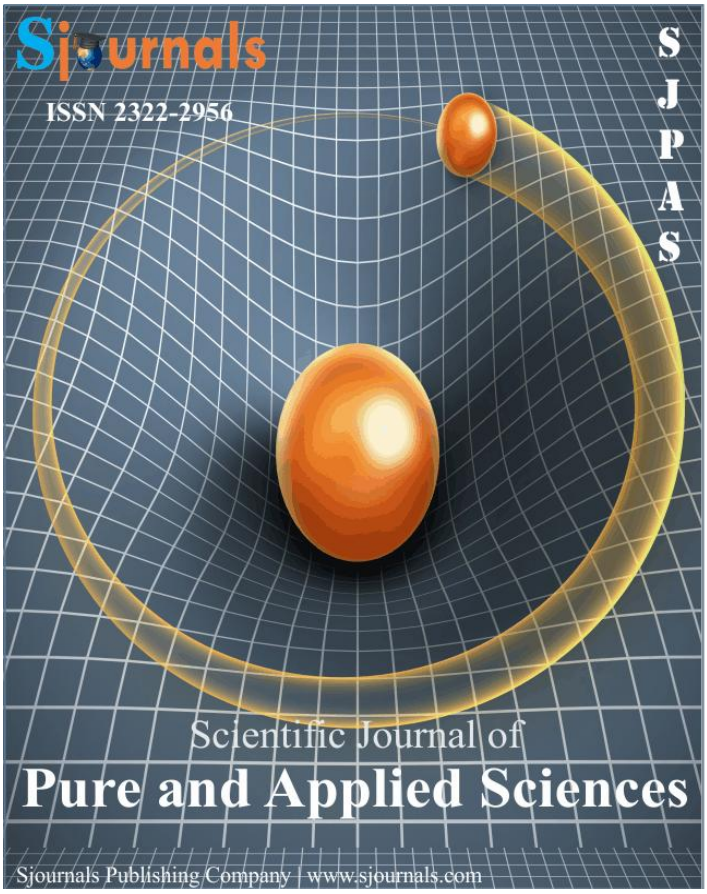
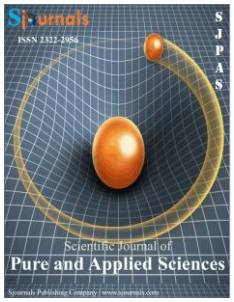


Sjournals
Scientific Journals
Scientific Journal of Pure and Applied Sciences (2016) 5(2) 360-366
ISSN 2322-2956
doi: 10.14196/sjpas.v5i2.2135



Contents lists available at Sjournals
Scientific Journal of **Pure and Appiled Sciences**
Journal homepage: www.Sjournals.com

Original article

Factors influencing the chances of making a Ph.D grade at the end of M.Sc: A logistic regression approach

Akinrinade Fatai*, Muritala Abdulkabir, Ahmadu Abdulaziz Oshioke
Department of Statistics, University of Ilorin, Nigeria.

*Corresponding author; Department of Statistics, University of Ilorin, Nigeria.

ARTICLE INFO

Article history,
Received 15 January 2016
Accepted 14 February 2016
Available online 21 February 2016
iThenticate screening 18 January 2016
English editing 12 February 2016
Quality control 18 February 2016

Keywords,
Waiting time
Ph.D grade
M.Phil grade
Binary logistic regression

ABSTRACT

Over some decades now, Students have been coming back for postgraduate studies after first degree in any University. The time of returning back varies from one graduate to the other depending on so many factors. At the beginning of this programme, the target or goal of every serious student will be to make a Ph. D grade. To this effect, the main objective is to determine which factor(s) (Waiting time (year), Gender, Marital status, B.Sc grades /Results and Institutions) influenced our chances of making a Ph. D grade using binary logistic regression approach, it observed that waiting time and grade of first Degree are the only factors that influenced the chance of making Ph. D grade. Graduates with good grade and with minimum waiting time have brighter chances of making Ph. D grade at the end of the programme.

