American Journal of Theoretical and Applied Statistics
2013; 2(6): 255-267
Published online December 20, 2013 (http://www.sciencepublishinggroup.com/j/ajtas)
doi: 10.11648/j.ajtas.20130206.24

Investigating predictors of examination result data using logistic regression (A case study of Imo State Polytechnic, Umuagwo, Imo State, Nigeria)

Opara Jude[1], Esemokumo Perewarebo Akpos[2, *], Iheagwara Andrew Ihuoma[3], Okenwe Idochi[4], OSUJI GEORGE A.[5]

[1]Department of Statistics, Imo State University, Owerri Nigeria
[2]Federal Polytechnic Ekewe, Yenagoa, Bayelsa State, Nigeria
[3]Department of Statistics, Imo State University, Owerri Nigeria
[4]Department of Statistics, School of Applied Sciences, Rivers State Polytechnic, Bori, Rivers State Nigeria
[5]Department of Statistics, Nnamdi Azikiwe University, PMB 5025, Awka Anambra State Nigeria

Email address:
judend88@yahoo.com (O. Jude), contactperes@yahoo.com (E. P. Akpos), andyiheagwara@yahoo.com (Iheagwara A. I.), nwonda@yahoo.com (O. Idochi), george.osuji99@yahoo.com (OSUJI G. A.)

To cite this article:
Opara Jude, Esemokumo Perewarebo Akpos, Iheagwara Andrew Ihuoma, Okenwe Idochi, OSUJI GEORGE A.. Investigating Predictors of Examination Result Data Using Logistic Regression (A Case Study of Imo State Polytechnic, Umuagwo, Imo State, Nigeria). *American Journal of Theoretical and Applied Statistics.* Vol. 2, No. 6, 2013, pp. 255-267. doi: 10.11648/j.ajtas.20130206.24

Abstract: This study tends to analyze the school examination results (scores) of 300 randomly selected students of Imo State Polytechnic, Umuagwo near Owerri, Imo State, Nigeria who offer English Language and Mathematics as general courses, using the binary logistic regression model with the aim of examining how some factors (variables) in secondary school level contribute to the performance of the students in the Polytechnic. The analysis is performed on the basis of the explanatory variables viz; gender, type of secondary schools, category of secondary schools, board of examinations and location of secondary schools, where scores of students in English Language and Mathematics are assumed to be the response variables. Applying the method of Correspondence Analysis revealed that there exist a significant correlation between board of examinations and location of schools, which made the analysis to be into two stages. The first stage is based on using English Language and Mathematics as a response variable with gender, type of secondary schools, category of secondary schools, and board of examinations as the explanatory variables. The second stage, on the other hand, English Language and Mathematics is the response variable, while gender, type of secondary schools, category of secondary schools, and location of schools are the explanatory variables. The odds ratio analysis compares the scores obtained in two examinations viz English language and Mathematics. The result of the analysis revealed that females are always showing best performances in Mathematics than English examination in all the two stages carried out in this paper. The study also showed that performances of students from girls' schools are found to be the best in English Language course examination than those of students from boys; secondary schools. Furthermore, the study revealed that government schools always show better performance in English course examination than in Mathematics.

Keywords: Odds Ratio, Wald Statistics, Logistic Regression Model, Correspondence Analysis

1. Introduction

Neither two students' nor two schools are identical. Students' differ in gender, culture, religion, language, home environment, financial status of parents etc., whereas schools differ in size of students, quality of teacher, infrastructure, location of the school, aid provided by the government etc. Obviously performance of the students measured in terms of scores or grades obtained by them in examination varies from student to student and school to school. The variability in scores is a function of social climate which has to be studied and analyzed scientifically. The history of analyzing the students' performance is as old as history of education. However formal presentation of analysis started around early thirties of the 20th century.

The performance measure corresponding to different independent variables may be analyzed using logistic regression analysis. Logistic regression has been successfully employed in social science, biostatistics, genetics and demographic issues, but as far as school examination is concerned, not many research articles are available.

This paper deals with presentation and analysis of examination results of Imo State Polytechnic students Umuagwo, Owerri Nigeria, who offer English language and Mathematics as general courses. The data collected consist of scores of the students in both the examinations in English, and Mathematics courses.

It is assumed that scores of students are affected by social environment controlled by the parameters viz:

(i) Gender (male, female)
(ii) Type of schools (boys', girls' and co-educational)
(iii) Category of schools (Government, non-Government)
(iv) Board of examinations i.e. West African School Certificate Examination (WASCE), National Examination Council (NECO) and National Business for Technical and Examination Board (NABTEB) and
(v) Location of schools (urban, rural).

The logistic regression approach has been adopted to study the examination scores under the variables mentioned above.

Scores of students are partitioned into two sets viz (0 - 49) and (50 - 100). Since in the above mentioned examination 50% and above marks indicate second class, hence the students are classified as belonging to two different categories as far as their scores are concerned. As a consequence idea of binary logistic regression analysis seems to be appropriate when scores are functions of independent variables mentioned above.

