Management Dynamics
Vol. 23, No. 1: 1-8, 2020
Shanker Dev Campus
Doi: https://doi.org/10.3126/md.v23i1.35514

Analyzing student performance in secondary education examination using logistic regressions

Krishna Prasad Acharya[*]

Abstract

The performance of the school examination results (CGPA) of 533 students who were admitted to higher secondary school in management stream of 2073 in Dolakha district was analyzed by examining their cumulative grade point average (CGPA) using binary logistic regression model. Factors affecting the CGPA were investigated. The factor that significantly influenced the CGPA in Secondary Education Examination (SEE) was types of schools. Other factors including gender and teaching language were found to be statistically not significant.

Keywords: *Likelihood ratio tests, logistic regression model, odds ratio, Wald statistics*

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 students measured in terms of scores or grades obtained by them in examinations 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 (Saha, 2012).

The performance measure corresponding to different independent variables may be analyzed using logistic regression analysis. This analysis has been successfully employed in the performance of the students. Factors affecting performance have been investigated by many researchers including Asampana et al. (2017); Mustapha et al. (2016); Urrutia-Aguilar et al. (2016); Ramosacaj et al. (2015); Mabula (2015); Sule & Saporu (2015); Aromolaran et al (2013); Luguterah & apam (2013).

[*] *Associate Professor, Shanker Dev Campus, Tribhuvan University, Nepal*

Previously the final examination of grade 10 was known as School Leaving Certificate (SLC). However, this examination got a new name as Secondary Education Examination (SEE) in 2073 B. S.. Another stepping stone in education sector took place in 2073 B.S. when the government decided to evaluate the students` performance on the basis of CGPA (Cumulative Grade Point Average) instead of the traditional percentage point system.

Dolakha district is located at 133 km east from the Capital, Kathmandu. It is a mountainous district which is known to be an important district for hydroelectricity production as it homes to some of the big and highly reputed hydropower plants in Nepal. Dolakha district covers an area of 2191 sq. km. As far as the geographical variations are concerned, the altitude of this district ranges from 762m to 7134m. According to the population census of 2068 B.S., the literacy rate of the people of the Dolakha district above the age of 5 years accounts for 62.28%, where the literacy rate of male population is 73.34% and the literacy rate of female population is only 53.67% (CBS, 2011). There are 51 Secondary Schools in this district that comprises of 42 Public, and 9 Private Schools.

Since in SEE the CGPA 2.41 and above is obtained by scoring at least 60% (first division) and CGPA between 2.40 to 1.80 considered as the second division (average), scores of these students were partitioned into two categories viz. [1.80 to 2.40] and [2.41 to 4.00]. Therefore, binary logistic regression analysis seems to be appropriate to analyze examination performance.

This study is aimed at finding out the performance of SEE students who were admitted to higher secondary in management stream of 2073 in Dolakha district by collecting their cumulative grade point average (CGPA). As far as I am aware, this type of study has not been done before. Therefore, the main objective of this study is to fill this gap.

Methods and materials

Data source

Five hundred and thirty three samples were collected from the entire district of Dolakha. In this survey from a total population of 51 schools, offerings of all Secondary schools were randomly selected and data related to examination scores were collected for analysis. Results of the secondary school examination were assumed to be influenced by the parameters viz. (i) gender (boys, girls), (ii) teaching language (English, Nepali), (iii) types of schools (Govt./public, private). The logistic regression approach has been adopted to analyze examination scores.

Model

The response variable of the study is the performance status of the students which is dichotomous variable with outcomes either below 2.41 (y =1) with probability $\pi\left(x\right)=\Pr\left(Y=1\,|\,X\right)$ or above 2.41 (y = 0) with probability $1-\pi\left(x\right)=\Pr\left(Y=0\,|\,X\right)$. The conditional probability that a household head is below 2.41 given X (set of predictor variables) is denoted by $\pi\left(x\right)=\Pr\left(Y\mid X\right)$

If there are *p* independent covariates $x_1, x_2, \ldots\ldots.., x_p$ the logistic regression model can be written as

$$\pi(x)=\frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+....+\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+....+\beta_px_p}} \qquad (1)$$

Its logit transformation in terms of $\pi(x)$ is:

$$g(x)=\ln\left[\frac{\pi(x)}{1-\pi(x)}\right]=\beta_0+\beta_1x_1+\beta_2x_2+....+\beta_px_p \qquad (2)$$

The logit transformation $g(x)$ possesses many of the desirable properties of linear regression model. The ratio π (x)/1-π (x) in the logit transformation is called the odds.

