportional odd or Parallel lines assumption):

- *Test for Proportional odds*

To test whether the observed deviations from the fitted ordinal logistic regression model are larger than what could be attributed to chance alone, this research work used the brant test to test **for** both the individual variables and for the whole model. The **r**esult of the brant test presented in Table XI was used to test if the parallel assumption (proportional odd assumption) of the ordinal logistic regression holds in the present model. The assumption is that, the relationship between each pair of outcome groups is the same. The null hypothesis for the brant test is that the parallel assumption holds. This means that the parallel assumption only holds for variables with p-values greater than the chosen significant value of 0.05. The Omnibus variable in the brant test output stands for the whole model. From the output presented in Table XI, the parallel assumption holds for the whole model as the p-value for the Omnibus variable is greater than the chosen significant value of 0.05. Also, the P-value for each of the independent variables are each greater than the chosen significant value of 0.05. This further showed that the parallel assumption holds for each of the significant independent variables used in the ordinal logistic regression model.

Table XI: Brant Test for Parallel Regression Assumption

Test for	X2	Df	Probability
Omnibus	-327.9	12	1
Prev_Sch	2.7	2	0.26
Fa_Incom	5.72	2	0.06
Mo_Incom	5.73	2	0.06
Edu_Mo	5.79	2	0.06
Residence	3.31	2	0.19
Stu_hrs	0.04	2	0.98
H0: Parallel Regression Assumption holds			
Test for	X2	Df	Probability
Omnibus	-327.89997507	12	1.000000e+00
Prev_Sch	2.70082645	2	2.591332e-01
Fa_Incom	5.72164088	2	5.722179e-02
Mo_Incom	5.73164088	2	5.734379e-02
Edu_Mo	5.78505052	2	5.543605e-02
Residence	3.31430048	2	1.906816e-01
Stu_hrs	0.04361476	2	9.784287e-01

V. RESULTS AND DISCUSSION

An ordinal logistic regression analysis was carried out to find out the main factors influencing students' academic performance at Njala University, Njala, Sierra Leone, West Africa. The dependent variable, students' academic performance, was categorized in order of magnitude of performance, ranging from poor performance to excellent performance. The ordinal logistic regression model was significant, as the test of the full model against a model with only the intercepts was statistically significant. This showed that at least one of the regression coefficients in the model is not equal to zero and that the predictors as a set reliably distinguished between students in their various academic performances (chi square = 57.333, $p < .05$ with df=11). The pseudo R^2 values (e.g. Negelkerke.= 0.841=84%) presented in Table VI indicates that the ordinal logistic regression model with its independent variables explained a relatively large proportion of the variation between students in their academic performances at Njala University. This further indicates that a model containing father's income level; mother's income level; mother's educational level and the number of student's study hours is most likely to be a very good predictor of the academic performance for any particular student. In addition, some statistical tests (e.g. brant and VIF tests) showed that all the main assumptions of ordinal logistic regression were satisfied. This showed that the result of the present ordinal logistic regression analysis can be generalized.

Several factors were initially considered as potential determinants of students' academic performance at Njala University. However, the result of the ordinal logistic regression analysis showed that, father's income level (Fa_Incom); mother's income (Mo_Incom) level; mother's educational level (Edu_Mo); and number of study hours (Stu_hrs).were the main factors influencing students' academic performance at Njala University.

The odd ratio for father's income was 4.05 and is positive. This implies that for any one unit increase in the level of father's income that is going from low level of income to high level of income, the odds of moving from , "poor" performance to " good" performance; from "good" to "very good" performance or from "very good" to "excellent" performance (or from the lower and middle categories to the high category) are 4.05 times greater, given that all of the other variables in the model are held constant. Reference [10] also discovered in their research findings that parent's income or social status positively affects the students' tests scores in examinations.