2. Review of Literature

Iheagwara et al (2013) carried out a research on investigating predictors of incident hypertention using logistic multiple regression. A random sample of 120 patients was selected from Federal Medical Center Owerri, Imo State Nigeria for the study. The data consists of three explanatory variables (Age, gender, BMI) and a response variable (hypertension binary variable). The Logistic regression was explicitly discussed. The SPSS Version 16.0 statistical software package was used for the analysis. From the ROC curve, the area under the curve is 0.903 with 95% confidence interval (0.843, 0.964), which implies that the curve is significantly different from 0.5 since p-value is 0.000 meaning that the logistic regression classifies the group significantly better than by chance. The result of the analysis further revealed that only age is significant, while BMI and gender are insignificant, which implies that among all the explanatory variables used in this paper, only age contributes to whether a patient is hypertensive or not. Thus age has an influence on the hypertensive variable of the patients used in this research work.

Soudarssanane et al (2006) carried out a research titled "A key predictors of high blood pressure and hypertension among adolescents: a simple prescription for prevention". A sample of 673 adolescents (351 males, 322 females) in the 15-19 years age group was used for the study. The univariate analysis followed by logistic regression was used as the statistical techniques. The results of the analysis revealed that mean Systolic Blood Pressure (SBP) and mean Diastolic Blood Pressure (DBP) were 113.6 and 74.3 mm Hg respectively (114.1 & 74.6 in males, 113.1 and 74.1 in females). Mean Blood Pressure (MBP) showed significant correlation with age. MBP and prevalence of hypertension increased with social class, salt intake, and parental history of hypertension, weight, height and BMI. Of these, BMI and higher salt intake emerged as independent predictors by multivariate analysis. Findings were confirmed by the case control study, and the major risk factors for hypertension among adolescents are BMI and higher salt intake.

Jewell (2004) admits that linear regression can be used to model risk difference (also called excess risk). However, he immediately followed this with a discussion of the potential for predictions outside of the 0-1 range and then went on to present logistic regression as the model of choice in general.

Kleinbaum and Klein (2002) make the following argument in favour of the logistic function being a reasonable formula for risk:

The S-shape of $f(z)$ has a nice epidemiological property. If we let $f(z)$ represent risk of disease, and let z represent a combination of risk factors, than an individual's risk is minimal for low z's until some *threshold* is reached. The risk then rises rapidly over a certain range of intermediate z values, and then remains extremely high (around 1) once z gets large enough. This threshold idea is thought by epidemiologists to apply to a variety of disease conditions, and so $f(z)$ is widely applicable for considering the multivariable nature of epidemiologic research questions.

Kleinbaum and Klein's argument provides a level of credibility to the logistic function, since logistic regression predicts risk. That is, logistic regression predicts the proportion of 1's in the outcome variable, while forcing the logistic function shape on the predicted values.

Abdalla (2012) researched on application of Multinomial Logistic Regression model which is one of the important methods for categorical data analysis. This model deals with one nominal/ordinal response variable that has more than two categories, whether nominal or ordinal variable. To identify the model by practical way, the researcher used real data on physical violence against children, from a survey of Youth 2003 which was conducted by Palestinian Central Bureau of Statistics (PCBS). Segment of the population of children in the age group (10-14 years) for residents in Gaza governorate, size of 66,935 had been selected, and the response variable consisted of four categories. Eighteen of the explanatory variables were used for building the primary multinomial logistic regression model. Model had been tested through a set of statistical tests to ensure its appropriateness for the data. Also the model had been tested by selecting randomly of two observations of the data used to predict the position of each observation in any classified group it can be, by knowing the values of the explanatory variables used. It was concluded by using the multinomial logistic regression model that one will be able to define accurately the relationship between the group of explanatory variables and the response variable, identify the effect of each of the variables, and as well can predict the classification of any individual case.

In this paper, we wish to use binary logistic regression to investigate predictors of examination result data in order to determine how some factors (variables) in secondary schools level contribute to the performance of students in tertiary education.

3. Methodology

The logistic regression model assumes that the log-odds of an observation y can be expressed as a linear function of the r input variables x;

$$\ln\left(\frac{P(x)}{1-P(x)}\right)=\sum_{j=0}^{r} b_j x_j \qquad (1)$$

In Equation (1), the constant term b_0 is added by setting $x_0 = 1$. This produces r + 1 parameters. The left hand side of Equation (1) is called the logit of P, thus which implies logistic regression.

Taking the exponent of both sides of Equation (1), we have

$$\frac{P(x)}{1-P(x)}=\exp\left(\sum_{j=0}^{r} b_j x_j\right) \qquad (2)$$

$$=\prod_{j=1}^{r} \exp(b_j x_j) \qquad (3)$$

Equation (3) tells us that logistic models are multiplicative in their inputs (instead of additive, like a linear model), and it produces a way to interpret the coefficients.

The logit equation can also be inverted to get a new expression for P(x).

$$P(x)=\frac{\exp v}{1+\exp v} \qquad (4)$$

where

$$v=\sum_{j=0}^{r} b_j x_j \qquad (5)$$

The right hand side of Equation (4) is the sigmoid of v, which maps the real line to the interval (0, 1), and is approximately linear near the origin. An important fact about P(v) is that the derivative P′(v) = P(v)[1 – P(v)]. The derivation goes as follows:

$$P(v)=\frac{\exp v}{1+\exp v}=(\exp v)(1+\exp v)^{-1} \qquad (6)$$

Solving Equation (6) using the product rule, we have

$$\mathrm{P}'(\mathrm{v})=(1+\exp \mathrm{v})^{-1}(\exp \mathrm{v})-(\exp \mathrm{v})(1+\exp \mathrm{v})^{-2}(\exp \mathrm{v}) \qquad (7)$$

$$=\frac{(1+\exp v)(\exp v)}{(1+\exp v)^2}-\frac{(\exp v)^2}{(1+\exp v)^2} \qquad (8)$$

$$=\frac{\exp v+(\exp v)^2-(\exp v)^2}{(1+\exp v)^2}-\frac{(\exp v)}{(1+\exp v)^2} \qquad (9)$$

