Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.11, No.1, 2021

MODELLING FACTORS AFFECTING AVERAGE NUMBER OF STUDENTS CARRYOVER: A POISSON REGRESSION APPROACH

[1]Umezurike Chinaegbomkpa*and [2] Agamba Chinedu K.
[1, 2]Department of Statistics, Michael Okpara University of Agriculture Umudike, Abia State, Nigeria.
E-mail: c.umezurike@mouau.edu.ng

Abstract

This research work investigated a Poisson regression model that fit the average number of carryover a student will have and factors affecting average number of students carryover such as cumulative grade point average (CGPA), Sex, Age, Marital status (Single and Married), Residence (Inside the campus and outside the campus), Choice of course of study, Jamb Mathematics score and Relationships (dating or not dating). Data on average number of carryover for students used in this work were collected from Department of Statistics, Michael Okpara University of Agriculture, Umudike while data on other factors were collected primarily by use of questionnaires. Poisson log-linear regression model was used to fit the data collected. It was discovered that students living inside campus have lesser number of carryover than students living off campus and also it was observed that students who are married have greater number of carryover than those who are not. The number of student's CGPA inside campus is greater than those students who live off campus. From the analysis obtained, we found out that the factors; CGPA, Marital status and Residence were the major contributors to the number of carryover for the students of statistics Department Michael Okpara University of Agriculture, Umudike.

Keywords: Poisson Regression, cumulative grade point average (CGPA), Age, Sex, Marital status, residence and relationships.

Introduction

Educating efficient and effective human forces is considered among the main duties of Universities. Every year, universities admit newcomer students and graduate some others; in this continuous cycle, education quality has a crucial position. Thus, increasing the quality of educational system is considered the most influential factor in developing the countries; this is because students achieve a position as a result of their academic success in which their maximum internal and external forces are used for achieving goals of higher education and obtaining necessary conditions for successful social life. On the other hand, lack of success in education paves the ground for several personal and social problems and deviation from achieving the goals of educational system. In this regard, one of the major problems of students in higher education centers is the word "carryover". Carry-over is similar to failure which not only leads to the waste of current expenditure and time but also generates mental-psychological, social and family problems for the university students. Carryover as we call it in Nigerian universities and polytechnics is the act of retaking a failed course in the next or another subsequent semester. According to studies, this problem is increasing every year so that many students cannot handle the curriculum (academic courses) or complete it in due time. Some woks exist in literature which has used Poisson regression to investigate social factors. Vijar *et al* (2016) in their research work, used Poisson regression method to solve the problem of factors affecting Children Ever born (CEB) in Botswana women of child bearing age (15-49 years) from the BFHS survey (BFHS,2007) as reported on Statistics Botswana website. Some basic socio-economic and demographic information of mother (respondent) were fitted into model. Their findings indicate that place of residence, age, education, marital status, work status, access to media, age at first child birth and use of condom by either partner have significant effect on fertility level in Botswana.

Felix *et al* (2004) in their finding, a simple random sample of drivers aged sixty-five years or older was selected from the Alabama Department of Public Safety Records. The data in the sample has information on many

variables including the number of accidents, demographic information, driving habits, and medication. They also made use of Poisson regression model to generalize application to accident data. Poisson regression (GPR) model is considered for identifying the relationship between the number of accidents and some covariates. In their research work about 59% of driver's quality was rated as average or below who are involved in automobile accidents and the drivers who take calcium channel blockers show a significantly reduced risk of about 34.5%.

Ferenc and Rita (2014) in their work applied Poisson regression model in the Sociology Study of Suicide. They explain how Poisson regression can be used to study dependent variable describes the number of occurrences of some rare event such as suicide. They were able to point out why ordinary linear regression is inappropriate for treating dependent variables of this sort, the reason they gave was that the dependent variable is a count data and they present the basic Poisson regression model and show how it fits in the broad class of generalized linear models.