1. Introduction

Due to some research carried out of recent on the state of higher education in Nigeria by the World Bank and United Nation Educational Scientific and Cultural Organization (UNESCO) has revealed the degradation of the Nigerian educational system. Nigeria used to be a country that produced world-class university graduates that could compete with their counterparts around the world and hard work was their watchword. Then, the fruits of hard work were imbedded in the heart and mind of every student because they saw how hardworking their

parents were on the farms and moral education was part of family responsibilities. Today, students are unable to compete in a standard academic atmosphere due to the waning and destructive role of the government, the university-community and the family.

Regrettably, most parents lack all ingredients to be called fathers and mothers because the family structure has been destroyed. These so-called parents hardly oversee the growth and development of their children, rather they chase after the fruits of the world. Family values and respect are now things of the past. Today, we only produce the worst set of uneducated tertiary graduates that cannot structure a simple sentence. Prior to the Nigerian oil industry evolution in the late 1960s, the Nigerian economy was totally-funded by revenue generated from cash crops such as groundnut, cocoa and palm oil, all

2. Materials and methods

The analysis of this research work will be limited to the variables available in the data collected. The variables are percentage scores (%) of the M.Sc graduates after the completion of the programme, the Waiting times (years) of the M.Sc students, Gender of the M.Sc students, Marital status of the M.Sc graduates, B.Sc grades/ Results of the M.Sc graduates and the Institutions that the M.Sc graduates attended during their First degree. Based on the above listed variables, binary logistic regression analysis will be carried out to know the relationship between the dependent variable (Ph.D status) against the set of independent variables (Waiting time, Gender, Marital status, B.Sc grade and Institutions of the M.Sc graduates).

2.1. Logistic regression

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

In logistic regression, the dependent variable is binary or dichotomous, i.e. it only contains data coded as Ph.D status (0 for below 60% and 1 for 60% & above). For the independent variables, the data are coded as follows: Waiting time (I for 1 year , 2 for 2 years, 3 for 3 years and 4 for 4 years & above), Gender (0 for male and 1 for female), Marital status (0 for single and 1 for married), B.Sc grades (1 for First Class, 2 for Second Class Upper, 3 for Second Class Lower, 4 for Third Class and 5 for PGDS) and Institutions (0 for Unilorin and 1 for Others).

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic}$$

Where “p” is the probability of having a Ph.D grade at the end of M.Sc or the probability that Y for cases equal 1, p(Y=1)

“1-p” is the probability of not having a Ph.D grade at the end of M.Sc or the probability that Y for cases equal 0, 1-p(Y=1)

“ln” is the natural logarithm, $\log_{exp}$, where e=2.71828

“p/(1-p)” is the odds

ln[p/1-p] is the log odds, or “logit”i.e

$$logit(p) = \ln\left(\frac{p}{1-p}\right).$$

2.2. Estimation of parameters

Rather than choosing parameters that minimize the sum of squared errors (like in ordinary regression), estimation in logistic regression chooses parameters that maximize the likelihood of observing the sample values.

2.3. Maximum likelihood estimation

Suppose we have a random sample of n observation y_i's and we let f(y_i) represents the probability that y_i = 1 or 0 then, the joint probability of observing the n , y_i's values is given as follows :

$$f(y_1, y_2, \ldots\ldots\ldots, y_n) = \prod f(y_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Since y_i ~ B (p) then

$$f(y_i) = pi^{yi}(1 - pi)^{1-yi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

hence (1) becomes :

$$f(y_1, y_2, \ldots\ldots\ldots, y_n) = \prod pi^{yi}(1 - pi)^{1-yi} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

(3) is known as Likelihood function of "Y" [L(y)]

Because the manipulation of (3) may be difficult to arrive at , we take the logarithm to make the simplification easier. Therefore, the log [L(y)] is given below:

$$\text{Log } L(y) = \sum [\, y_i \ln (p_i/1\text{-}p_i) + \sum \ln (1\text{-}p_i) \,] \quad \ldots\ldots\ldots\ldots\ldots (4)$$

In terms of model parameter (4) becomes

$$\text{Log } L(y) = \beta \text{ ' } X' \, Y - \sum \ln (1 + e^{X'\beta}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

To estimate the parameter of the model β_i, I = 0,1,2,……n we differentiate (5) partially with respect to the parameter of interest and equate the result to zero. The solution of the resulting equation will give an estimate of the model parameter of interest.