From Equation (9), we have

$$P'(v)=\frac{\exp v}{1+\exp v}\cdot\frac{1}{1+\exp v} \qquad (10)$$

$$= \mathrm{P}(\mathrm{v})\,[1-\mathrm{P}(\mathrm{v})] \qquad (11)$$

The solution of a Logistic Regression problem is the set of parameters β that maximizes the likelihood of the data, which is expressed as the product of the predicted probabilities of the N individual observations.

$$L(X\mid P)=\prod_{i=1,y_i=1}^{N} P(x_i)\prod_{i=0,y_i=0}^{N}[1-P(x_i)] \qquad (12)$$

(X, y) is the set of observations; X is a r + 1 by N matrix of inputs, where each column corresponds to an observation and the first row is 1; y is an N-dimensional vector of responses; and (x_i, y_i) are the individual observations.

Taking the log of Equation (12) we get

$$L(X\mid P)=\sum_{i=1,y_i=1}^{N}\log P(x_i)+\sum_{i=0,y_i=0}^{N}\log[1-P(x_i)] \qquad (13)$$

Maximizing the log-likelihood will maximize the likelihood. The quantity -2*log-likelihood is called the deviance of the model. It is analogous to the Residual Sum of Squares (RSS) of a linear model. Ordinary least square minimizes RSS; logistic regression minimizes deviance.

$$Psuedo-R^2=1-\frac{deviance}{null\ deviance} \qquad (14)$$

To maximize the log-likelihood, we take its gradient with respect to β;

$$\nabla_\beta L=\sum_{\substack{i=0\\y_i=1}}^{N}\frac{P_i'}{P_i}x_i-\sum_{\substack{i=0\\y_i=1}}^{N}\frac{P_i'}{1-P_i}x_i \qquad (15)$$

where P_i is shorthand for $P(x_i)$. the maximum occurs where the gradient is zero.

Recall that P′ = P (1 – P), therefore Equation (15) becomes

$$\nabla_\beta L = \sum_{\substack{i=1 \\ y_i=1}}^{N} \frac{P_i(1-P_i)}{P_i} x_i - \sum_{\substack{i=0 \\ y_i=0}}^{N} \frac{P_i(1-P_i)}{1-P_i} x_i \quad (16)$$

$$\nabla_\beta L = \sum_{\substack{i=1 \\ y_i=1}}^{N} (1-P_i)x_i - \sum_{\substack{i=1 \\ y_i=0}}^{N} P_i x_i \quad (17)$$

$$= \sum_{i=1}^{N} [y_i(1-P_i)-(1-y_i)P_i]x_i \quad (18)$$

Equation (18) merges the two cases (y_i = 1 and y_i = 0) into a single sum. This produces the set of simultaneous equations that are true at the optimum;

$$\sum_{i=1}^{N} y_i x_i - P_i x_i = 0 \quad (19)$$

It should be noted that from Equation (19), the sum of probability mass across each coordinate of the x_i vectors is equal to the count of observations with that coordinate value for which the response was true. For instance, supposing the jth input variable is 1 if the subject is hypertensive, 0 if the subject is not hypertensive. Then

$$\sum_{i=1}^{N} y_i x_{ij} - P_i x_{ij} = 0 \quad (20)$$

$$\sum_{i=1}^{N} y_i - P_i = 0 \quad (21)$$

$$\sum_{\substack{i=1 \\ x_{ij=1}}}^{N} y_i = \sum_{\substack{i=1 \\ x_{ij=1}}}^{N} P_i \quad (22)$$

The coefficients β can be solved using the Newton's method.

Supposing we have a vector function f:y = f(β), and we want to obtain the value β_{opt} such that $f(\beta_{opt}) = 0$. Assuming that we start with an initial guess β_0, we can take the Taylor expansion of f around β_0:

$$F(\beta 0 + \Delta) \approx f(\beta 0) + f'(\beta 0)\Delta \quad (23)$$

f is a matrix, and it is the Jacobean of first derivatives of f with respect to β. Setting the left hand side to zero, we then solve for Δ as follows:

$$\Delta 0 = -[f'(\beta 0)]-1f(\beta 0) \quad (24)$$

We then update our estimate for β:

$$\beta 1 = \beta 0 + \Delta 0$$

and iterate until convergence.

In this paper, f is the gradient of the log-likelihood, and its Jacobean is the Hessian (the matrix of second derivative) of the log-likelihood function.

$$H = \frac{\partial \nabla_\beta L}{\partial \beta} = -\sum_{i=1}^{N} x_i \nabla_\beta P_i \quad (26)$$

$$= \sum_{i=1}^{N} x_i P_i (1-P_i) x_i^T \quad (27)$$

$$= XWX^T \quad (28)$$

Where W is a diagonal matrix of the derivatives P'_i and the ith column of X corresponds to the ith observation. So we can solve for Δ at each iteration as

$$\Delta r = (XWrX^T)^{-1}X(y - Pr) \quad (29)$$

Where W is the current matrix of derivatives, y is the vector of observed responses, and P_k is the vector of probabilities as calculated by the current estimate of β.

