ISSN 2229-5518

Determinants of Academic Performance-A Multinomial Logistic Regression Approach

Sharmin Sharker*, MD. Mujibur Rahman

Abstract- The aim of this study is to examine the influence of some selected socio-economic, demographic, familial, individual students' scholastic and institutional factors on the academic achievement in the 1st semester of undergraduate students. A survey was conducted by using a self-administered questionnaire for data gathering. The study participants consisted of 140 graduating students from six different departments of International University of Business Agriculture and Technology, Bangladesh. Some factors including: student's seriousness about study during their 1st semester, time spend in study, and whether they had to face problem in understanding the courses have been identified as significant determinants of academic success of students. The seriousness about study found to be more significant than the rest of the variables. It has been also found out that the student's current grade also significantly depends on their 1st semester's grade. The findings of the study would help students to understand their strength and weakness and act properly for better academic achievement. It would also assist the parents and university authorities to have a deeper understanding of the factors influencing academic performance of students and take necessary actions.

Index Terms- Academic performance, Determinants, Grade, Logistic regression, Multinomial logistic regression, Seriousness, Study hour.

---

1. Introduction

There have been many researches conducted to examine the factors that influence student's academic performance. The determinants of academic performance have been widely researched. Some of those researches have concentrated on specific topics while others focus on more general topics across the disciplines.

Researcher usually apply an educational production function to explore these relationships, where academic achievement is a function of level of attendance, student ability, time devoted to learning, parents education various attributes on an individual level and on an aggregate level the relationship between school resource variables, student background characteristics and school outcomes.

- *Sharmin Sharker is currently working as a Lecturer in department of statistics in International University of Business Agriculture and Technology, Dhaka, Bangladesh. E-mail: sharmin.rume@iubat.edu*
- *Md. Mujibur Rahman is currently working as a Lecturer in department of statistics in International University of Business Agriculture and Technology, Dhaka, Bangladesh. E-mail: mujibur@iubat.edu*

There are many factors that affect to their higher CGPA. These factors are: gender, college of study, pre-admission high school grade, level of attendance, probation status, time spends in study, father's education, and parental support and involvement (M Islam, 2014).

With due respect to the determinants of academic performance, there are several studies based on the ordinary least squares (OLS) estimation. Spector and Mazzeo (1980) produced the first study that applied a qualitative model to determine academic performance. However, their ordered probit analysis concentrated on the probability of getting a letter grade of A versus the probability of not getting an A. because there are more than two categories in a course grade, a multinomial qualitative model would render useful information that is somewhat different from that obtained with ordered probit. Two such models exist: the multinomial logit and the multinomial probit. Because of the relative computational ease, multinomial logit has been widely used, even though the error term has a logistic distribution and the restrictive assumption of the "independence of irrelevant alternatives."

The main objective of the study is to use the multinomial logit model to identify the major determinants of a student's semester grade.

2. Logistic Regression Modeling

Logistic regression employs "logit", a type of sigmoid or "ʃ" shaped logistic function or curve, which represents the probability or likelihood of an event, expressed as a categorical "dependent" variable (Zerai & Banks, 1999). In contrast to LR and DA, logistic regression does not require stringent assumptions and even if those assumptions are satisfied, logistic regression still performs well (Brod & Lundt, 1996). When a dependent variable is discrete not appropriate because basic assumptions of the OLS are violated, as shown by Pindyck and Rubinfeld (1981) (such as course grade), the OLS estimation is and Spector and Mazzeo (1980). To circumvent these problems, researchers can convert the discrete variable, the course grade, into a continuous variable, such as the probability of getting a grade. The next step is to choose an appropriate functional form to estimate this probability. A linear probability model may underestimate the true regression slope when there are many observations of a variable characterized by extreme values of choice probability (0 and 1). When there more than 2 alternatives of choice (e.g., the probability of getting an A, B, C, D or F) Multinomial or Polychotomous logistic regression is appropriate. In the logit model, the dependent variable is the odds (or, more precisely, the logarithm of the odds) that a particular event will occur given specific values of the explanatory variables.

