PREDICTING STUDENT COMPLETION STATUS USING LOGISTIC REGRESSION ANALYSIS

Albert Luguterah, PhD
Department of Statistics, University for Development Studies, Ghana
Benjamin Apam, BSc
Department of Statistics, Bolgatanga Polytechnic, Ghana

Abstract

To fulfil the mandate of the Polytechnics, which is to fill the gap in the manpower supply needs of the country, in an appropriate and timely manner, students must perform well and successfully complete their course of study in a timely manner. A failure of a student in a subject could among other things, increase the number of credit hours a student takes in a particular semester, affect the academic performance of the student and delay the time of certification of that student. Using longitudinal data of students' academic performance in each semester, this study used the logistic regression technique to study the influence of Gender, mode of entry, study intensity and the Age of the student on the likelihood of a student completing without ever failing in any subject in the Bolgatanga Polytechnic. The results showed that Gender, Mode of entry, Study intensity and the Age of the student, all significantly influenced the completion status of a student. Among other things, Female students were shown to be twice as likely to complete their study without ever failing in any subject as compared to their male counterparts. The results suggest that some targeted interventions, for particular student segments, could minimize the possibility of student failing.

Keywords: Completion status, Logistic regression, Random-intercept model, Polytechnic Education, Ghana.

Introduction

Polytechnics are higher educational institutions responsible for the training of students in Scientific and Technical subjects. Following the industrial development policy and rapid technological progress in a broad range of areas in the 1960's, technical education became a necessity for the country. The major focus of the Polytechnics in Ghana has been to fill the gap in the manpower supply needs of the country (Nsiah-Gyabaah, 2005a). To fulfil this mandate in an appropriate and timely manner, students must

perform well and successfully complete their course of study in a timely manner.

Many studies have been conducted on the programme orientation of the Polytechnics and academic performance of students in Ghana (Abledu, 2012; Nsiah-Gyabaah, 2005a; Nsiah-Gyabaah, 2005b; Aryeetey, 2000; Agodzo & Songsore, 2005; Nyarko, 2011; Owusu-Agyeman, 2006). Some studies have also been done on the factors that influence the successful completion of a student. Marvin (2006) found that college completion was positively associated with an institution's tuition revenue as a percentage of the total revenue whereas de Valero (2001) established that selected departmental characteristics were associated with longer times to degree and lower completion rates among doctoral students.

Given the high costs associated with Polytechnic education in recent times, the current national climate of diminishing resources for higher education, and an increased competition for these resources between the Universities, Polytechnics and the Training Colleges, understanding and examining the factors that affect students' ability to complete successfully without having a trail in a timely manner becomes crucial.

A failure of a student in a subject could among other things, increase the number of credit hours a student takes in a semester, affect the academic performance of the student and delay the time of certification. Luguterah and Apam (2013) showed that an increase in the number of credit hours taking by the student in a semester, negatively affected the performance of the student.

The importance attached to the improvement of the quality and quantity of the human resource coming from the Polytechnics underline the need to maintain a continuous flow of employable Polytechnic graduates. It is therefore important to study the factors that affect the successful completion of students of the Polytechnics.

Materials and Methods

Data on students' grade point average, demographic and Academic characteristics pertinent to this study, was obtained from the Examinations Department and the Students Affairs Unit of the Bolgatanga Polytechnic. The cohort of students, admitted in 2009, and who had successfully reached their final year in 2012, formed the basis of this study.

The logistic regression was employed to study the relationship between the completion status of the student and Age, Gender, total number of credit hours (intensity of the course) and the mode of entry of the student into the Polytechnic. The completion status of a student was categorised as having completed without ever failing any subject (1) and completed but ever failed at least one subject (0).

The logistic function describes the mathematical form on which the logistic model is based. According to Hosmer & Lemeshow (2000), this function is given by

$$f(z) = \frac{1}{1+e^{-z}}, \quad -\infty < z < +\infty \quad\text{.......................................} (1)$$

where

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad\text{................................} (2).$$

The probability being modelled can be denoted by the conditional probability statement $P(G=1/X_1, X_2, \cdots, X_k)$, conveniently referred to as P(X).

$$P(X) = \frac{1}{1+e^{-(\beta_0 + \sum \beta_i X_i)}} \quad\text{...} (3)$$

where the terms β_0 and β_i in this model, represents unknown parameters that needs to be estimated based on the data obtained on the covariates (age, sex, entry mode, and study intensity).