Comparing this to the solution of a linear regression:

$$y = X^T\beta$$

$$Xy = XX^T\beta$$

$$\beta = (XX^T)^{-1}Xy$$

Comparing the two, we can see that at each iteration, Δ is the solution of a weighted least square problem, where the "response is the difference between the observed response and its current estimated probability of being true.

4. Data Analysis

In this paper, binary logistic regression analysis is performed with dependent variables of examination scores of students in English and Mathematics courses of National Diploma (one) in Imo State Polytechnic, Umuagwo, in presence of explanatory variables; gender, type of secondary schools, board of examinations, category of secondary schools and location of secondary schools.

By applying the method of Correspondence Analysis (CA), it is observed that, there exists a significant correlation between board of examinations and location of schools with r = 0.68. Thus, we perform binary logistic regression analysis.

4.1. Interpretation of Data

Table 1. *Coding of Variables Affecting Imo Polytechnic Results*

Data Variables	Data Explanation	Data Type	Conditioned used
Dependent Variable			
Examination Scores in Imo State Polytechnic	Not-satisfied Satisfied	Binary	0 – Not satisfied 1 – Satisfied

Data Variables	Data Explanation	Data Type	Conditioned used
Independent Variables			
Gender	Gender of student	Binary	0 – Female 1 – Male
School Type	Type of Schools	Categorical	0 – Girls' 1 – Boys' 2–Co-educational
Board	Board of Examinations	Categorical	0 – WASCE 1 – NECO 2 – NABTEB
School Category	Category of Schools	Binary	0 – Government 1–Non-Government
Location	Location of Schools	Binary	0 – Urban 1 – Rural

Table 2. *Categorical Variables Affecting Imo Polytechnic Students' Results*

Variables	Category	Frequencies	%
Gender	Female	186	62.00
	(Male)	114	38.00
School Type	Girls'	55	18.36
	Boys'	79	26.33
	(Co-Education)	166	55.34
Board	WASCE	148	49.33
	NECO	92	30.67
	(NABTEB)	60	20.00
School Category	Government	157	52.33
	(Non-Government)	143	47.67
Location	Urban	138	46.00
	(Rural)	162	54.00

Here the reference category is shown in parenthesis (2nd column in Table 2)

Table 3 (a). *Classification of Total Marks in English Language Course Examination*

Classification	Frequency	Percentage (%)
Not-Satisfied	90	30
Satisfied	210	70
Total	300	100

Table 3 (b). *Classification of Total Marks in Mathematics Course Examination*

Classification	Frequency	Percentage (%)
Not-Satisfied	84	28
Satisfied	216	72
Total	300	100

By dropping one of the variables which is correlated, i.e., board of examinations and location of schools are not used simultaneously as explanatory variables, we carry out binary logistic regression analysis in two stages. It should be recalled that the second column of the following tables "B" represents the coefficient for the constant (also known as the "intercept") in the model.

4.2. Stage I Analysis

In this stage, we consider the examination scores in English course as a dependent variable; where as gender of students, school type, board of examinations, and school category are explanatory variables.

Table 4. *Logistic Regression Analysis of Examination Results in English Course of Students in Imo State Polytechnic, Umuagwo*

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Gender	.138	.253	.298	1	.585	1.148	.700	1.883
	School Type			7.250	2	.027			
	School Type (1)	.771	.356	4.680	1	.031	2.162	1.075	4.349
	School Type (2)	.566	.290	3.800	1	.051	1.762	.997	3.113
	Board			3.257	2	.196			
	Board (1)	.336	.241	1.942	1	.163	1.399	.872	2.245
	Board (2)	.440	.281	2.450	1	.118	1.553	.895	2.696
	School Category	.326	.231	2.001	1	.157	1.386	.882	2.177

[a]. Variable(s) entered on step 1: Gender, School Type, Board, School Category.

The Wald Statistic and the corresponding significance level test, the significance of each of the covariate and dummy explanatory variables in the model are shown in Table 4. It should be noted that if the Wald Statistic is significant (i.e. less than 0.05), then the parameter is significant in the model. Among the explanatory variables, gender of students, board of examinations, and school category are insignificant, whereas type of schools have significantly affected the results of students in English course of Imo State Polytechnic, Umuagwo.

As shown in Table 4, type of schools (i.e. girls' and boys' schools) as a whole is a significant factor with p-value 0.027 corresponding to students' results. Also it is noticed that, performance of girls' and boys' schools are 2.162 and 1.762 times better than the co-educational schools respectively.

The performance of female students is 1.148 times higher than that of the performance of male students, though it is insignificant to the model.

As far as the board of examination is concerned, it is observed that WASCE and NECO boards are 1.399 and 1.553 times better than that NABTEB board. Similarly, it is observed that performances of government schools are 1.386 times better than, the non-Government schools.

Examining the length of confidence interval of estimated odds, we observe that gender (female) of students is estimated with 95% confidence having shortest interval length.

Here, we consider the examination scores in Mathematics course as a dependent variable; whereas gender of students, school type, board of examinations, and school category are explanatory variables.