A two-category (dichotomous) logit model can be used to determine the probability of getting a specific event, for example, poor health status. The model can be specific in the logistic functional form as

$$P = F(\alpha + \boldsymbol{\beta X}) = \frac{1}{[1+\exp(-\alpha-\boldsymbol{\beta X})]}, \qquad (1)$$

where P is the probability that an individual will have a poor health status, given the values in a vector of explanatory variables $\mathbf{X}\,(x_1,\ x_2,\ x_3, \dots, x_n)$. The $\boldsymbol{\beta}$ is a vector of the coefficients$(\beta_1,\ \beta_2,\ \beta_3, \dots,\ \beta_n)'$, and exp represents the natural logarithm. Multiplying both sides of equation (1) by $1 + \exp\left(-\alpha - \mathbf{X\beta}\right)$, dividing by P, and subtracting 1 leads to

$$\exp\left(-\alpha - \mathbf{X\beta}\right) = (1/P) - 1 = (1-P)/P \qquad (2)$$

or

$$\exp\left(\alpha + \mathbf{X\beta}\right) = P/(1-P). \qquad (3)$$

After taking the natural logarithm of both sides, the estimation model is given by

$$\log\left[P/(1-P)\right] = \alpha + \mathbf{X\beta} + u, \qquad (4)$$

where u is the error term exhibiting a logistic distribution. Now the dependent variable in equation (4) is the logarithm of the odds that poor health status is obtained.

The logit model expresses the logarithm of the odds of one grade versus another as a linear function of the explanatory variables. With five possible outcomes, the conditional logits to be estimated as follows:

$$\log(P_A/P_F) = \alpha_{AF} + X\beta_{AF} + U_{AF} \qquad (5)$$

$$\log(P_B/P_F) = \alpha_{BF} + X\beta_{BF} + U_{BF} \qquad (6)$$

$$\log(P_C/P_F) = \alpha_{CF} + X\beta_{CF} + U_{CF} \qquad (7)$$

$$\log(P_D/P_F) = \alpha_{DF} + X\beta_{DF} + U_{DF}. \qquad (8)$$

Other conditional logits can be derived from the above equations-for example,

$$\log(P_A/P_B) = \log(P_A/P_F) - \log(P_B/P_F). \qquad (9)$$

This polychotomous logit model can be estimated by generalized least squares (Theil 1970) or by maximum likelihood methods (Nerlove and Press 1973). The successful application of generalized least squares estimation requires not only a large sample but also the use of either categorical explanatory variables or the arbitrary categorization of continuous explanatory variables.

3. Analysis and Discussions

The data for our study were based on responses to a questionnaire of 140 undergraduate students. Respondents were from various departments. The variables which were considered to have some impact on the academic achievement and were added to the questionnaire. They are based on the respondent's demographic, geographic and issues related to their academic involvement. The CGPA of the 1st semester is considered as the dependent variable with categories: A, B, C, D and F containing marks 90% above, 80-89%, 70-79%, 60-69% and below 60% respectively. The independent variables consist of the students average CGPA, their exactly previous semester's CGPA, their percentage of attendance during the 1st semester, their average study hour (weekly), their father's educational qualification and place of residence.

The analysis shows that, 2.9% of the students who were selected as sample belong to 3rd semester, 10.7% belong to 4th semester, 41.4% belong to 5th, 17.9% belong to 6th and 27.1% student belong to 7th semester or more. During their 1st semester only 5% of them got A, 17.1% students got B, 43.6%, 16.4%, and 17.9% students got C, D, and F respectively. While, their immediately previous semester's result reflects that the percentage of getting "A" has been increased to 29.3% and the percentage of getting F is

ISSN 2229-5518

decreased to 17.1%. It has been found out that 13.6% students attended to 98%-100% of classes, 32.1% students attended to 90-97% of classes, 36.4% students attended to 80-89% of classes and 7.9% students attended to 70-79% of classes and 70% attended below 10% of the classes during their 1st semester.

