Bulletin of the *Transilvania* University of Braşov
Series VII: Social Sciences • Law • Vol. 8 (57) No. 1 - 2015

APPLICATION OF ORDINAL LOGISTIC REGRESSION IN THE STUDY OF STUDENTS' ACHIEVEMENT IN EXTERNAL TESTING

Sadri ALIJA[1]

Abstract: *The logistic regression describes the relationship between a binary (dichotomous) response variable and explanatory variables. In this study, we have implemented logistic regression to evaluating the probability of passing the external Testing to students in Economics' Secondary School "8 Shtatori" in Tetovo, Macedonia. Drawing on SEC (State Examination Center) for external testing database, Gender, Ethnicity, School year and Subjects of the Course are selected as predictive variables. The results verify that students Macedonian has 7.3 times more chance to gain a passing grade in External Testing compared to a student Albanian, students of second-year has 12.2% less chance to gain a passing grade in External Testing compared to a students of first-year, also students of third-year have 12.2% less chance to gain a passing grade in External Testing compared to a students of second-year.*

Key words: *Binary logistic regression, Odds Ratio, External Testing, Students' Achievement.*

1. Introduction

Logistic regression analysis is a popular and widely used analysis that is similar to linear regression analysis except that the outcome is dichotomous (e.g., success/failure or yes/no or died/lived). Logistic regression and linear discriminant analyses are multivariate statistical methods which can be used for the evaluation of the associations between various covariates and a categorical outcome.

The achievement measure corresponding to different independent variables may be analyzed using logistic regression analysis. Logistic regression has been successfully employed in the social sciences [1], [5], [10], [15], in medical and health sciences [2], and in education-al research-especially in higher education [3], [4], [13].

The Mini-Mental State Examination (MMSE) was introduced as a brief assessment of cognitive functioning [6]. The 3MS extends the MMSE. It adds four additional sub-tests (date and place of birth, word fluency, similarities, and delayed recall of words).

Several research groups have used a regression-discontinuity design to examine the educational consequences of failing an exit examination. These analyses examine

[1] Faculty of Business and Economics, SEEU, Tetovo, Macedonia-FYROM.

whether barely failing the exit examination affects the probability of high-school graduation for students on the margin of passing. Papay, Murnane, & Willett [12] and Ou [11] find that failing an exit examination reduces the probability of high-school graduation significantly and substantially for low-income students in Massachusetts and New Jersey, respectively.

The central mathematical concept that underlies logistic regression is the logit-the natural logarithm of an odds ratio. The simplest example of a logit derives from a 2×2 contingency table.

Logistic regression solves these problems by applying the logit transformation to the dependent variable. In essence, the logistic model predicts the logit of Y from X.

As stated earlier, the logit is the natural logarithm (ln) of odds of Y, and odds are ratios of probabilities (π) of Y happening (i.e., a Student who passed the course) to probabilities ($1-\pi$) of Y not happening (i.e., a Student who failed the course).

Macedonia recognized the need to reform its education system after emerging from civil conflict in 2001. The government enlisted the International Bank for Reconstruction and Development's (IBRD) aid to help schools better cope with a range of challenges, from poor teacher development to weak enrollment policies. As a result, Macedonia has boosted secondary school enrollment, established training options for teachers, and now reports higher student achievement.

The implementation in Macedonia of a standards-based assessment system for secondary qualifications since 2008 has changed the focus of assessment in the last three years of secondary schooling. Students are now assessed against both internally assessed (school-based) and externally assessed (national examinations) standards.

Based on the Secondary Education Act, one of the activities of the National Examinations Centre is organizing and administering external assessment of the students in the primary and secondary education. The external assessment is organized and administered with the support of MES (Ministry of Education and Science), the Bureau for Development of Education, Centre for Vocational Education and Training and the State Inspectorate for Education.

In the school year 2009/10 and 2010/11 pilots testing was carried out in schools in the Republic of Macedonia.

External testing, as indicate by the MES, is significant and long-term reform that aims at creating educational system and knowledge society and that the reform will depend on the current political and parochial agendas of anyone.

This paper examines the results survey data for 2331 in External Testing to students in Economics' Secondary School "8 Shtatori" in Tetovo (grade 1 to 3), in Academic Year 2013-2014 [14].