Table 5. *Logistic Regression of Mathematics Examination Results*

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Gender	.515	.251	4.216	1	.040	1.673	1.024	2.735
	School Type			.378	2	.828			
	School Type (1)	-.186	.317	.346	1	.557	.830	.446	1.544
	School Type (2)	-.084	.278	.092	1	.761	.919	.533	1.584
	Board			7.827	2	.020			
	Board(1)	.470	.236	3.953	1	.047	1.600	1.007	2.543
	Board(2)	.711	.278	6.526	1	.011	2.036	1.180	3.513
	School Category	.084	.223	.143	1	.705	1.088	.702	1.686

[a]. Variable(s) entered on step 1: Gender, School Type, Board, School Category.

Here we consider Mathematics result as the dependent variable, where as gender of students, school type, board of examination, and school category are explanatory variables.

Looking at Table 5, it is observed that gender of students, and board of examination contribute significantly to the model; where type of schools and category of schools are found not to be significant in explaining the satisfactory results in Mathematics.

It is obvious that the performances of female students are 1.673 times higher than that of the performance of male students. On the other hand for type of students, it can be seen that, it is not a significant factor with p-value 0.828 corresponding to examination scores. Hence, it is observed that, performance of girls' and boys' schools are 0.830 and 0.919 times better than the co-educational schools respectively.

As far as board of examinations is concerned, it is observed that, WASCE and NECO boards are 1.600 and 2.036 times better than that of NABTEB board. On the other hand, it is observed that the performances of Government schools are 1.088 times better than the non-Government schools.

Looking at the length of confidence interval of estimated odds, we find that category of school is estimated with 95% confidence having shortest interval length.

Observing the Figure 1, we may conclude that, female performance is better in mathematics examination than that of English Language examination. It is interesting to know that boys' schools also show better performance in English Language examination. Similarly, performance of WASCE shows better results in English Examination than in Mathematics examination; where as, NECO shows better performance in Mathematics examination than in English examination. Government schools show better performance in English than in mathematics examination.

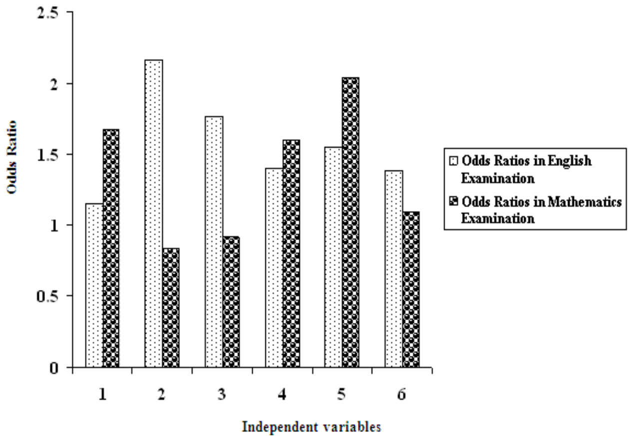

Figure 1. *Comparison of Odds Ratios obtained in English Language and Mathematics Examination*

4.2. Stage II Analysis

In this stage, we consider the examination scores in English course as a dependent variable, where as gender of students, school type, school category and location of schools are explanatory variables.

Table 6. *Logistic Regression Analysis of Examination Results in English Course of Students in Imo State Polytechnic, Umuagwo*

		Variables in the Equation							
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95.0% C.I.for EXP(B)**	
								Lower	**Upper**
Step 1[a]	Gender	.304	.238	1.641	1	.200	1.356	.851	2.159
	School Type			12.798	2	.002			
	School Type (1)	.949	.348	7.449	1	.006	2.582	1.307	5.103
	School Type (2)	.758	.286	7.027	1	.008	2.133	1.218	3.736
	School Category	.517	.218	5.602	1	.018	1.677	1.093	2.572
	Location	-.025	.211	.014	1	.906	.975	.645	1.476

[a]. Variable(s) entered on step 1: Gender, School Type, School Category, Location.

Looking at Table 6, it is observed that school type and school category contribute significantly to the model; where as gender of student and location of schools are found to be insignificant in explaining the satisfactory results in the institution.

Particularly, the performances of female students are 1.356 times higher than that of the performance of male students. On the other hand, for school type (i.e., girls' and boys' schools) as a whole is significant factor with p-value 0.002 corresponding to examination results. Hence, it is observed that, performance of girls' and boys' schools are 2.582 and 2.133 times better than the co-educational schools respectively.

On the other hand, the performances Government schools comparing to the other i.e. non-Government schools are better (with odds ratios 1.677). Similarly, the performances of urban students are 0.975 times better than the rural students.

Looking at the length of confidence interval of estimated odds, we observe that, location of school is estimated with 95% confidence having shortest interval length.

Table 7. *Logistic Regression Analysis of Mathematics Results*

		Variables in the Equation							
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95.0% C.I.for EXP(B)**	
								Lower	**Upper**
Step 1[a]	Gender	.714	.236	9.121	1	.003	2.041	1.285	3.244
	School Type			.400	2	.819			
	School Type(1)	.055	.304	.033	1	.855	1.057	.582	1.919
	School Type(2)	.169	.270	.393	1	.531	1.184	.698	2.009
	School Category	.312	.209	2.223	1	.136	1.366	.906	2.060
	Location	.088	.205	.184	1	.668	1.092	.731	1.632

[a.] Variable(s) entered on step 1: Gender, School Type, School Category, Location.

Here, we consider the examination scores in Mathematics course as a dependent variable; where as gender of students, school type, school category and location of schools are explanatory variables.

