



Multinomial Logistic Regression Analysis of the Determinants of Students' Academic Performance in Mathematics at Basic Education Certificate Examination

Gaspar Asampana¹, Korah Kassim Nantomah¹, Evans Ayagikwaga Tungosiamu²

¹Department of Statistics, Bolgatanga Polytechnic, Bolgatanga, Ghana

²Zamse Senior High/Technical, Bolgatanga, Ghana

Email address:

gaspil@yahoo.com (G. Asampana), nantomahk@yahoo.com (K. K. Nantomah), evapeace30@yahoo.com (E. A. Tungosiamu)

To cite this article:

Gaspar Asampana, Korah Kassim Nantomah, Evans Ayagikwaga Tungosiamu. Multinomial Logistic Regression Analysis of the Determinants of Students' Academic Performance in Mathematics at Basic Education Certificate Examination. *Higher Education Research*. Vol. 2, No. 1, 2017, pp. 22-26. doi: 10.11648/j.her.20170201.15

Received: November 1, 2016; **Accepted:** December 14, 2016; **Published:** January 16, 2017

Abstract: The focus of the study is to use multinomial logistic regression model to analyze the determinants of students' academic performance in mathematics. A simple random sample of 393 students was selected from a cohort of first year students of Zamse Senior High/Technical in the Bolgatanga Municipality. The students were admitted in the 2015/2016 academic year to pursue various programmes in the school. A questionnaire was used to gather data from the students. The results indicate that the occurrence of good performance in mathematics is largely dependent on sex of students with male students showing significantly good performance than female students. Another significant predictor of good academic performance in mathematics was the age of students; with younger students exhibiting good academic performance than older students. Mother's employment also contributes significantly to good performance in mathematics with students whose mothers are employed showing good academic performance than their counterparts whose mothers are not employed.

Keywords: Basic Education Certificate Examination, Multinomial Logit Model, Academic Performance and Regression

1. Introduction

Mathematics is an integral part of the Ghana's school curricula because of its wide application in commerce, agriculture, health and industry. It is one of the major requirements for gaining admission into educational institutions of higher learning. As a result, some studies have been done on factors that affects students' performance in mathematics [7] [12] [10]. However, the effect of each factor to students' academic performance is rarely measured in Ghana. As such, this study used multinomial logistic regression which is capable of finding the effect of each factor to students' performance in mathematics [11] [6] [2] to predict the determinants of students' academic performance in mathematics at Basic Education Certificate Examination.

2. Related Literature

[1] determined the academic performance of undergraduate students using multinomial logistics regression and their

findings showed that sex of students and mode of admission significantly affect the level of academic performance and that male students are likely to have high performance than female students. Residential status of students was also found to have a significant effect on academic performance of undergraduate students with residential students exhibiting a higher performance compared to non-residential students. [5] Identified factors that affect undergraduate students' academic performance as gender, college of study, high school grade, class attendance, time spent in study, father's education, and parental support and involvement.

[13] identified students' seriousness to study, time spent studying and difficulty students' face in understanding courses as significant determinants of academic success. Students' seriousness about study was found to be the most significant predictor of academic performance. [8] used multinomial logit to determine students' grade in an undergraduate money and banking course. Their results

indicate that student's previous cumulative GPA and their percentile rank in a college entrance examination were major determinants of academic performance. Less significant predictors were student's class attendance and the value students place on the course.

3. Methods and Materials

In this study multinomial logistic regression (MLR) was used to predict the academic performance of students in mathematics at Basic Education Certificate Examination. Multinomial logistic regression model is an extension of binary logistic regression [3] [4] and it is effective where we have polychotomous categorical dependent variable. In an MLR model, the estimates of parameters can be identified and compared to a baseline-category of the dependent variable [6]. The baseline-category logit model with a predictor x is

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \alpha_j + \beta_j x, j = 1, 2, \dots, J - 1. \quad (1)$$

Let P_{ij} = the probability that a student i will have a j^{th} performance

Where $i = 1, 2, \dots, 393$ and $j = 1, 2, 3$. Then, the two multinomial logit equations are as follows:

$$\log\left(\frac{p(y_i=1|x_iSEX+\dots+x_iSTB)}{p(y_i=3|x_iSEX+\dots+x_iSTB)}\right) = \beta_0 + \beta_1 x_iSEX + \dots + \beta_8 x_iSTB \quad (2)$$

$$\log\left(\frac{p(y_i=2|x_iSEX+\dots+x_iSTB)}{p(y_i=3|x_iSEX+\dots+x_iSTB)}\right) = \beta_{10} + \beta_{12} x_iSEX + \dots + \beta_{18} x_iSTB \quad (3)$$