The odds ratio is the primary parameter estimated when fitting the logistic regression. A transformation of P(X) that is central to the application of the logistic regression is the logit transformation. This logit links the logistic function to a linear model and is defined, in terms of P(X), as:

$$g(x) = \ln\left[\frac{P(X)}{1-P(X)}\right]$$

$$g(x) = \beta_0 + \sum \beta_i x_i \quad\text{…………………………………………………} (4)$$

Amending the standard logistic regression to have random effects for each department, g(x) becomes

$$Y = X\beta + Zu + \varepsilon \quad\text{…………………………………………………..} (5)$$

Where **Y** is the $n \times 1$ vector of responses, **X** is an $n \times p$ covariate matrix for the fixed effects β and **Z** is the $n \times q$ covariate matrix for the random effects **u**. The $n \times 1$ vector of errors, ε, is assumed to be multivariate normal with mean zero and variance matrix $\sigma_\varepsilon^2 R$.

The random portion, $Zu + \varepsilon$, assumes that **u** has variance – covariance matrix **G** and that **u** is orthogonal to ε so that

$$\text{var}\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{pmatrix} G & 0 \\ 0 & \sigma_\varepsilon^2 R \end{pmatrix}.$$

The random effects **u** are not directly estimated, but instead are characterized by the elements of the variance component (**G**), that are estimated along with the overall residual variance σ_ε^2 and the residual – variance parameters that are contained within **R**.

We considered the random intercept model and again, extended it to allow for a random slope on sex with the aim of comparing the two to see the model that best fits the data. Thus;

The Random – Intercept model on sex for graduation status is

$$g(x) = \beta_0 + \beta_1 sex_{ij} + \beta_2 age_{ij} + \beta_3 cr_{ij} + u_i \ \dots\dots\dots\dots\dots\dots\dots\dots \ (6)$$

While the Random – coefficient model on sex for graduation status is

$$g(x) = \beta_0 + \beta_1 sex_{ij} + \beta_2 age_{ij} + \beta_3 cr_{ij} + u_{0i} + u_{1i} sex + \varepsilon_i \ \dots\dots\dots\dots (7)$$

The variables **sex** is the gender of the respondents, **age** is the age of the respondents, and **cr** is the number of credit hours taken in a semester by the student.

The random effects, u_i, serves to shift the logistic regression line up or down according to each department.

Results and Discussion

The results of the logistic regression analysis, for both the Random-Intercept and Random-Slope models are presented in Table 1. Both models showed that at the 5% significance level, female students have a higher chance of completing successfully without ever having failed a course than their male counterparts and that an older student was more likely to complete his studies without ever having trailed in a subject than a younger student. A student with a higher number of credit hours is less likely to complete successfully without ever having failed a subject, whereas students admitted on the basis of a decisive factor and through the matured system were less likely to complete successfully without ever having failed in at least one subject as compared to students who were admitted through the direct admission process.

The lower is better model fit criteria of the Log-Likelihood, BIC and AIC tests, favour the random–intercept model as a better fit. Thus a departmental-specific shift better fit the data than a random department-specific logistic regression.

The Random – Intercept model is given by

$$P(GRD = 1) = \frac{e^Z}{1 + e^Z}$$

Where
$$Z = 5.533 + 0.662 * SEX + 0.083 * AGE - 0.20 * CR - 1.449 * NQ - 0.775 * QM$$
and GRD represents the completion status of a student.

The odds ratio are shown in Table 2: The chances of the Female student in Bolgatanga Polytechnic completing successfully without ever having failed in at least one subject is approximately two times that of their Male counterparts. This may not be a contradiction of the findings of Downs

& Becker (2011) that showed that the odds that a student will graduate within the three years of study did not significantly different by gender as it is conceivable that a student could fail in at least a subject, take them again and pass, within the three year period.

Whilst the likelihood of a student completing successfully without ever having failed in any subject decreased with an increase in the number of total credit hours taken in a semester, the converse was true for an increase in the age of a student. The effect of an increase in the credit hours (intensity of a course) taken by a student on his/her completion status, is expected and consistent with the findings of Luguterah & Apam (2013) and Downs & Becker (2011).

The odds ratios also indicated that students who were given admission directly based on their qualification were about four times more likely to complete successfully without ever having failed in any subject than students who were admitted by a decisive factor, and about two times more likely to complete successfully without ever having failed in any subject than students who were admitted through the Matured Student process. Consequently, if continued recruitment of students based on a decisive factor and through the matured system is planned, then a more dynamically targeted policy of academic support and guidance for these students will be required to reduce the risk of students from the Polytechnics completing without certificates to enable them be employed.