In order to estimate regression coefficients of Eq. (2), likelihood equations can be formulated that are based on the principle of maximum likelihood. These equations are as follows

$$\sum_{i=1}^{n}\left[y_i-\pi(x_i)\right]=0 \qquad (3)$$

$$\sum_{i=1}^{n}x_{ij}\left[y_i-\pi(x_i)\right]=0 \qquad (4)$$

Where *j= 1, 2,…p* representing for p number of covariates. Please see Hosmer and Lemeshow (2000); Agresti, A. (1996); Hosmer, D. and Stanley, L. (1989); Neter, Kutner, Nachtsheim and Wassserman, (1996); Afifi (2004); Kleinbaum (2010) for details about likelihood functions.

Bivariate and binary logistic regressions were conducted to analyze the factors affecting the performance. Regressions were carried out using SPSS Statistical Packages (version 20).

Model evaluation

Goodness-of-fit statistics were calculated to evaluate the fit of model (2) to the data. In addition, Omnibus test, Wald test, Nagelcerke and Cox & Snell R^2 and Hosmer-Lemeshow test were also conducted to check the model fit. To examine the predictive power of the model, classification and discrimination tests were also conducted. Phi-coefficient was also calculated to check the effectiveness of Chi-square test (Table 1).

Results and discussion

Descriptive statistics

The Phi-coefficient of teaching language and types of school were found to be -0.22 and -0.23 respectively, which is virtually the same. Gender of student was not significant. However, male students show significantly good performance than female students (Asampana et al., 2017 and Sattayanuwat, 2015) and on the contrary, Yousef (2011) and Alfan & Othman (2005) concluded that female students performed better than male students (Table 1).

Table 1

Association of covariates with response variable

Dichotomous covariates with category & coding schemes	% of distribution of students	Association of covariates with performance			Phi-coefficient
		Percentage of students within category	Chi-square value	p-value	
Teaching Language: English (0) Nepali (1)	 77.3 22.7	 37.0 12.4	26.24	<0.001	-0.22
Types of school: Private (0) Govt./Public (1)	 80.6 19.4	 36.6 9.7	27.88	<0.001	-0.23
Gender: Bows (0) Girls (1)	 40.8 59.2	 29.0 32.7	0.80	0.37	0.04

Binary logistic regression model

Estimated values of coefficients along with their associated p values and other fit statistics are presented in Table 2. The estimates for parameters were statistically significant as shown by the omnibus test. The model fitted the data very well as shown by Hosmer-Lemeshow test. The coefficient of covariate (types of school) was statistically significant as the p-value was less than 0.001. The 95% confidence intervals for odds ratio (OR) did not include one. The value of Cox & Snell and Nagelkerke R^2 were 0.07 and 0.09 respectively. The signs of all coefficients were negative. This indicated that each covariate has negative impact on the performance of students.

Table 2

Estimated variables of coefficients along with odds ratio (OR), corresponding standard errors, p values, and 95% confidence intervals for model (2).

Characteristics	Beta	Wald test	OR	S.E.	P -value	95% C.I. for OR
Teaching Language:						
English			1.00			
Nepali	-0.67	2.64	0.51	0.41	0.10	(0.23,1.15)
Types of school:						
Private			1.00			
Govt./Public	-1.14	5.58	0.32	0.48	<0.001	(0.13, 0.82)
Constant	-0.51	24.34	0.60	0.10	<0.001	
-2Loglikelihood (Null Model) = 661.26, -2Loglikelihood (Full Model) = 625.44 LR Chi-square = 35.82, p<0.001, Cox & Snell R^2 = 0.07, Nagelkerke R^2 = 0.09 Hosmer and Lemeshow χ^2 test = 1.45, p = 0.23, n = 533						

The overall effectiveness of the model was assessed using the chi-square statistic. The chi-square value was 35.82 and its respective p-value was less than 0.05. This indicated a significant relationship between the dependent and the independents variable in the final model. Similar results were found by Asampana et al. (2017).

The odds ratio of the Govt. /Public school was 0.32. This means the Govt. /public school is 68% less performance than the private school, keeping the other covariate teaching language is constant . Similar results was Yousef (2011) and he found that the performance of the students who attained private schools was higher than those who attained public schools. Rijal and Shrestha (2019) also admitted that type of school was found to be significant. However, Ramosacaj et al. (2015) and Mabula (2015) found that the types of school were not statistically significant. On the other hand, the teaching language was not significant. The insignificant relationship between performance and the teaching language may be attributed to the fact that both types of school in Dolakha used Nepali as teaching language.

Classification and discrimination of the model

In order to assess the predictive power of the models, a classification table of correct and incorrect prediction was constructed, based on the predicted probability as average. A probability equal or greater than 0.5 was interpreted as being a good student (first division). The values of sensitivity, specificity and correct classification of the model are presented in Table 3 for the cutoff point of 0.5.