Simple logistics regression ,

$$\Lambda_i = \text{Log } (p_i/1\text{-}p_i) = \beta_0 + \beta_1 X_i$$

Where Λ_i is the logit of the probability of having a Ph.D grade or not having a Ph.D grade at the end of M.Sc

The logistic model above can be fitted by a linear model

$$E(\gamma) = E\{ \text{Log } (p_i/1\text{-}p_i)\} = \beta_0 + \beta_1 X_i$$

Where $\gamma = \text{Log } (p_i/1\text{-}p_i)$ are the observed logits (p_i's are the estimates of π_i).

$$\text{Therefore } E(\gamma) = \beta_0 + \beta_1 X_i$$

We can compare the above with the classical simple linear regression model

$E(\gamma) = \beta_0 + \beta_1 X_i$ where consistent estimates of the parameters β_0 and β_1 can be easily obtained by ordinary least square (OLS) method. The OLS assumes that the y_i 's are homoscedastic, that is the y_i has constant variance denoted by σ^2 and hence the OLS estimates are given by:

$$\beta = (X'X)^{-1} X'Y$$

$$\text{Where } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

2.4. Odds ratio

The odds ratio is one of a ranging statistics used to assess the risk of a particular outcome if a certain factor is present or not. The odds ratio is a relative measure of risk, telling us how much more likely it is that someone who is exposed to the factor under study will develop the outcome as compared to the person who is not exposed to it. Odds are ways of presenting probabilities. The odds of an event happening is the probability that the event will happen divided by the probability that the event will not happen.

Odds ratio is directly derived from regression co-efficient in logistic model. If β_i represents the regression co-efficient for independent X_i, then exponentiation β_i yields the odds ratio when all other factors are held constant. It means the change in the odds of Y given a unit change in X_i. An odds ratio of one indicates that the condition or event under study is equally likely to happen in both groups. An odds ratio greater than one indicates that the event under study or consideration is more likely to happen in the first group and an odds ratio of less than one indicates that the condition is less likely to happen in the first group.

3. Analysis

Binary logistic regression was used to analysed the data collected and the following results were obtained as SPSS outputs. The level of significant used is 0.05 with 95% accuracy and the decision is based on the critical region that we reject the null hypothesis if the p-value is less than 0.05, otherwise we accept the null hypothesis.

3.1. Estimating the parameters for the full model

The model that expressed all the factors and their parameters is as follows:

3.1.1. Variables in the Equation

Table 1
Showing the model that expressed all the factors and their parameters.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Waiting time			5.991	3	0.112	
	Waiting time(1)	1.862	1.100	2.868	1	0.090	6.439
	Waiting time(2)	1.679	0.879	3.652	1	0.056	5.362
	Waiting time(3)	2.362	0.992	5.673	1	0.017	10.608
	Gender(1)	0.745	0.774	.926	1	0.336	2.106
	Marital status(1)	0.674	0.740	.828	1	0.363	1.961
	B.Sc Grade			21.865	4	<0.001	
	B.Sc Grade(1)	19.819	1.353E4	<0.001	1	0.999	$4.050x10^{8}$
	B.Sc Grade(2)	1.953	1.169	2.792	1	0.095	7.050
	B.Sc Grade(3)	-1.809	1.014	3.184	1	0.074	0.164
	B.Sc Grade(4)	-2.581	1.232	4.391	1	0.036	0.076
	Institution(1)	0.657	0.741	0.788	1	0.375	1.930
	Constant	-1.571	1.276	1.515	1	0.218	0.208

3.2. Selection of the best model

The backward selection technique with Wald criterion was used to select the best fitting model that describes the relationship between the probability of the outcome variable and the set of independent variables. It involves four steps. At the first step, all the independent variables were involved. At the second step, the most insignificant independent variable “Institution “was removed. At the third step, the variable “Gender “was removed and finally at the fourth step, the variable Marital status was removed leaving the two significant variables “Waiting time & B.Sc grade “.