For identifying the most possible reason for having an unexpected poor result in the 1st semester, 60.7% students claimed that they were not that much serious about their study during the first semester, 36.4% students claimed that they faced problems in understanding the courses, 28.6% students claimed that they faced problem in communicating with the respective course teacher and failed to follow their instructions properly, 8.6% students claimed that they suffered from lack of confidence in a new environment. The analysis also shows that 20% students spent more than 14 hours in study during their 1st semester, 52.9% students spent 10-14 hours, 10.7% students 5-10 hours, and 16.4% spent less than 5 hours.

It has been also found out that, the students who have got "A" in their most recent examination, 50% of them usually spend more than 14 hours for study and 30% spend 10-14 hours for study.

The analysis shows that, 17.1% of the students are involved with different types of part-time jobs and 7.1% are involved with full time jobs. Where, 73.6% students think that getting involved with any kind of jobs does not affect their study and 26.4% think that it affects their study.

The demographic information shows that, 15% student's fathers have completed their post graduation, 36.4% completed graduation and 8.6%have completed the primary education. It has been also found that 47.9% students came from urban area, 23.6% from semi urban and 28.6% students came from rural areas of Bangladesh.

From the Multinomial logistic regression analysis it has been found out that among the independent variables, only 3 variables have significant effect on the result of the 1st semester at 5% level of significance. They are, (1) whether the students were less serious about their course curriculum or not, (2) whether they faced problems in understanding the courses or not and (3) the average study hour (weekly) during their 1st semester. The results are shown in the table 1.

Table 1: Likelihood ratio test results

Effect	P-value
Lack of seriousness	0.001
Problem in understanding the courses	0.049
Study hour	0.006

In all four logistic regression lines (Table2), the grade "F" category forms the baseline group against which the other four grading groups are directly compared. This table displays the parameter estimates of the four logistic regression equations that predict membership in the four separate grade groups from categorical measures of lack of seriousness, problems in understanding the courses and the average study hour in the first semester. The table displays the B column (the unstanderdized regression slopes) for each of the four regression equations, followed by the corresponding standard error of the slopes, Wald statistics, and the significance level.

Table 2: Parameter estimates for the Multinomial logistic regression

Grade 1st	B	Std. Error	Wald	df	Sig.	Exp(B)

semesetr							
A	Intercept	-5.094	1.865	7.460	1	.006	
	[seriousness=.00]	4.268	1.308	10.642	1	.001	71.351
	[seriousness=1.00]	0b	.	.	0	.	.
	[course=.00]	2.591	1.369	3.582	1	.048	13.345
	[course=1.00]	0b	.	.	0	.	.
	[study1.new=1.00]	1.823	1.557	1.372	1	.241	6.192
	[study1.new=2.00]	-1.556	1.627	.915	1	.339	.211
	[study1.new=3.00]	.295	1.735	.029	1	.865	1.343
	[study1.new=4.00]	0b	.	.	0	.	.
B	Intercept	-2.177	.911	5.713	1	.017	
	[seriousness=.00]	2.303	.797	8.351	1	.004	10.004
	[seriousness=1.00]	0b	.	.	0	.	.
	[course=.00]	1.264	.729	3.003	1	.083	3.540
	[course=1.00]	0b	.	.	0	.	.
	[study1.new=1.00]	2.509	1.189	4.455	1	.035	12.296
	[study1.new=2.00]	.244	1.009	.059	1	.809	1.277
	[study1.new=3.00]	.036	1.219	.001	1	.977	1.036
	[study1.new=4.00]	0b	.	.	0	.	.
C	Intercept	-.490	.575	.726	1	.394	
	[seriousness=.00]	1.675	.702	5.696	1	.017	5.340
	[seriousness=1.00]	0b	.	.	0	.	.
	[course=.00]	1.171	.563	4.324	1	.038	3.226
	[course=1.00]	0b	.	.	0	.	.
	[study1.new=1.00]	.593	1.017	.340	1	.560	1.810
	[study1.new=2.00]	.622	.690	.812	1	.367	1.862
	[study1.new=3.00]	-.909	.941	.934	1	.334	.403
	[study1.new=4.00]	0b	.	.	0	.	.
D	Intercept	-.761	.636	1.429	1	.232	
	[seriousness=.00]	1.802	.758	5.653	1	.017	6.061
	[seriousness=1.00]	0b	.	.	0	.	.
	[course=.00]	.378	.655	.333	1	.564	1.459
	[course=1.00]	0b	.	.	0	.	.
	[study1.new=1.00]	.584	1.144	.261	1	.609	1.794
	[study1.new=2.00]	.033	.796	.002	1	.967	1.034
	[study1.new=3.00]	-.144	.935	.024	1	.878	.866
	[study1.new=4.00]	0b	.	.	0	.	.

