

A DUMMY VARIABLE REGRESSION ON STUDENTS' ACADEMIC PERFORMANCE

Okewole D. Modupe

Department of Mathematical Sciences, Redeemer's University

Abstract

A good formal education system is judged by the performance of the students involved. However, there are several factors also influencing this performance. Studies that have been carried out on this has identified Economic, environmental and social factors. Most of these researches focused on the public schools. This paper is a report on a survey carried out in a Nigerian private University (Redeemer's University) having the undergraduates as the target Population. A dummy variable regression was carried out on the cumulative grade point average of the respondent. The factors considered were; the study hours, reading habit, mother's level of education, monthly allowance and involvement in extra-curricular activities. Of all these factors, reading habit and mother's level of education were significant at the 5% level. Specifically, reading on a daily basis was the significant category of the reading habit factor in the dummy variable regression while educated mother is the significant category in the mother's education factor.

Key Words: Dummy variable, Regression analysis, significant factor, Academic Performance, survey, respondents

1 Introduction

The quality of human resources, in which formal education plays major role, is very important in the socio-economic development of a society. This quality depends among other factors, on academic performance while undergoing a formal education. Statistical methods have been known to be the basis of researches whose conclusions and inferences are reliable. One of such methods is the Regression with dummy variable applied in this research work.

The research was carried out on undergraduate students of a private University. The data was collected with the use of questionnaire designed to capture the objectives of the research, which is to find out the factors influencing students' academic performance and the extent of the effect of these factors. The explanatory variables considered are: study hours, reading habit, monthly allowance, mother's education, and involvement in extra-curricular activities, all of which produce categorical data. These variables were represented with dummy variable and then regressed on the academic performance. A dummy variable is one that takes the values 0 or 1 to indicate the absence or presence of some categorical effect that may be expected to shift the outcome. The regression analysis is carried out normally as in multiple regression.

2 Theoretical background/Literature Review

Regression with dummy variable is one of the methods for studying relationship between a quantitative dependent variable and one or more qualitative explanatory variable. Some of the factors affecting students' academic performance are qualitative in nature, hence the applicability of this method.

There have been many researches on factors affecting students' academic performance. All of the research reviewed support the view that student performance depends on different socio-economic, psychological, and other environmental factors. Checchi (2000) and Agus and Makhbul (2002) in their separate researches both concluded that family income influences academic performance positively, that is, the higher the income the better the performance. On the other hand, Tahir Hijazi and S.M.M Raza Naqvi (2006) found that there is negative relationship between student performance and student family income. Parents' education is another factor found in literature to have a positive relationship with students' academic performance. According to Ermisch and Francesconi (2001), there is significant relationship between each parent's education level and their child's academic performance. Relative to a parent with no qualifications, mother's education has a stronger association with her child's academic performance than the education of the father. This result is supported by Agus and Makhbul (2002). They indicate that the level of education of mother has been found to exert the strongest influence on academic achievement as compared to level of education of father. Involvement in Extra-curricular activities has also been found out to influence academic performance (Marsh & Kleitman, 2002 and Guest and Schneider, 2004). A study on extra-curricular activities and academic achievement indicates that participation in some activities improves achievement, while participation in others reduces achievement.

The focus of this study is on the undergraduates of the private University. The use of the dummy variable regression, among other available methods is to present its application for better understanding.

3 Dummy variable Regression

A dummy variable is a numerical variable used in regression analysis to represent subgroups of the sample in a study. In research design, a dummy variable is often used to distinguish different treatment groups. In the simplest case, there is a 0,1 dummy variable where a subject is given a value of 0 if they are in the reference group or a 1 if they are in the group of interest. Dummy variables are useful because they enable us to use a single regression equation to represent multiple groups. This means that we don't need to write out separate equation models for each subgroup. The dummy variables act like '**switches**' that turn various parameters on and off in an equation.

A simple dummy variable model is given below:

$$y_i = B_0 + B_1 Z_i + e_i$$

Where y_i = the dependent variable, e_i = residual for the i^{th} unit

B_0 = the intercept

B_1 = the regression coefficient

$Z_i = 1$ if the i^{th} unit is in the group/category of interest

0 if the i^{th} unit is in the reference group

If we have more than one independent variable say (2), in which one of the variable is categorical with 3 levels for example, each level is treated as a dummy variable with its regression coefficient, further explanation on this is contained in the methodology for this research work. The analysis is done as in multiple regression models

4 Methodology

A sample size of 500 was selected from the university's student population. The sampling was done by the use of stratified sampling technique where the sample was

proportionally allocated to all the departments in the university. Questionnaires were conveniently given to respondents within each of the departments in the university. The data analysis was done with the use of SPSS