Table 2
Showing step 1.

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1	Waiting time			5.991	3	0.112	
	Waiting time(1)	1.862	1.100	2.868	1	0.090	6.439
	Waiting time(2)	1.679	0.879	3.652	1	0.056	5.362
	Waiting time(3)	2.362	0.992	5.673	1	0.017	10.608
	gender(1)	0.745	0.774	0.926	1	0.336	2.106
	Marital status(1)	0.674	0.740	0.828	1	0.363	1.961
				21.865	4	<0.001	
	B.Sc grade(1)	19.819	1.353E4	<0.001	1	0.999	$4.050x10^{8}$
	B.Sc grade(2)	1.953	1.169	2.792	1	0.095	7.050
	B.Sc grade(3)	-1.809	1.014	3.184	1	0.074	0.164
	B.Sc grade(4)	-2.581	1.232	4.391	1	0.036	0.076
	institution(1)	0.657	0.741	0.788	1	0.375	1.930
	Constant	-1.571	1.276	1.515	1	0.218	0.208

Table 3
Showing step 2.

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 2	Waiting time			6.532	3	0.088	
	Waiting time(1)	2.008	1.081	3.450	1	0.063	7.452
	Waiting time(2)	1.786	0.873	4.187	1	0.041	5.965
	Waiting time(3)	2.437	0.994	6.018	1	0.014	11.443
	gender(1)	0.635	0.759	0.700	1	0.403	1.886
	Marital status(1)	0.760	0.740	1.054	1	0.305	2.137
	B.Sc grade			21.623	4	<0.001	
	B.Sc grade(1)	19.971	1.340E4	<0.001	1	0.999	$4.715x10^{8}$
	B.Sc grade(2)	1.910	1.135	2.831	1	0.092	6.752
	B.Sc grade(3)	-1.786	0.984	3.295	1	0.069	0.168
	B.Sc grade(4)	-2.704	1.210	4.997	1	0.025	0.067
	Constant	-1.077	1.098	0.962	1	0.327	0.341

Table 4
Showing step 3.

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 3	Waiting time			6.333	3	0.096	
	Waiting time(1)	1.945	1.056	3.394	1	0.065	6.997
	Waiting time(2)	1.830	0.873	4.395	1	0.036	6.237
	Waiting time(3)	2.282	0.953	5.738	1	0.017	9.799
	Marital status(1)	0.678	0.730	.863	1	0.353	1.970
	B.Scg rade			21.995	4	<0.001	
	B.Sc grade(1)	20.092	1.336E4	<0.001	1	0.999	$5.318x10^{8}$
	B.Sc grade(2)	2.059	1.112	3.424	1	0.064	7.834
	B.Sc grade(3)	-1.730	0.964	3.223	1	0.073	0.177
	B.Sc grade(4)	-2.520	1.163	4.694	1	0.030	0.080
	Constant	-0.614	0.929	0.437	1	0.509	0.541

3.3. Model summary

Table 5
Showing the best model.

Predictors	B	S.E	Wald	d.f	Sig.	Exp(B)
Waiting time(1)	2.189	1.024	4.565	1	0.033	8.922
Waiting time(2)	1.948	0.869	5.024	1	0.025	7.013
Waiting time(3)	2.304	0.948	5.906	1	0.015	10.018
B.Sc grade(First Class)	20.350	1.362E4	<0.001	1	0.999	$6.888x10^{8}$
B.Sc grade(2/1)	2.184	1.115	3.837	1	0.050	8.881
B.Sc grade (2/2)	-1.568	0.954	2.707	1	0.100	0.208
B.Sc grade(3^{rd} class)	-2.331	1.140	4.183	1	0.041	0.097
Constant	-0.683	0.936	0.532	1	0.466	0.505

4. Interpretation of odds ratio

$e^{2.189}$ = 8.922, it means that a student who had stayed at home for 1 year has a chance which is 8.922 times that of a student who had stayed at home for 4 years (reference) of making a Ph.D grade at the end of the programme.