Jonas *et al* (2005) considers changes in climate and their effects in the Public health. In their research work, they make used of Poisson regression to analyze the daily number of hospitalizations in São Paulo City, Brazil, in the period of January 01, 2002 to December 31, 2005. This data set relates to pneumonia, coronary ischemic diseases, diabetes and chronic diseases in different age categories. In order to verify the effect of climate changes the following covariates are considered such as atmosphere pressure, air humidity, temperature, year season and also a covariate related to the week day when the hospitalization occurred.

Boateng (2012) in his work used Poisson and negative regression model, in modeling the occurrence of Malaria cases (a case study of Obuasi Government hospital 2007 to 2010), to model the incidence of severe malaria cases given age , gender and time in years and lastly to validate the two models using negative binomial regression model. Both models indicated that malaria is independent of gender.

Nakaya *et al* (2004) researched on the topic geographically weighted Poisson regression for disease association mapping. In their work reveal that geographically weighted Poisson regression (GWPR) and its semi-parametric variant as a new statistical tool for analyzing disease maps arising from spatially non-stationary processes. From their findings, it indicates that there are significant spatial variation (that is variation beyond that expected from random sampling) in the relationships between working-age, mortality and occupation segregation and working-age, mortality and unemployment through the Tokyo metropolitan area and that, consequently the application of traditional 'global' models would yield misleading results.

George *et al* (2016) they work make use of Logit Modeling, to work on some factors affecting the Academic Performance of Graduating Science Students in University Of Uyo, Nigeria. The data showed the graduating cumulative grade point average (CGPA) of 2011/2012 graduates, together with some factors affecting their academic performance which includes, gender, Mode of entry which is through jamb or remedial programme, type of secondary school attended whether it is government or private and the student's Department. The results of the analysis showed that female students, students who were admitted into university through the Joint Admission and Matriculation Examinations Board (JAMB) examinations and students that attended government secondary school performed better academically. The student's department has a negative effect on the academic performance of students. They concluded that there is need for the total over haul of the entire Nigeria education system at all level to improve the overall performance of the system.

Lisa (2009), Philip and Sebastian (2015) and Ataharul and Rafiqul (2017) are popular literatures in this research area with stunning results.

Adeleke *et al* (2010) in their study explores the distribution of the performance of graduate distance learners in the University of Lagos. The graduating Cumulative Grade Point Average (CGPA) of five hundred and sixty-five students and their final Grade Point Average (GPA) are used for the study.

Aromolaram *et al* (2013) conducted a survey to determine the socio-economic factors influencing student academic performance in Yaba College of Technology, Yaba, Lagos. The students' academic performance was measured using variable CGPA categorized into two poor (CGPA between 0 and 2.49) and good (CGPA between 2.50 and 4.00). Four factors; mothers' education level, living togetherness of parents, student class and weekly income/allowance; are found to influence students' academic performance.

It is apparent that there is an astronomical decline in students' academic achievement in Nigeria Universities. According to Igbo (2007), lack of success can lead to inappropriate behavior and frustration on the part of students. Statistical record has it that every year a lot of students fail out of college. In this work, we want to consider and find out the causes of carryover among student in higher institutions and at the same time to determine the relationships if any that exist between the various causes of student carryover. The study sought to analyze the factors affecting the average number of carryover of students using Poisson regression.

Methodology

The data used in this work is primary and secondary data. The primary data were collected through questionnaire and the secondary data were collected from Statistics Department, Michael Okpara University of Agriculture, Umudike. Poisson regression modeling is adopted. Y the response variable is the number of carryovers while the independent variables are represented thus.

X1	CGPA
X_2	Sex (male or female)
X_3	Age
X_4	Marital status (single or married)
X_5	Residence (inside campus or outside campus)
X_6	Jamb mathematics score
X_7	Choice of study
X_8	Relationship (dating or not dating)

Poisson regression model

$$\Pr[Y = y] = \frac{e^{-\mu}\mu^{y}}{y!} \quad \mu > 0, y = 0,1,2,... \qquad 1$$

where the log of the mean μ is assumed to be a linear function of the independent variables. That is,

$\ln(\mu) = X'\beta$ and $X' = \{1, X_1, X_2, ...X_8\}$

$$\ln(\mu) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_8 X_8 = X'\beta \qquad 2$$

The exp $\{\beta_i\}$ is also known as Incidence Rate Ratio (IRR) which quantifies the direction and strength of relationship between predictors and dependent variable (carry over).