The equations (2) and (3) give the odds ratios of a student obtaining good and average academic performance in mathematics relative to low academic performance and their corresponding probabilities equations (4), (5) and (6) are:

$$P_{i, \text{good academic performance}} = \frac{e^{\beta_1 x_i}}{1 + \sum_{j=1}^3 e^{\beta_j x_i}} \quad (4)$$

$$P_{i, \text{average academic performance}} = \frac{e^{\beta_2 x_i}}{1 + \sum_{j=1}^3 e^{\beta_j x_i}} \quad (5)$$

$$P_{i, \text{low academic performance}} = \frac{e^{\beta_3 x_i}}{1 + \sum_{j=1}^3 e^{\beta_j x_i}} \quad (6)$$

The explanatory variables include: age of student, sex of student, mother's education, mother's employment, parent/guardian religion, parent/guardian residential status and supply of textbooks. A simple random sample of 393 students was selected from a cohort of first year students of Zamse Senior High/Technical School in Bolgatanga Municipality in the Upper East Region. These were students admitted in 2015/2016 academic year.

A questionnaire was used to obtain the data. The questionnaire was structured to avoid students' bias in answering the response variable by specifying the interval within which student's actual performance falls. The data was analyzed using SPSS version 16.0 to calculate the estimated

The model has $J - 1$ logit equations, with separate parameters for each. The effects differ according to the category paired with the baseline [2]. It predicts the effect of the individual independent variables on the dependent variable. When there are J categories of the dependent variable, the model will consist of $J - 1$ logit equations which are fit simultaneously.

In this study, dependent variable (academic performance) is categorized into good performance, average performance, and low performance with low performance serving as a reference category.

Let y denote the level of student's academic performance. Thus:

$$y = 1: \text{High performance}$$

$$y = 2: \text{Average performance}$$

$$y = 3: \text{Low performance}$$

multinomial logit model parameter coefficients.

4. Analysis and Discussion

The study used a multinomial logit model to predict the determinants of students' academic performance in mathematics. The maximum likelihood method was used to calculate the parameter coefficients of the model.

4.1. Model Fitting Information

The overall effectiveness of the model was assessed using the Chi-square statistic. The Chi-square value of 45.954 and its respective p-value of 0.002 is less than 0.05 (see Table 1). This indicates a significant relationship between the dependent variable and the set of independent variables in the final model.

Table 1. Model Fitting Information.

	-2log likelihood	Chi-square	df	p-value
Intercept only	502.857			
Final	456.903	45.954	22	0.002

The Chi-square statistic value could not show the strength of the association between the response variable and predictor variables hence, the Pseudo R-Square measures were employed as shown in Table 2.

Table 2. Pseudo R-Square.

Cox and Snell	0.123
Nagelkerte	0.149
McFadden	0.075

It is indicated in Table 2 that all the three measures (Cox

and Snell, Nagelkerte and McFadden) values indicate weak correlations between the dependent variables and the set of independent variables.

The classification accuracy of the model was assessed and presented in Table 3.

Table 3. Classification.

Observed	Predicted			Percent Correct
	Good performance	Average performance	Low performance	
Good performance	3	65	0	4.4%
Average performance	1	229	2	98.7%
Low performance	1	46	4	7.8%
Overall Percentage	1.4%	96.9%	1.7%	67.2%

The model in general predicted 67.2% of the cases correctly in to their respective memberships. In terms of groups wise, 4.4% of cases were correctly predicted into good academic performance, 98.7% and 7.8% of the cases were correctly predicted into average and low academic performance respectively.

4.2. The Multinomial Logit Models

The study presents two multinomial logit equations with student's low academic performance in mathematics serving as reference category/baseline category. The first multinomial

logit model compares student's good academic performance to low academic performance in mathematics regarding some predictor variables (age of student, sex of student, mother's education, mother's employment, parent/guardian religion, parent/guardian residential status, number of children in the family and supply of textbooks). The second multinomial logit model also compares student's average academic performance to low academic performance in mathematics with the same predictor variables. The first multinomial logit model was presented in Table 4. It presents parameter values obtained from maximum likelihood estimates.

Table 4. Coefficients of multinomial logistic regression-good versus low performance.