$e^{1.948}$ = 7.013, it means that a student who had stayed at home for 2 years has a chance which is 7.013 times that of a student who had spent 4 years at home (reference) of making a Ph.D grade at the end of the programme.

$e^{2.184}$ =8.881, it means that a student that had Second Class Upper (2/1) during his/ first degree has a chance which is 8.881 times that of a student who passed through PGDS (reference) of making a Ph.D grade at the end of the programme.

$e^{-2.331}$ =0.097, it means that a student who came through PGDS (reference) has a chance which is 0.097 times more than a student who had Third Class during his/her first degree of making a Ph.D grade at the end of the programme.

The final model is given as

$$log_e\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \beta_1 X_{1I} + \beta_4 X_{4J}$$

Where

$$\beta_1 = \begin{bmatrix} 2.189 \\ 1.948 \\ 2.304 \end{bmatrix} \text{ and } \beta_4 = \begin{bmatrix} 2.184 \\ -2.331 \end{bmatrix}$$

Also X_{1i} represent the Waiting times (year) and X_{4j} represent B.Sc grades.

5. Results and discussion

Majority of the Students were admitted for M.Sc after 2 years of graduation. We have More Males coming for the M. Sc. program compared to the females. Most of the M.Sc. students were married. Students that had Second Class Upper(2.1) in their first degrees are showing more interest in the program and about 77.6% of the total M.Sc students graduated from University of Ilorin**.**

6. Conclusion

The chance of having a Ph.D Grade at the end of M.Sc only depend on the Waiting time of the Students as well as the B.Sc grades /Results of the students at 0.05 level of significance. Making a Ph. D. grade from the M. Sc. Programme has nothing to do with the Gender, the Marital Status and the Institutions of the students. 82.2% of the M.Sc Students under consideration were correctly classified using our final model.

References

Affero, I., Norhasni, Z.A., 2009. The importance of graduates' needs on supervisory contribution in a Malaysian public University. The social sciences, 4(4), 355-365. ISBN 0134-548.

Afolakemi, O.O., 2008. Postgraduate students' supervision and training in Nigerian tertiary institutions: A comparative study.

Aspland, T., Edwards, H., O'Leary, J., Ryan, Y., 1999. Tracking new directions in the evaluation of postgraduate supervision. Innovative Higher Education, 24(2), Winter 127 – 146.

Baker, R., 1999. Towards a sustainable relationship between supervisor and student, in A. Holbrook and S. Johnston, eds. Supervision of postgraduate research in education, Review of Research in Education No 5,. Australian Association for Research in Education, Coldstream, Vic.

Boyd, C.R., Tolson, M.A., Copes, W.S., 1987. Evaluating trauma care: The TRISS method. Trauma score and the injury Severity Score. J. Trauma. 27(4), 370-378.

Brain, E., 2002. Postgraduate supervision: Is having a PhD enough? A paper presented at the Australian association for research in education conference. Problematic Futures: Educational Research in an Era of …Uncertainty, 1 – 5, December.

Grant, B., 2001. Fighting for space in supervision: Fantasies, fairytales, fictions and fallacies, Innovation and Links Conference: Research Management and Development & Postgraduate education, Auckland University of Technology, 26-27th November.

Harrel, Frank, E., 2001. Regression modeling strategies. Springer- Valley. ISBN 0-387-9523-2.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer 6.
Knoke, D., Burke, P., 1980. Log –linear models (Quantitative Applications in Social Science).

How to cite this article: Fatai, A., Abdulkabir, M., Oshioke, A.A., 2016. Factors influencing the chances of making a Ph.D grade at the end of M.Sc: A logistic regression approach. Scientific Journal of Pure and Applied Sciences, 5(2), 360-366.