2. Logistic Regression with Binary Response

Let Y be a binary response variable, which is coded as 0 or 1, referred to as fail or pass, respectively. Then the logistic regression model is given as follows:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \qquad (1)$$

$\pi(x)$ Represents the conditional mean of Y given *x*, i.e. $E(Y \setminus x)$. The value of response variable given *x* can be expressed as $y = \pi(x) + \varepsilon$, ε is the error term. If y =1, then $\varepsilon = 1 - \pi(x)$ with probability $\pi(x)$ and if y = 0, $\varepsilon = -\pi(x)$ with

probability $1-\pi(x)$. Therefore, ε follows a binomial distribution with mean 0 and variance $\pi(x)\left[1-\pi(x)\right]$. A transformation of $\pi(x)$ which is called logit function is required:

$$g(x)=\ln\left[\frac{\pi(x)}{1-\pi(x)}\right]=\beta_0+\beta_1 x \qquad (2)$$

The unknown parameters are estimated by the method of maximum likelihood estimation with given likelihood function for $\beta=(\beta_0,\beta_1)$ given as $L(\beta)=\prod_{i=1}^{n}\pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i}$.

2.1. Fitting Logistic Model with Binary Explanatory Variables

Let us consider the interpretation of the coefficients for logistic regression model with the case where explanatory variables are at the nominal level of measurement. Assume that X is coded either 0 or 1. Then the difference between logit function when x=1 and x=0 is given as $g(1)-g(0)=\beta_1$. To interpret this result, a measure of association called odds ratio (OR) is required:

$$OR=\frac{\pi(1)/\left[1-\pi(1)\right]}{\pi(0)/\left[1-\pi(0)\right]}=e^{\beta_{11}} \qquad (3)$$

Odds ratio provides an approximation how much more likely or unlikely it is for the response variable to occur among those with x = 1 than among those with x = 0. For details, one can see Hosmer and Lemeshow [7-9].

3. Numerical Results

This study aims to analyze whether a Student's passed the course or a Student's who failed the course in External Testing by using categorical principal component logistic regression.

The dependent variable "passing" is a binary variable coded 1 for the student's has taken a passing grade and 0 when the student has taken a failing grade in relevant courses. The effects of some explanatory variables on passing grade are investigated. The explanatory variables used are: Gender (X1; 1 for male, 2 for female), Ethnicity X2; 1 for Albanian, 2 for Macedonian), School Year (X3; 1 for first-year students, 2 for second-year students, 3 for third-year students) and Subjects of the Course (X4; 1 for General Course, 2 for Vocational Course.

One of these measures is Hosmer-Lemeshow [8] test shows that the model ensures better fit than a null model with no explanatory variables. If the test statistic is not significant, then it means that the model adequately fits the data. According to Table 1, Hosmer-Lemeshow goodness of fit test statistics is not significant (p-value is 0.194) which implies that the estimated model fit the data at a convenient level.

Goodness of Fit Statistics for Logistic Regression Table 1

Model Summary				**Hosmer and Lemeshow Test**		
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square	Chi-square	df	Sig.
1	1276.413	0.29	0.57	11.141	8	0.194

To summarize the logistic regression model, there are some approximations for coefficient of determination R2, called pseudo R2. However these are not goodness-of-fit tests but rather measure strength of association. One of them is Cox & Snell R2, indicates 28.65% of the variation in the response variable is explained by the model. However, there is more reliable measure, Nagelkerke R2 indicates a strong relationship of 57.49% between explanatory variables and the prediction. Both pseudo measures tends to be lower than the traditional R2.

In the following table, we present the estimation results for logistic regression in the case where principal variables are used as explanatory variables.

Estimation Results for Logistic Regression Table 2

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
GENDER(1)	.021	.153	.020	1	.889	1.022	.756	1.380
ETHNICITY(1)	1.988	.367	29.309	1	.000	7.298	3.554	14.986
Course_Subjects	.112	.159	.492	1	.483	1.118	.818	1.528
School_Year	-.122	.098	1.544	1	.214	.885	.731	1.073
Constant	2.068	.215	92.546	1	.000	7.912		

a. Variable(s) entered on step 1: GENDER, ETHNICITY, Course Subjects, School Year.