Predictor variables	Multinomial Logit model								
	Good performance versus low performance							95% C. I. for odds ratio	
	B	SE	Wald	df	p-value	Odds ratio	Lower Bound	Upper Bound	
Sex									
Male	1.401	0.427	10.746	1	0.001*	4.059	1.756	9.379	
Female ^{RC}	0 ^b			0					
Age									
1-18 years	1.274	0.505	6.352	1	0.012*	3.574	1.327	9.623	
19+ years ^{RC}	0 ^b			0					
Parents/guardians religion									
Christianity	-2.024	0.851	5.657	1	0.017*	0.132	0.025	0.700	
Islamic	-1.711	0.933	3.363	1	0.067	0.181	0.029	1.125	
Traditionalists ^{RC}	0 ^b			0					
Mother's education									
Educated	-0.737	0.428	2.967	1	0.085	0.479	0.207	1.107	
Not educated ^{RC}	0 ^b			0					
Employment of mother									
Employed	0.858	0.415	4.280	1	0.039*	2.260	1.046	5.321	
Not employed ^{RC}	0 ^b			0					
Parents/guardians Residential status									
Self-contained apartment	-0.605	0.410	2.180	1	0.140	0.546	0.244	1.219	
Compound house ^{RC}	0 ^b			0					
Number of Children in household									
1-3 children	0.237	0.603	0.154	1	0.695	1.267	0.388	4.132	
4-6 children	0.124	0.444	0.078	1	0.779	1.133	0.474	2.705	
7+ children ^{RC}	0 ^b			0					
Supply of mathematics textbooks									
Adequate	1.004	0.577	3.034	1	0.082	2.730	0.882	8.451	
Inadequate	0.598	0.468	1.631	1	0.202	1.819	0.726	4.554	
Not all ^{RC}	0 ^b			0					

RC=Reference category, sig. categories at 0.05 probability level

The results showed that male students are more likely to have good academic performance relative to low academic

performance in mathematics than female students with a significant odd ratio of 4.059. In order words, a male student

is approximately 4 times more likely to have a good performance than a female student. It is also observed that, students who are less than or equal to 18 years are more likely to have good academic performance relative to low academic performance in mathematics than those who are older than 18 years with a significant odd ratio of 3.574.

Also, students whose parents/guardians are Christians are less likely to have good academic performance relative to low academic performance in mathematics than students whose parents/guardians are traditionalists with a significant odd ratio of 0.132. In other words, students whose parents/guardians are traditionalists are more likely to have good academic performance relative to low academic performance in mathematics with a significant odd ratio of 7.569 than their counterparts whose parents/guardians are Christians.

Interestingly, the study indicates that students whose parents/guardians are traditionalists are more likely to have good academic performance relative to low academic performance in mathematics than their counterparts whose parents are Muslims with an odd ratio of 5.534, though this is not significant. In other words students with Muslims parents/guardians are less likely to have good academic performance relative to low academic performance than their

counterparts whose parents/guardians are traditionalists with an insignificant odd ratio 0.181.

The study further indicates that students whose mothers are educated are less likely to have good academic performance relative to low academic performance than their counterparts whose mothers are not educated with an insignificant odd ratio of 0.479. Similarly, students whose mothers were employed are more likely to have good academic performance relative to low academic performance than their counterparts whose mothers are not employed with a significant odd ratio of 2.260 as illustrated in Table 4.

The findings indicate that students who were adequately supplied with textbooks have approximately 3 times good academic performance relative to low academic performance than their peers who do not have textbooks at all. Lastly, students with inadequate supply of textbooks were also approximately two times more likely to have good academic performance relative to low academic performance than their counterparts who never had a text book. Both were, however not significant contributors to the model.

The estimates of the second multinomial logit model were presented in Table 5. The parameter coefficients of model were equally obtained from maximum likelihood estimates.

Table 5. Coefficients of multinomial logistic regression-average versus low performance.

Predictor variables	Multinomial Logit model							
	Average performance versus low performance							
	B	SE	Wald	df	p-value	Odds ratio	95% C. I. for odds ratio	
							Lower Bound	Upper Bound
Sex								
Male	1.117	0.348	10.312	1	0.001*	3.054	1.545	6.038
Female ^{RC}	0 ^b			0				
Age								
1-18 years	1.046	0.390	7.198	1	0.007*	2.846	1.326	6.109
19+ years ^{RC}	0 ^b			0				
Parents/guardians religion								
Christianity	-1.164	0.797	2.135	1	0.144	0.312	0.065	1.488
Islamic	-1.117	0.865	1.667	1	0.197	0.327	0.060	1.784
Traditionalists ^{RC}	0 ^b			0				
Mother's education								
Educated	-0.485	0.351	1.914	1	0.167	0.615	0.309	1.224
Not educated ^{RC}	0 ^b			0				
Employment of mother								
Employed	0.192	0.339	0.319	1	0.572	1.211	0.623	2.356
Not employed ^{RC}	0 ^b			0				
Parents/guardians Residential status								
Self-contained apartment	-0.341	0.336	1.029	1	0.310	0.711	0.368	1.374
Compound house ^{RC}	0 ^b			0				
Number of Children in household								
1-3 children	0.726	0.502	2.094	1	0.148	2.067	0.773	5.524
4-6 children	0.567	0.364	2.428	1	0.119	1.763	0.864	3.597
7+ children ^{RC}	0 ^b			0				
Supply of mathematics textbooks								
Adequate	0.185	0.479	0.149	1	0.700	1.203	0.471	3.075
Inadequate	0.141	0.370	0.144	1	0.704	1.151	0.558	2.376
Not at all ^{RC}	0 ^b			0				