On the other hand, the odd ratio for mother's income was 0.27 and is less than one (<1). This is not surprising as the coefficient for mother's income (in Table VII) is negative. This implies that for a unit increase in the level of mother's income, that is going from low level of income to middle or high level of income, the odds of moving from , "poor" performance to " good" performance; from "good" to "very good" performance or from "very good" to "excellent" performance (or from the lower and middle categories to the high category) are 0.27 times lower, given that all of the other variables in the model are held constant. This is in line with the findings of [20] that, the mothers' employment status has a negative impact except those who are employed in the teaching profession. This implies that income earned by mothers from any other employment apart from teaching or teaching work related, may have a negative influence on students' academic performance as money spent lavishly on children by some mothers may take the place of time spent to help the children to cope with their academic work.

Again, the odd ratio for mother's educational level is **1.5.** *This implies that for students whose mothers attained higher level of education, (secondary or tertiary education) the odds of good, very good or excellent academic performance (i.e., versus poor or very poor academic performance) are* ***1.51*** *times greater than students whose mothers only had primary or no formal education holding all other variables constant. This is in line with the sturdy carried out by [17]. The findings of this study suggested that, education of both father & mother have a positive impact on the academic performance of the students. The study further states that Mother education, however, has a greater impact on the academic outcomes of the students as compared to father's education.*

Above all, the odd ratio for students' sturdy hours is 3.69. This implies that, for every one unit *increase* in student's study hours, the odds of high academic performance(good, *very* good, or *excellent* versus poor or very poor) is 3.69 times greater, holding constant all other variables. In other words, for every one unit *decrease* in student's study hours, the odds of poor academic performance are 3.69 times greater, holding constant all other variables. This is supported by [15] who observed that, students who are very successful in their desired career have longer study time. Reference [11] also discovered in their sturdy that the academic performance of the long study time behavior students was significantly different from that of their short study time counterparts

Finally, the variable residence was also included in the analysis as a possible determinant of students' academic performance. However, the result of the ordinal logistic regression analysis showed that residence is not a significance determinant of students' academic performance (P-value>0.05). This is supported by the findings of [13], who discovered in his paper that there is no significance difference in the academic performance of the students residing on campus and those residing outside the school environment.

VI. CONCLUSION

The objective of the sturdy was to identify the main factors influencing students' academic performance at Njala University. A stratified random sampling method was employed to collect data from 284 students using structured questionnaires. Due to the ordinal nature of the dependent variable, an ordinal logistic regression was used in the analysis. The test result used to check if the final (present) model with explanatory variables included is an improvement over the intercept only model was statistically significant at the 5% significance level. The ordinal logistic regression model used in the analysis satisfied all the necessary assumptions. Several factors were initially considered as potential determinants of students' academic performance. However, the result of the empirical analysis revealed that, number of study hours; father's income level; mother's educational level and mother's income level are the main factors influencing undergraduate students' academic performance at Njala University.; Increase in the number of study hours increases students' academic performance; increase in the father's income level increase students' academic performance; increase in the mother's educational level increase student's academic performance while increase in the mother's income level decreases students' academic performance.

VII. RECOMMENDATIONS

Among the significant factors that may influence students' academic performance, the number of sturdy hours is the only factor used in the ordinal regression analysis that a student may have absolute control over. That is, a student can decide to increase his or her academic performance by simply increasing the number of sturdy hours.

The researchers, therefore, recommend that students should increase the number of study hours so as to increase their academic performances.

Also, mothers, as role models should be encouraged to spend more time with their children to help them cope with their academic works.

REFERENCE

[1] Anderson, G., Benjamin, D., & Fuss, M. (1994). The Determinants. of Success in University Introductory Economics Courses. *The Journal of Economic Education, 25*(2), 99-119. doi:10.2307/1183277

[2] Ahawo, H. (2009). Factors enhancing student academic Achievement in public mixed day Secondary schools in Kisumu East District Kenya. (Unpublished Master thesis). Maseno.

[3] Alias, M., & Zain, A. F. M. (2006). Relationship between Entry Qualifications and Performance in Graduate Education. International Education Journal, 7, 371-378.

[4] Ali, N., Jusoff, K., Ali, S., Mokhtar N., & Salamat A. S. A. (2009). The Factors Influencing Students' Performance at University Technology, MARA Kedah, Malaysia. Management Science and Engineering. ISSN 1913-0341 Vol.3 No.4

[5] C. V Ananth, D. G Kleinbaum (1997) Regression models for ordinal responses: a review of methods and applications**,** *International Journal of Epidemiology*, Volume 26, Issue 6, Pages 1323–1333.