The parameter estimation of the logistic regression model, Wald statistic and the corresponding significance level test, the significance of each of the covariate and dummy independent variables in the model are shown in the above table. If the Wald statistic is significant (i.e., less than 0.05) then the parameter is significant in the model. Of the independent variables, gender of students, Course Subjects and School Year is insignificant, whereas Ethnicity of students have significantly affected the results of students in External Testing. Then, the logistic regression model for this study can be expressed as follows:

$$\pi(x)=\frac{\exp\left(2.068+.021Gen+1.988Eth+.122CSb-.122SchY\right)}{1+\exp\left(2.068+.021Gen+1.988Eth+.122CSb-.122SchY\right)} \tag{4}$$

The Exp (B) column presents the extent to which raising the corresponding measure by one unit influences the odds ratio. We can interpret EXP (B) in terms of the change in odds. If the value exceeds 1 then the odds of an outcome occurring increase; if the fi gure is less than 1, any increase in the predictor leads to a drop in the odds of the outcome occurring. For example, the odds ratios for variable is associated with Ethnicity is 7.298>1, i.e. student Macedonian, has 7.3 times more chance to gain a passing grade in External Testing compared to a student Albanian. The estimated odds-ratio is 0.878 for School Year, indicates that students of second-year has (1-0.878) · 100 = 12.2% less chance to gain a passing grade in External Testing compared to a students of first-year, also students of third-year have 12.2% less chance to gain a passing grade in External Testing compared to a students of second-year.

4. Conclusions

The achievement of female students is approximately 1.022 times higher than that of the achievement of male students. As far as Ethnicity is concerned, it can be seen that, the achievement of students Macedonian are 7.3 times better than that of the student Albanian.

As shown in the above table, Ethnicity (i.e., Albanian and Macedonian students) is a significant factor with p value 0.000 corresponding to examination scores in External Testing.

Also it is observed that, achievement of Vocational Course is 1.118 times better than the General Course. Similarly, the achievement of first-year students are 0.885 times better than for the second-year students, and the achievement of second-year students are 0.885 times better than the for third-year students.

Other information may be obtained from the address**:** s.aliji@seeu.edu.mk.

References

1. Agresti, A.: *An Introduction to Categorical Data Analysis*. Hoboken. John Wiley and Sons, Inc., 1996.
2. Antonogeorgos, G, Panagiotakos, D.B., Priftis, K.N., Tzonou, A.: *Logistic Regression and Linear Discriminant Analyses in Evaluating Factors Associated with Asthma Prevalence among 10- to 12-Years-Old Children: Divergence and Similarity of the Two Statistical Methods.* In: International Journal of Pediatrics (2009). Available at: ttp://www.ncbi.nlm.nih.gov/ pubmed/20049344.
3. Austin, J.T., Yaffee, R.A., Hinkle, D.E.: *Logistic regression for research in higher education.* In: *Higher Education: Handbook of Theory and Research*, Paulsen, M. (ed.). Springer, New York, 1992, p. 379–410.
4. Cabrera, A.F. *Logistic regression analysis in higher education: An applied perspective.* In: Higher Education: Handbook of Theory and Research (Vol. 10, 225–256.
5. Chuang, H.L.: *High school youth's dropout and re-enrollment behavior.* In: Economics of Education Review **16** (1996) No. 2, p. 171–186
6. Folstein, M.F., Folstein, S.E., McHugh, P.R.: *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* In: J Psychiatr Res. **12** (1975) No. 13, p. 189–198.
7. Hosmer, D., Stanley, L.: *Applied Logistic Regression.* Hoboken. John Wiley and Sons, Inc., 1989.
8. Hosmer, W.D., Lemeshow, S.: *Applied Logistic Regression*. New York. John Wiley & Sons, 1989.
9. Hosmer, W.D., Lemeshow, S.: *Applied Logistic Regression. 2nd ed.* New York. Wiley-Interscience Publication, 2000.
10. Janik, J., Kravitz, H.M. *Linking work and domestic problems with police suicide.* In: Suicide and Life Threatening Behavior **24** (1994) No. 3, p. 267–274.
11. Ou, D.: *To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam.* In: Economics of Education Review **29** (2010*)* No. 2, p. 171 - 186.
12. Papay, J.P., Murnane, R.J., Willett, J.B.: *The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts.* In: Educational Evaluation and Policy Analysis ***32*** (2010) No. 1, p. 5-23.

13. Peng, C.Y., So, T.S., Stage, F. K., St. John, E.P.: *The use and interpretation of logistic regression in higher education journals: 1988–1999.* In: Research in Higher Education (2002) No. 43, p. 259–293.
14. State Examination Center. Available at: http://www.dic.edu.mk/en.
15. Tolman, R.M., Weisz, A.: *Coordinated community intervention for domestic violence: The effects of arrest and prosecution on recidivism of woman abuse perpetrators.* In: Crime and Delinquency **41** (1995) No. 4, p. 481–495.