Goodness of fit:

Overall performance of the fitted model can be measured by two different chi-square tests. There is the Pearson Statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - \exp\left\{X_i\hat{\beta}\right\}\right)^2}{\exp\left\{X_i\hat{\beta}\right\}} \quad 3$$

And the deviance Stastistic.

$$D = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\exp\left\{X_i\hat{\beta}\right\}} \right) - \left(y_i - \exp\left\{X_i\hat{\beta}\right\}\right) \right] \quad 4$$

Both of these statistics are approximately chi-square distributed with *n* - *k* degrees of freedom. When a test is rejected, there is a statistically significant lack of fit. Otherwise, there is no evidence of lack of fit. The high p-values indicate no evidence of lack of fit.

The Akalike information criteria (AIC), a goodness of fit measure defined as (-2ln L+2k) where k is the number of parameters in the model and L is the likelihood function of the final model. Bayesian information criteria (BIC) a goodness of fit measure defined as

$$AIC = -2\ln L + 2k \quad 5$$

$$BIC = \frac{-2\ln L + K\ln(n)}{n} \quad 6$$

The variance inflation factor (VIF) identifies correlation between independent variables and the strength of that correlation. The variance inflation factor for the j[th] explanatory variable is defined $VIF_j = \frac{1}{1-R_j^2}$ 7

In practice, a VIF >5 or 10 indicates that the associated regression coefficients are poorly estimated because of multcollinarity.

Result and Discussion

Minitab 18 was explicitly used for all analyses in this study and statistical tests conducted were done at 5% level of significance.

Table 1: Deviance Table

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	8	93.253	11.6567	93.25	0.000
X_1	1	42.666	42.6658	42.67	0.000
X_2	1	0.864	0.8639	0.86	0.353
X_3	1	0.134	0.1344	0.13	0.714
X_4	1	6.218	6.2176	6.22	0.013
X_5	1	5.499	5.4990	5.50	0.019
X_6	1	0.071	0.0714	0.07	0.789
X_7	1	1.328	1.3281	1.33	0.249
X_8	1	0.365	0.3649	0.36	0.546
Error	111	161.345	1.4536		
Total	119	254.599			

From Table 1, it could be observed that three predictors variables (X_1 = CGPA, X_4 = Marital status, X_5 = Residence) are statistically significant at the 0.05 level. We can conclude that changes in these variables are associated with changes in the response variable.

Table 2: Coefficients

Term	Coefficient	SE Coefficient	VIF
Constant	2.015	0.950	
X_1	-0.813	0.126	1.63
X_2	0.159	0.173	1.08
X_3	-0.0088	0.0241	1.13
X_4	1.066	0.509	1.01
X_5	0.432	0.190	1.09
X_6	-0.00202	0.00759	1.72
X_7	-0.171	0.148	1.07
X_8	-0.092	0.152	1.10

The general model is defined as;

$$\hat{\mu} = e^{\alpha+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+...+\beta_8 X_8}$$

From the analysis, we obtained;

Y = exp(Y')

Y = 2.015 - 0.813 X_1+ 0.159 X_2 - 0.0088 X_3 + 1.066 X_4 + 0.432 X_5 - 0.00202 X_6 - 0.171 X_7- 0.092 X_8

where α= 2.015

$\beta_i X_i$= the predictor variables with their respective coefficients.

From table 2, given the students CGPA, having the same sex, having the same age, having the same choice of study, living in the same hostel, having the jamb mathematics score and marital status, the average number of carryover is estimated to be $e^{-0.813}$ = 0.4435 times.

Given male and female who are of the same age, living in the same hostel, having same choice of study, having same jamb mathematics score, same CGPA and Marital status, the average number of carry over a male student will have is estimated to be $e^{0.159}$ = 1.1723 times of that of female student.