Looking at Table 7, gender is the only variable found to be significantly contributing to the model.

The performances of female students are 2.041 times higher than that of the performances of male students; whereas performances of Government schools are 1.366 times better than, the non-Government school. Similarly, the performances of urban students are 1.092 times better than the rural students.

As indicated in Table 7, type of schools (i.e. girls' and boys' schools) as a whole is insignificant factor with p-value 0.819 corresponding to examination scores. Hence, it is observed that, performance of girls and boys' schools are 1.057 and 1.184 times better than the co-educational schools respectively.

Looking at the length of confidence interval of estimated odds, we observe that location of school is estimated with 95% confidence having shortest interval length.

Looking at Figure 2, we may conclude that, female performance is better in mathematics examination than that of English Language examination. It is also interesting to know that boys' schools also show better performance in English Language examination. Government schools show better performance in English than mathematics examination. Similarly, urban schools show slightly better performance in mathematics examination.

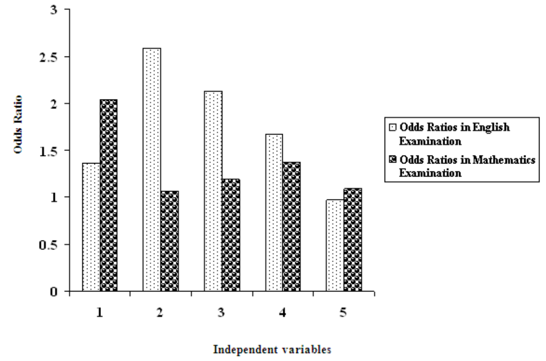

Figure 2. *Comparison of Odds Ratios obtained in English Language and Mathematics Examination*

5. Conclusion

This study attempts to examine how some factors (variables) contribute to the performances of undergraduate students in English and Mathematics courses examination, in Imo State Polytechnic, Umuagwo. From the analysis carried out in this paper for the two stages, we may conclude that, females are always showing best performances in Mathematics examination than English examination in all the two stages. The performances of girls' schools are found to be best in English course examination. Similarly, Government schools always show better performance in English course examination.

Appendix

S/N	English	Maths	Gender	School type	Board	School Category	Location
1	1	0	0	0	2	0	1
2	1	1	1	2	0	1	0
3	1	0	0	1	0	1	0
4	0	0	1	0	2	0	1
5	1	1	1	1	1	1	0
6	0	1	0	2	0	0	1
7	1	1	1	2	0	1	1
8	0	1	0	1	1	0	1
9	1	0	0	0	1	1	1
10	1	1	0	1	0	0	0
11	0	0	0	2	0	0	1
12	1	1	1	2	0	0	0
13	1	1	0	2	0	0	1
14	1	1	0	2	1	1	1
15	1	1	0	0	0	1	0
16	1	0	0	1	1	0	1
17	0	0	0	2	0	1	1
18	1	1	1	1	0	1	0
19	1	1	1	2	0	0	0
20	0	0	0	1	1	0	1
21	1	0	0	2	1	0	1
22	1	1	0	0	1	1	0
23	1	1	0	1	0	0	1
24	0	0	1	2	2	0	0
25	1	1	1	2	0	1	0
26	1	1	0	0	0	0	0
27	1	0	1	1	0	0	0
28	0	0	0	2	0	0	1
29	1	1	0	0	1	0	1
30	0	0	0	2	0	1	0
31	1	1	0	1	0	0	0
32	0	0	0	2	0	1	0
33	1	1	0	0	2	1	0
34	0	0	1	2	2	0	1
35	1	1	1	0	0	0	0
36	1	1	1	2	0	0	1
37	1	1	0	1	2	1	1
38	0	0	1	0	1	1	1
39	1	1	0	1	0	0	1
40	1	1	0	2	1	0	1
41	1	1	1	1	0	1	0

S/N	English	Maths	Gender	School type	Board	School Category	Location
42	0	1	1	2	2	1	0
43	1	1	1	2	0	0	0
44	1	1	0	2	1	1	1
45	1	0	0	2	1	1	1
46	0	0	0	1	0	0	1
47	0	1	0	1	0	0	0
48	1	1	0	2	2	1	0
49	1	1	0	0	1	0	1
50	1	0	0	2	0	1	0
51	0	0	1	0	2	0	0
52	1	1	1	0	0	0	1
53	1	1	1	1	2	1	1
54	1	1	0	2	0	1	1
55	0	0	1	0	1	0	1
56	1	1	1	2	0	1	1
57	1	0	0	1	1	1	1
58	1	1	0	2	0	0	0
59	0	1	0	1	0	1	0
60	1	1	0	2	0	0	0
61	0	0	0	2	2	1	1
62	1	1	0	1	2	0	1
63	1	1	0	2	1	1	1
64	1	1	0	0	1	0	1
65	0	1	0	1	0	0	0
66	1	0	0	2	0	1	0
67	1	1	1	1	1	0	1
68	0	0	1	2	0	1	1
69	1	1	1	2	0	1	0
70	1	1	1	1	2	0	1
71	0	0	1	1	0	1	0
72	1	1	1	1	2	0	0
73	1	1	1	0	0	1	0
74	0	0	0	0	0	0	1
75	0	0	0	0	1	1	1
76	1	0	0	2	2	1	1
77	0	0	0	2	1	0	1
78	1	1	0	2	0	0	0
79	1	1	1	1	0	1	1
80	1	1	1	2	1	0	0
81	1	1	1	2	1	1	1
82	1	1	1	1	0	1	1
83	0	1	0	0	0	0	1
84	1	1	1	0	0	0	0
85	0	0	0	0	1	1	1
86	1	0	0	0	0	1	1
87	1	1	0	2	0	0	1
88	1	1	0	1	2	0	0
89	0	1	0	2	1	0	0
90	1	1	1	2	0	0	1
91	1	1	1	0	0	0	1
92	1	0	1	1	2	0	0
93	1	1	0	2	1	0	1
94	1	1	1	0	0	1	0
95	0	1	0	2	2	0	0
96	1	1	1	2	1	0	1
97	0	1	1	1	0	0	0
98	1	1	0	0	2	1	1
99	1	1	0	2	1	0	1
100	0	1	0	1	2	0	1
101	1	0	0	1	0	1	0
102	1	1	0	2	2	0	1
103	1	0	0	0	0	1	1
104	1	1	1	1	0	1	1
105	0	1	1	2	1	0	0