Each of the four equations in Table 2 includes the intercept and the slopes for the predictors. The first equation's intercept is the log of the ratio of the probability of a student having grade "A" to the probability of that student having "F". Among the grades, each of the four subgroups, i.e., A, B, C and D are contrasted with the baseline group of "F". For example, the first equation (Z_1) shows the slopes that predict a student with grade A (compared to that of a student with grade F) and would appear as:

ISSN 2229-5518

$$Z_1 = -5.094 + 4.268(seriousness) + 2.591(course) + 1.823(study\ hr\ 1) - 1.556(study\ hr\ 2) + 0.295(study\ hr\ 3) \quad (10)$$

The slope for the predictor, lack of seriousness (seriousness) is positive and significant for all of the equations. This indicates that the students who were serious about their study during the 1st semester were significantly more likely to have grade A, B, C or D rather than F.

The odds ratio indicates that the students who were serious about their study during the 1st semester were 71.351 time more likely to have grade A than F. and 47.7 times more likely to have grade B than F.

Classification of Observed and Predicted frequencies:

Overall the model correctly predicted 53% of the students. The model was particularly good at predicting students with grade C (79%) and grade B (50%), since they respectively constituted 43.6% and 17.2% of the students.

4. Conclusions

In this study, multinomial logistic regression was used to identify those variables that determine a student's grade in the 1st semester of the undergraduate study. The results suggest that the key determinants are the student's seriousness about their study, problems in understanding the courses during the 1st semester and their average study hour. It was the best findings of the study that, generally effort and intelligence determine the grade. Demographic variables, such as place of residence, father's educational qualification do not seem to contribute. Measures of counter effort, such as involving with any kind of part-time or full-time jobs, do not seem important. It has been also found out that student's current grade is fully dependent on their 1st semester's grade. Multinomial logistic regression has shown how each subgroup of student's grades are compared to a given baseline group in regard to any number or combination of potentially-constraining conditions or predictor variables.

References

[1] Islam, M. M. (2014). Factors Influencing the Academic Performance of Undergraduate Students in Sultan Qaboos University in Oman. *Journal of Emerging Trends in Educational Research and Policy Studies (JETERAPS)* 5(4):396-404 (ISSN:2141-6990).

[2] Kang H. Park & Peter M. Kerr. (1990). Determinants of Academic Performance: A Multinomial Logit Approach. *The Journal of Economic Education*, Vol. 21, No. 2 (Spring), pp. 101-111.

[3] Nerlove, M., & S. J. Press. 1973. *Univariate and multivariate log-linear and logistic models*. Santa Monica, Cal.: Rand Corp.

[4] Pindyck, R., and D. Rubinfeld. (1981). *Econometric models and economic forecasts*. 2nd ed. New York: McGraw-Hill.

[5] Spector, L. C., and M. Mazzeo. (1980). Probit analysis and economic education *Journal of Economic Education* 11 (Spring): 37-44.

[6] Theil, H. 1970. On the estimation of relationships involving qualitative variables. *American Journal of Sociology* 76 (july): 103-154.