Given the age of the students who are of the same sex, living in the same hostel, having same choice of study, having same jamb mathematics score, same CGPA and Marital status, the average number of carryover is estimated to be $e^{-0.0088}$ = 0.9912 times.

Given the students who are single, of the same sex, living in the same hostel, having same choice of study, same jamb mathematics score and same CGPA, the average number of carryover is estimated to be $e^{1.066}$ = 2.9037 times of those who are married.

Given the students living in the hostel and off campus who are of the same sex, having the same age, having the same choice of study, having same jamb mathematics score, same CGPA and Marital status, the average number of carryover is estimated to be $e^{0.432}$ = 1.5403 times of that of those living off campus.

Given the students jamb mathematics score, who are of the same sex, having the same age, living in the same hostel, having the same choice of study, same CGPA and Marital status, the average number of carryover is estimated to be $e^{-0.00202}$ = 0.9980 times.

Given students who apply statistics, having the same sex, of same age, living in the same hostel, having the same jamb mathematics score, same CGPA and Marital status, the average number of carryover is estimated to be $e^{-0.171}$ = 0.8428 times of those who do not apply statistics.

Given male and female who are in relationship, of the same age, living in the same hostel, having same choice of study, having same jamb mathematics score, same CGPA and Marital status, the average number of carryover a student that is not in a relationship (not dating) will have its estimate to be $e^{-0.092}$ = 0.9121 times of that of students that are in a relationship (dating).

From the analysis obtained, statistically, the variables X_1 = CGPA, X_4 = Marital status, X_5 = Residence (inside campus or outside campus) are significant at 0.05 level of significance while X_2 =Sex, X_3 = Age, X_6 = Jamb mathematics score, X_7 = Choice of study and X_8 = Relationship (dating or not dating) are not statistically significant at 0.05 level of significance.

The VIF from table 2 indicate that there is no multicollinearity between the independent variables since they have values that are less than five(5).

Table 3: Model Summary

Deviance R-Sq	Deviance R-Sq (adj)	AIC
36.63%	33.49%	375.63

From table 3, it shows that 36.63% of the variations are explained by the model.

Table 4: Goodness-of-Fit Tests

Test	DF	Estimate	Mean	Chi-Square	P-Value
Deviance	111	161.34547	1.45356	161.35	0.001
Pearson	111	134.05574	1.20771	134.06	0.067

From table 4, the Goodness of fit test for p-value of the Deviance test 0.001 is less than the usual significance level of 0.05, is sufficient evidence to conclude that the predicted factors affecting the average number of carryover of students do not deviate from the observed average number of carryover. While the p-value for the Pearson test is 0.067 which is higher than the usual significance level of 0.05, is insufficient evidence to conclude that the predicted factors affecting the average number of carryover of students deviate from the observed average number of carryover.

Since the Deviance is strong enough to reject the null hypothesis at significance level of 0.05. We conclude that the model is correctly specified.

Conclusion

This research work investigated a model that fit the factors affecting average number of students carryover. We consider some selected factors such as CGPA, Sex, Age, Marital status, Residence (inside campus or outside campus), Jamb Mathematics score, Choice of study and Relationship (dating or not dating).

In all the factors considered we found out that only three factors which are CGPA, Marital status and Residence (inside campus or outside campus) have an effect on the student's number of carryover in higher institutions. This is because they are significant. The following findings were made.

i. The average number of carryover of students living inside the school is less than those living outside the school with the estimate of 2.

ii. The average number of carryover students who are married is greater than those who are not married with the estimate of 3.

iii. The average number of carryover student with lower CGPA inside the school is greater than those living outside the school.

Similarly, from the analysis obtained in chapter four, it was discovered also that;

iv. The average number of carryover male students is higher than that of the female students. i.e. the estimated average number of carryover for male is 1.

v. The average number of carryover male students between ages 17-35 is higher than that of the female students. i.e. the estimated average number of carryover for male is 1.

vi. The average number of carryover male students with Jamb mathematics score is higher than that of the female students. i.e. the estimated average number of carryover for male is 1.

vii. The average number of carryover students who choose statistics as their course choice of study is less than that of those students that did not choose statistics as their course choice of study.

viii. The average number of carryover students who are in relationship (dating) is higher than that of those students who are not in relationship (not dating). i.e. the estimated number of carryover students is 1.