[6] Annie, W., Howard, W.S. & Mildred, M. (1996), "Achievement and Ability Tests - Definition of the Domain",Educational Measurement 2, University Press of America, pp. 2–5, ISBN 978-0-7618-0385-0

[7] *Ananth, C.V.* and *D.G. Kleinbaum*, 1997. *Regression model for ordinal responses*: A review of methods and applications. Int. J. Epidemiol., 26: 1323-1332.

[8] Anderson JA ((1984) Regression and ordered categorical variables. J R Stat Soc B 1984, 46:1-30.

[9] Brant R (1990), Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics 1990, 46:1171-8.

[10] Considine, G. &Zappala, G. (2002). Influence of social and economic disadvantage in the academic performance of school students in Australia. Journal of Sociology, 38, 129-148.

[11] D. E. Ukpong & I. N. George (2013), Length of Study-Time Behaviour and Academic Achievement of Social Studies Education Students in the University of Uyo, International Education Studies; Vol. 6, No. 3; 2013 ISSN 1913-9020 E-ISSN 1913-9039

[12] Gupta, K., &Maksy, M. M. (2014). Factors associated with student performance in an investments course: An empirical study. Journal of Finance and Accountancy. Volume 16

[13] lker Etikan, Kabiru Bala, Ogunjesa Babatope, Meliz Yuvalı, İsmail Bakır (2017) Influence of residential setting on student outcome, Biometrics & Biostatistics International Journal, Volume 6 Issue 4

[14] Jeynes (2002) Parental socio economic status and students' academic performance, *Journal of Personality and Social Psychology*, *4*(3), 586-598.

[15] Kunal, D. S. (2008). Cultivating competence, self-efficacy and intrinsic interest through proximal self-motivation. Journal of Personality and Social Psychology, 4(3), 586-598.

[16] Lee J. (1992) Cumulative logit modelling for ordinal response variables: applications to biomedical research. Comput Appl Biosci. 1992 Dec; 8(6):555-62. doi: 10.1093/bioinformatics/8.6.555. PMID: 1468011.

[17] Mariam Abbas Soharwardi, (2020), impact of parental socioeconomic status on academic performance of students: a case study of bahawalpur, pakistan, Journal of Economics and Economic Education Research, Volume 21, Issue 2,

[18] McCullagh P, Nelder JA (1989), Generalized Linear Models New York: Chapman andHall;

[19] McCullagh, Peter (1980). "Regression Models for Ordinal Data"Journal of the Royal Statistical Society Series, B (Methodological). **42** (2): 109–142. *JSTOR 2984952*..

[20] Mohammad Morshedul Hoque, Sultana Tanjima Khanam & Mohammad NurNobi (2017), The Effects of Mothers' Profession on their Children's Academic Performance: An Econometric Analysis, Global Journal of human-social science: Economics Volume 17 Issue 2 Version 1.0 Year 2017, Online ISSN: 2249-460x & Print ISSN: 0975-587X

[21] Oloo, M .A. (2003). Gender Disparity in student performance in day Secondary Schools, Migori, Unpublished Master of Education Thesis, Maseno University.

[22] Rigney T.J. (2003). A Study on the Relationship between Entry Qualifications and Achievement of Third Level Business Studies Students, The Irish Journal of Management, 117-138.

[23] Stafolani, S. & Brat, M., (2006). Student tme allocaton and educational producton functons,.University of Ancona, Department of Economics Working Paper No.170.

[24] Yusuf, Taofeek Ayotunde; Onifade, C A.; and Bello, O S. (2016) "Impact of Class Size on Learning, Behavioral and General Attitudes of Students in Secondary Schools in Abeokuta, Ogun State Nigeria," *Journal of Research Initiatives*: Vol 2: Iss.1 Article 12.

[25] Zezekwa, N., & Mudavanhu, Y., (2011). The effects of entry qualification on students 'performance in university science courses: The case of Bindura University of Science Education. African Journal of Education and Technology. Vol. 1, No. 3, pp. 32-39.