RC=Reference category, sig. categories at 0.05 probability level

The logit model indicates that male students are more likely to have average academic performance relative to low academic performance than female students with a

significant odd ratio of 3.054. This means that male students are approximately three times more likely to show average academic performance relative to low academic performance

in mathematics than female students.

The analysis further indicates that students who were 18 years old or younger are more likely to obtain average academic performance relative to low academic performance in mathematics than students above 18 years old with a significant odd ratio of 2.846. Interestingly, students whose parents/guardians are traditionalists are more likely to obtain average academic performance relative to low academic performance than students whose parents were Muslims and Christians with insignificant odd ratios of 3.056 and 3.203 respectively. The model also predicts that students whose mothers were employed are more likely to obtain average academic performance relative to low academic performance in mathematics than students whose mothers were not employed with an insignificant odd ratio of 1.211. This implies that student's mother employment status slightly predicts students' average academic performance.

Students who were supplied with adequate textbooks and inadequate textbooks were more likely to exhibit average academic performance relative to low academic performance in mathematics than students who did not have textbooks at all with insignificant odd ratios of 1.203 and 1.151 respectively.

5. Conclusion

Multinomial logit equations were used to predict the determinants of students' academic performance in mathematics. The results showed that male students have higher chance of exhibiting good academic performance in mathematics than female students. Also, students below 19 years old appeared to be more likely to achieve good and average academic performance in mathematics than students with 19 and above years old. By extension, the younger you are the better your performance in mathematics at the BECE level. Pupils who are able to write their BECE before age 19 have a higher chance of performing well in mathematics than those who do at age 19 or later according to the findings of this paper. In addition, students whose mothers were employed also exhibited good performance in mathematics. As intimated by [5] parental support especially financially has an influence in the performance of the student.

References

- [1] Abiodun, O. O. & Issaiah, F. A. (2015). Academic Performance, Relationship with Gender and mode of Admission. *Journal of Research & Method in Education*, 5 (6): 59-66.
- [2] Agresti, A. (2007). An Introduction to Categorical Data Analysis (2nd ed.). New Jersey: John Wiley & Sons, Inc.
- [3] Aldrich, J. H. & Nelson, F. D. (1984). *Linear probability, logit and probit models*. Newbury Park, CA: Sage publications.
- [4] Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York: Wiley.
- [5] Islam, M. M. (2014). Factors Influencing the Academic performance of Undergraduate Students in Sultan Qaboos University in Oman. *Journal of Emerging Trends in Educational Research and Policy Studies*, 5 (4): 396-404.
- [6] Long, J. S. (1997). *Regression Models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- [7] Nantomah, K. K. & Asampana, G. (2015). Teachers' perception of the causes of students' poor performance in mathematics at the Basic Education Certificate Examination. *Research Journal's Journal of Education*, 3 (9): 1-14.
- [8] Park, K. H. & Kerr, P. M. (1990). Determinants of Academic Performance: A Multinomial Logit Approach. *The Journal of Economic Education*, 21 (2): 101-111.
- [9] Petrucci, C. J. (2009). A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression. *Journal of Social Service Research*, 35 (2): 193-205.
- [10] Tachie, S. A., & Chireshe, R. (2013) High Failure Rate in Mathematics Examinations in Rural Senior Secondary Schools in Mthatha District, Eastern Cape: Learners' Attributions. *Stud Tribes Tribals*, 11 (1): 67-73.
- [11] Tomar, D. & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio-Technology*, 5 (2013), pp. 241-266.
- [12] Tshabalala, T., & Ncube, A. C. (2013). Causes Of Poor Performance Of Ordinary Level Pupils In Mathematics In Rural Secondary Schools In Nkayi District: Learner's Attributions. *Nova Journal of Medical and Biological Sciences*, 1 (1): 4-14.
- [13] Sharker, S. & Rahman, M. D. M. (2015). Determinants of Academic Performance-Multinomial Logistic Regression Approach. *International Journal of Scientific & Engineering Research*, 6 (11): 1212-1216.