Conclusion

The purpose of this study was to determine the factors that influence the likelihood of a student completing without ever failing in any subject, in the Ghanaian Polytechnics. The results support the conclusion that Gender, Age and the number of credit hours taken by a student, were significant determinants of a student's likelihood of completing successfully without ever having failed any subject. Also, students admitted on the basis of a decisive factor and through the matured entrance examination are less likely to complete without ever having failed any subject, as compared to students with direct admission.

References:
Abledu, G. K. (2012). Multiple Regression Analysis of Assessment of Academic Performance of Students in the Ghanaian Polytechnics. *Research on Humanities and Social Sciences*, 2(9), 15-25.
Agodzo, S. K., & Songsore, J. (2005). Competency-Based Learning: The case of Wa Polytechnic. *Journal of Polytechnics in Ghana.*, 1(1), 1-8.
Aryeetey, E. (2000). A Diagnostic study of Research and Technology Development in Ghana. *ISSER Technical Publication*, University of Ghana, Legon. No.60.

de Valero, Y. F. (2001). Departmental factors affecting time-to-degree and completion rates of doctoral students at one land-grant research institution. *Journal of Higher Education*, 72(3), 341-367.

Downs, G., & Becker, P. (2011). What factors affect academic performance in an online Technology Management bachelors program? *Technology Management in the Energy Smart World (PICMET), 2011 Proceedings of PICMET* (pp. 1-3). IEEE.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression.* New York: Wiley.

Luguterah, A., & Apam, B. (2013). Determinants of Academic Performance in Ghanaian Polytechnics. *Journal of Mathematical Modeling and Theory*. 3(7), 7-10

Marvin, T. A. (2006). Understaning college degree completion of students with low socio-economic status: The influence of institutional financial context. *Research in Higher Education*, 47(4), 371-398.

Nsiah-Gyabaah, K. (2005a). Polytechnic Education in Ghana: The Past, Present and the Future. *A paper presented at the Kick-Off Conference, NPT/UCC Project.* University of Cape Coast.

Nsiah-Gyabaah, K. (2005b). Migration and Brain Drain - Implications for Capacity Building and Institutional Strengthening in the Polytechnics in Ghana. *Journal of Polytechnics in Ghana*, 1(1), 49-59.

Nyarko, D. A. (2011, March 23). Polytechnic Education in Ghana: The Challenges and Prospects. *NABPTEX/Polytechnics Meeting*. Accra.

Owusu-Agyeman. (2006). Government Access-Policies on Polytechnic Education in Ghana: The Relevance for Cape-Coast Polytechnic.

Appendix

Table 1: Mixed-effects model for Graduation Status

Estimation method	MODEL			
	Random-Intercept model (1)		Random - Slope model (2)	
Fixed-effects parameter	*Estimate (SE)*	p-value	*Estimate (SE)*	p-value
Intercept	5.533 (1.036)	0.000***	5.544 (1.045)	0.000***
Female	0.662 (0.201)	0.001**	0.691 (0.349)	0.047*
Age	0.083 (0.030)	0.006**	0.082 (0.030)	0.007**
Credit Hours (CR)	-0.200 (0.023)	0.000***	-0.201 (0.023)	0.000***
Dummy variable on Decisive Factor (NQ)	-1.409 (0.309)	0.000***	-1.401 (0.311)	0.000***
Dummy variable on qualified matured (QM)	-0.775 (0.262)	0.003**	-0.774 (0.263)	0.003**
Covariance parameter	*Estimate (SE)*		*Estimate (SE)*	
σ^2 (Random Intercept)	1.220 (0.787)		- 0.1159 (0.881)	
σ^2 (Sex)			0.0933 (0.372)	
σ^2 (Residual Variance)			1.3013 (0.492)	
Model fit criteria				
Log-likelihood	-530.468		-530.440	
BIC	1114.125		1129.265	
AIC	1074.9		1078.9	
Likelihood ratio test				p-value
Model (1) nested in Model (2)				0.9724

$^{*}p < 0.05,\ ^{**}p < 0.01,\ ^{***}p < 0.001$

Table 2: Odds ratios

	Odds Ratio (SE)	p-value
Female	1.996 (0.696)	0.001
Age	1.086 (0.033)	0.006
Credit hours	0.818 (0.019)	0.000
Decisive Factor (NQ)	0.246 (0.077)	0.000
Matured entrance Examination (QM)	0.461 (0.121)	0.003