From the foregoing, the following recommendations were made.

i. The school management of the universities in Nigeria are advised to call the attention of the government so that more conducive hostels will be built and ensure that the students occupy the hostels rather than living outside campus.

ii. The students both male and female are advised not to marry while they are still in the university.

iii. The university admission board should ensure that students are admitted into the institution based on the course they applied not by any means changing their choice of course in order for it not to affect their results and CGPA in the long run.

REFERENCES

Adeleke, I. A, Ogundeji R.K, and Edesiri, Aham (2010). Analysis of Graduating CGPA of Distance Learning Students

Ahani, E. Abass .O and Okafor Ray O. (2010).Application of Logistic Regression Model to Graduating CGPA of University Graduates.*Journal of Modern Mathematics and Statistics Vol. 4 Issue: 2 Pp. 58-62.*

Aromolaran, Adeyemi D., Oyeyinka, Isaiah. K., Olukotun, OluseyiOkon, Benjamin E. (2013). Binary Logistic Regression of Students Academic Performance in Tertiary Institution in Nigeria by Socio-Demographic and Economic Factors.*International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2(4) pp. 590 – 596*

Atalharul, M. Islam, Rafiqul I. Chowdhury (2017). A Generalized Right Truncated Bivariate Poisson Regression Model with Applications to Health Data.*PLoS ONE 12(6): e0178153. https://doi.org/10.1371/journal.pone.0178153.*

Boateng, A. (2012). Modelling the Occurrence and Incidence of Malaria.Master's Thesis of Department of Mathematics, Kwame Nkrumah University of Science and Technology.

Felix, Famoye, John T. Walu, Jr. and Karom P. Smgh. (2004). On the Generalized Poisson Regression Model with an Application to Accident Data " *Journal of data Science 2 (2004), 287-295*

Ferenc, Moksony–Rita Hegedűs (2014). The use of Poisson Regression in the Sociological Study of Suicide" *corvinus journal of sociology and social policy 2 (2014)*

George, Uchechukwu, Ibiok, Elizabeth Uwemedimoh and Nwabueze, Joy Chioma (2016).Logit Modeling of Some Factors Affecting the Academic Performance of Graduating Science Students. *International Journal of Science and Technology Volume 5 No.3.*

Igbo, J.N. (2007). Strategies for Enhancing Teaching and Learning in Tertiary Institutions: A study of University of Nigeria, in Nworgu, B.G. (Ed) Optimization of Service Delivery in the Education Sector: Issues and Strageties.

Jonas, Badin Alonso; Jorge Alberto Achcar; LuizKoodiHotta; (2005).Climatic Changes and their Effects in Public Health Using Poisson Regression Models.

Lisa, A. White (2009). Predicting Hospital Admissions with Poisson Regression Analysis.

Nakayam T., Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2004).Geographically Weighted Poisson Regression for Disease Association Mapping.

Noor Abdulrazzaq, Mustapha Kamal, AshjanMuhsen, Ali Tareq, Riyadh Al Zubaidi, Ali Al Mousawi (2017). Academic Failure and Students Viewpoint : The Influence of Individual, Internal and External Organizational Factors. *Scientific journal of medical research vol.1, issue.1,pp,1-5,winter, (2017)*

Philip Nilsson, Sebastian Nilsson (2015). Application of Poisson Regression on Traffic Safety.

Rodriguez, G. (2007). Poisson Models for Count Data.

Shalabh, IIT Kanpur Regression Analysis, *pp. 1-10*

Vijai Kumar Dwivedi, ThapeloSediadie and Njoku Ola Ama (2016). Factors Affecting Children Ever Born (CEB) in Botswana: Application of Poisson Regression Model. *Research Journal of Mathematical and Statistical Sciences Vol. 4(10), 1-9, November (2016).E-ISSN 2320-6047*