S/N	English	Maths	Gender	School type	Board	School Category	Location
106	1	1	0	2	2	1	1
107	1	1	1	2	0	0	0
108	1	1	1	1	1	1	1
109	1	1	0	2	1	0	1
110	0	1	1	0	0	1	0
111	1	0	1	1	1	0	1
112	1	1	1	2	0	1	0
113	1	1	0	1	1	0	1
114	1	1	1	2	1	0	1
115	1	1	0	1	2	1	0
116	0	1	1	2	1	1	1
117	1	1	0	2	2	0	1
118	1	1	0	2	2	0	1
119	1	1	0	0	0	1	0
120	0	0	0	1	0	0	0
121	1	1	0	2	1	1	1
122	1	1	0	1	1	0	1
123	1	1	0	2	1	1	1
124	1	1	1	1	1	0	1
125	0	0	1	2	2	1	0
126	0	0	1	1	0	0	0
127	1	1	0	2	0	1	0
128	0	0	1	1	2	0	1
129	1	1	0	0	0	1	0
130	0	0	0	1	2	0	1
131	1	0	0	0	2	0	1
132	1	1	0	0	2	1	1
133	0	0	0	2	0	0	0
134	1	1	0	1	0	0	0
135	0	0	0	2	2	1	1
136	1	1	1	1	1	0	1
137	1	1	1	2	0	0	0
138	1	1	1	1	0	0	0
139	0	0	1	2	0	1	0
140	1	1	0	1	0	0	1
141	1	1	1	0	0	0	0
142	1	0	0	0	1	0	1
143	0	1	0	2	0	0	0
144	1	1	0	2	1	1	1
145	0	1	0	1	1	0	1
146	1	1	0	2	1	1	1
147	1	1	1	1	1	0	1
148	0	1	1	2	1	1	1
149	1	1	0	1	0	0	0
150	1	1	1	0	0	0	0
151	1	1	1	1	2	1	1
152	0	1	1	0	0	0	0
153	1	1	0	2	2	0	0
154	0	1	0	1	1	1	1
155	1	1	0	2	1	1	1
156	1	1	1	1	1	0	0
157	1	0	1	2	1	0	0
158	1	1	0	2	0	0	0
159	0	1	1	1	1	0	1
160	0	1	0	2	2	1	1
161	1	1	1	2	1	0	1
162	1	1	0	0	0	1	0
163	0	1	0	2	1	1	1
164	1	1	0	0	1	0	1
165	1	1	1	0	2	0	1
166	1	1	1	1	1	1	1
167	1	0	1	1	0	1	1
168	0	1	0	2	1	0	1

S/N	English	Maths	Gender	School type	Board	School Category	Location
169	1	1	0	2	1	1	1
170	0	0	0	1	2	0	0
171	1	1	1	1	1	1	1
172	1	1	1	2	1	0	1
173	1	1	0	1	0	0	0
174	1	1	0	2	2	1	1
175	1	1	0	2	0	1	0
176	1	1	0	1	1	0	1
177	0	0	0	2	0	0	1
178	1	1	0	1	2	1	0
179	1	1	1	2	1	0	1
180	1	1	0	1	2	1	1
181	1	1	1	0	1	0	0
182	1	1	1	2	0	0	0
183	0	0	1	1	1	1	1
184	1	0	0	2	0	0	0
185	1	1	0	1	1	0	1
186	0	1	0	2	1	0	1
187	1	1	1	2	0	1	0
188	1	1	1	1	0	0	0
189	0	1	0	2	2	1	1
190	1	1	0	1	0	0	0
191	0	1	1	2	1	0	1
192	1	1	1	1	1	0	0
193	1	1	0	2	1	1	1
194	1	1	0	0	1	0	0
195	0	1	0	1	1	0	1
196	1	1	0	2	1	1	1
197	1	1	0	0	0	0	0
198	1	1	0	1	2	1	0
199	1	0	1	2	0	0	0
200	0	1	1	1	1	0	1
201	1	1	0	2	0	1	0
202	1	1	0	1	2	0	1
203	1	1	1	2	1	1	0
204	1	1	0	1	1	0	1
205	1	1	0	2	0	1	0
206	0	0	0	2	2	1	1
207	1	1	0	1	1	0	0
208	1	1	0	2	0	1	1
209	0	1	0	0	0	0	0
210	1	1	1	2	1	0	1
211	1	1	1	1	1	0	0
212	0	1	1	2	0	1	0
213	1	1	0	1	2	0	1
214	1	1	1	2	0	1	1
215	1	0	1	1	0	0	0
216	1	1	0	2	1	0	0
217	1	1	0	1	0	0	1
218	0	1	0	2	0	1	0
219	1	1	0	1	1	0	1
220	1	1	1	2	1	0	1
221	1	1	1	2	2	1	0
222	0	1	0	2	0	1	0
223	1	1	0	2	0	0	1
224	1	1	0	0	1	1	1
225	1	1	0	2	0	1	0
226	0	1	1	2	0	0	0
227	1	1	0	2	2	1	1
228	1	1	0	2	0	1	1
229	1	0	0	1	1	0	0
230	0	1	0	2	1	1	1
231	1	1	0	2	0	1	1
232	1	1	1	2	0	0	0
233	1	1	1	2	0	1	0
234	0	0	0	1	1	1	1
235	1	1	1	0	0	0	0