Table 3

Sensitivity, specificity and correct classification values

Cutoff	Sensitivity	Specificity	Correct Classification	Area Under Curve (AUC)
0.50	94.0%	25.5%	68.5%	0.60

Table 3, that the percentage of average cases correctly predicted by model is 94.0% while the percentage of good (first division) cases correctly predicted by the model is 25.5%. The overall correct classification of the model for considering cutoff value 0.50 is 68.5% and the AUC is 0.60.

Conclusion

In this study, I investigated factors affecting the performance of the students in SEE Examinations in Dolakha district using the primary data. Bivariate analysis and Binary logistic regression showed that types of school had a significant association with the performance status of the students. This is the first study ever done in Dolakha district by using the best available methods to analyze the factors accounting for performance of the students.

Acknowledgement

I would like to acknowledge Mr. Binod Adhikari for his support in data collection.

References

Afifi, A., Virginia, A. C., & Susanne, M. (2004). *Computer Aided Multivariate Analysis*. New York, Chapman & Hall/CRC.

Agresti, A. (1996). An Introduction to Categorical Data Analysis. *John Wiley and Sons, Inc.*

Alfan, E., & Othman, N. (2005). Undergraduate students' performance: the case of University of Malaya. *Quality assurance in education*.

Aromolaran, A. D., Oyeyinka, I. K., Olukotun, O., & 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)*, *2*(4), 4-590596.

Asampana, G., Nantomah, K. K., & Tungosiamu, E. A. (2017). Multinomial logistic regression analysis of the determinants of students' academic performance in mathematics at basic education certificate examination. *Higher Education Research*, *2*(1), 22-26.

CBS (2011). Nepal Living Standard Survey 2010/11, Statistical Report, *Central Bureau of Statistics*. National Planning Commission Secretariat, Government of Nepal.

Hosmer and Lemeshow (2000). Applied Logistic Regression. Second Edition. *John Wiley & Sons, Inc.*

Hosmer, D. and Stanley, L. (1989). Applied Logistic Regression, *John Wiley and Sons, Inc.*

Jayanthi, S. V., Balakrishnan, S., Ching, A. L. S., Latiff, N. A. A., & Nasirudeen, A. M. A. (2014). Factors contributing to academic performance of students in a tertiary institution in Singapore. *American Journal of Educational Research*, *2*(9), 752-758.

Kleinbaum D.G. & Klein M.(2010). *Logistic Regression: A Self Learning Text.* New York, *Springer Publications*.

Luguterah, A., & Apam, B. (2013). Predicting Student Completion Status Using Logistic Regression Analysis. *European Scientific Journal*, *9*(20).

Mabula, S. (2015). Modeling Student Performance in Mathematics Using Binary Logistic Regression at Selected Secondary Schools a Case Study of Mtwara Municipality and Ilemela District. *Journal of Education and Practice*, *6*(36), 96-103.

Mustapha, M., Usman, F. W., & Yusuf, S. (21016). A logistic regression model on academic performance of students' in Mathmatics. *Continental J. Applied Science* 11(2): 1-15, 2016.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). *Applied Linear Statistical Models*. Fourth Edition. MCB MCGraw-Hill.

Ramosacaj, M., Hasani, V., & Dumi, A. (2015). Application of logistic regression in the study of students' performance level (Case Study of Vlora University). *Journal of Educational and Social Research*, *5*(3), 239.

Rijal, T. D., & Shrestha, G. (2019). Multinomial Logistic Regression Model to Identify the Factors Associated with Academic Performance of Hearing Impaired Students of Some Selected Districts of Nepal. *Nepalese Journal of Statistics*, *3*, 41-56.

Saha, G. (2012). Stochastic modeling of the grading pattern in presence of the environmental parameter. *Electronic Journal of Applied Statistical Analysis*, *5*(1), 108-120.

Sattayanuwat, W. (2015).Determinants of Students Performance in International Trade Course. *American Journal of Educational Research*, Vol. 3, No. 11, 1433-1437.

Sule, B. O., & Saporu, F. W. O. (2015). A Logistic Regression Model of Students Academic Performance in University of Maiduguri, Maiduguri, Nigeria. *Mathematical Theory and Modeling*, *5*(10), 124-136.

Urrutia-Aguilar, M. E., Fuentes-García, R., Martínez, V. D. M., Beck, E., León, S. O., & Guevara-Guzmán, R. (2016). Logistic regression model for the academic performance of first-year medical students in the biomedical area. *Creative Education*, *7*(15), 2202.

Yousef, D. A. (2011). Academic performance of business students in quantitative courses: A study in the faculty of business and economics at the UAE University. *Decision Sciences Journal of Innovative Education*, *9*(2), 255-267.