S/N	English	Maths	Gender	School type	Board	School Category	Location
236	1	1	0	2	2	1	0
237	0	0	0	2	1	0	1
238	0	0	0	2	0	1	1
239	1	1	0	2	1	0	0
240	0	0	0	2	0	1	1
241	1	1	0	2	2	0	1
242	1	1	0	2	0	1	0
243	0	0	1	2	1	1	1
244	1	0	0	2	2	0	1
245	1	1	1	2	0	0	1
246	0	0	1	2	0	0	0
247	1	1	0	2	0	0	0
248	1	1	0	0	1	0	1
249	1	1	1	2	2	0	0
250	0	1	1	2	0	1	1
251	0	1	1	2	2	1	0
252	0	1	0	2	0	1	1
253	1	1	0	2	0	1	0
254	0	1	0	2	0	0	1
255	1	1	0	2	0	0	0
256	1	0	0	2	2	1	0
257	0	1	0	2	2	1	1
258	1	1	0	0	0	1	0
259	1	1	1	0	2	1	0
260	1	1	0	2	2	0	1
261	0	0	1	2	0	1	0
262	1	1	1	2	2	0	1
263	1	1	1	2	0	1	0
264	1	1	1	2	0	1	0
265	0	0	0	2	0	0	1
266	1	1	0	2	2	1	0
267	1	1	0	2	0	0	0
268	1	0	0	0	0	1	1
269	0	0	0	0	0	1	0
270	1	1	0	0	0	0	0
271	1	1	0	2	1	1	1
272	0	0	0	2	1	1	0
273	1	1	0	2	1	1	0
274	1	1	0	2	0	1	0
275	1	1	1	2	2	0	1
276	0	1	0	2	0	1	0
277	1	1	1	2	0	1	0
278	1	1	0	2	0	1	1
279	1	1	0	0	0	1	0
280	0	1	1	2	0	0	1
281	1	0	1	2	0	1	0
282	0	0	0	2	0	1	1
283	1	1	0	2	0	1	0
284	0	0	0	2	0	0	1
285	1	1	0	2	0	1	0
286	1	1	0	0	0	1	0
287	0	0	1	2	0	0	1
288	1	1	0	2	0	0	0
289	1	1	1	0	0	1	0
290	0	0	0	2	0	1	0
291	1	1	1	2	2	1	0
292	0	0	1	2	0	1	1
293	1	0	0	2	0	0	0
294	1	1	0	2	0	1	1
295	0	1	1	2	0	0	1
296	1	1	0	2	2	1	1
297	1	1	0	0	0	1	0
298	0	0	0	2	0	0	0
299	1	1	0	2	0	1	0
300	1	1	0	0	0	1	0

References

[1] Abdalla M.E. (2012): An Application on Multinomial Logistic Regression Model. Pok.j.stat.oper.res. Vol. VIII No. 2 2012 pp 271-291.

[2] Iheagwara, A.I., Nwobi, A.C, Esemokumo, P.A, and Opara, J. (2013): Investigating Predictors of Incident Hypertension Using Logistic Multiple Regression. International Journal of Applied Mathematics and Modeling, IJA2M@KINDI PUBLICATIONS. Vol.1, No.4, 1-17. October, 2013. ISSN: 2336-0054.

[3] Jewell NP. (2004). *Statistics for Epidemiology*. New York, Chapman & Hall/CRC.

[4] Kleinbaum DG, Klein M. (2002). *Logistic Regression: A Self-Learning Text*. 2nd Ed. New York, Springer-Verlag.

[5] Sarma R and Sarmah, P. (1999): Analysis of Results Based on Grades, in Proceedings of the Second International Conference on Operations and Quantitative Management in the Global Business Environment (ICOQM), Ahmadabad, India, 3 – 6th January 1999, p.282 – 290.

[6] Saha, G. and Sarmah, P. (2010): Statistical Analysis of School Examination Result with Special Reference to the State of Tripura: North-East India, Journal of Statistics.

[7] Soudarssanane M.B., Karthigeyan S., Stephen A. and Sahai, A. (2006): Key Predictors of High Blood Pressure and Hypertension among Adolescents: A Simple Prescription for Prevention. Indian Journal of Community Medicine Vol. 31, No. 3, July – September, 2006.