The dependent variable in this case is the student academic performance (CGPA)

The independent/explanatory variables considered were:

(1). Study hours (SH). (2) Reading habit (RH). (3) Monthly Allowance (ALW)

(4) Mother's education (5) Involvement in Extracurricular activities (ECA)

The regression model reflecting the dummy variables for the explanatory variables is thus;

$$cgpa = \beta_0 + \beta_1 SH_1 + \beta_2 SH_2 + \beta_3 RH_1 + \beta_4 RH_2 + \beta_5 ME_1 + \beta_6 ME_2 + \beta_7 ECA + \beta_8 ALW_1 + \beta_9 ALW_2$$

The interpretation of the dummy variables is as follows:

$$SH_1 = \begin{cases} 1 & \text{if student studies between 2 and 3 hours} \\ 0 & \text{otherwise} \end{cases}$$

$$SH_2 = \begin{cases} 1 & \text{if student studies between 4 and 6 hours} \\ 0 & \text{otherwise} \end{cases}$$

$$RH_1 = \begin{cases} 1 & \text{if student studies every other day} \\ 0 & \text{otherwise} \end{cases}$$

$$RH_2 = \begin{cases} 1 & \text{if student studies daily} \\ 0 & \text{otherwise} \end{cases}$$

$$ALW_1 = \begin{cases} 1 & \text{if student receives between N11,000 and N19,000} \\ 0 & \text{otherwise} \end{cases}$$

$$ALW_2 = \begin{cases} 1 & \text{if student receives between N20,000 and above} \\ 0 & \text{otherwise} \end{cases}$$

$$ME_1 = \begin{cases} 1 & \text{if mother's education is primary/secondary level} \\ 0 & \text{otherwise} \end{cases}$$

$$ME_2 = \begin{cases} 1 & \text{if mother is educated above the secondary level} \\ 0 & \text{otherwise} \end{cases}$$

$$ECA = \begin{cases} 1 & \text{YES} \\ 0 & \text{NO} \end{cases}$$

A Comparison of the effects of the categories of each variable

We make use of one of the variables (Reading habit) for illustrative purpose. This is to compare the effect of studying near test/exam, every other day and daily, by pair wise comparison of the three mean CGPAs, which will be denoted as $\mu_{rh,nt}$, $\mu_{rh,ed}$, $\mu_{rh,d}$ respectively. If we set

$$y_i = \mu_i + e_i$$

$$= \beta_0 + \beta_1 RH_1 + \beta_2 RH_2 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA + e_i$$

$$\mu_{rh,nt} = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA$$

$$= \beta_0 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA$$

$$\mu_{rh,ed} = \beta_0 + \beta_1(1) + \beta_2(0) + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA$$

$$= \beta_0 + \beta_1 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA$$

$$\mu_{rh,d} = \beta_0 + \beta_1(0) + \beta_2(1) + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA$$

$$= \beta_0 + \beta_2 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA$$

These equations imply that

$$\begin{aligned} \mu_{rh,ed} - \mu_{rh,nt} &= (\beta_0 + \beta_1 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA) - \\ &\quad (\beta_0 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA) \\ &= \beta_1 \end{aligned}$$

$$\begin{aligned} \mu_{rh,d} - \mu_{rh,nt} &= (\beta_0 + \beta_2 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA) - \\ &\quad (\beta_0 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA) \\ &= \beta_2 \end{aligned}$$

$$\begin{aligned} \mu_{rh,d} - \mu_{rh,ed} &= (\beta_0 + \beta_2 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA) - \\ &\quad (\beta_0 + \beta_1 + \beta_3 SH_1 + \beta_4 SH_2 + \beta_5 ALW_1 + \beta_6 ALW_2 + \beta_7 ME_1 + \beta_8 ME_2 + \beta_9 ECA) \\ &= \beta_2 - \beta_1 \end{aligned}$$

5 Findings and Policy implications

The analysis showed that RH_2 and ME_2 are the only significant explanatory variables in the model. RH_2 represent studying on a daily basis while ME_2 represent mother's is education above secondary school level. These findings imply that reading everyday contributes more to higher CGPA than other less frequent reading habit, also that students whose mothers are educated have better performance than others. Brief descriptive Statistics on the two variables reading habit and mother's education is that 20 out of the 500 respondents have their mothers not educated, 68 are educated at the primary or secondary school level while 405 are educated above the secondary school level. Also that 217 read near test/exam, 79 read every other day and 204 read daily. Although the percentage of students whose mother's are not educated is small, the significance of the test already attaches importance to this issue of mother's influence on the academic performance of their children. On the other hand, the study shows that most students read near test or exam, a habit that should be discouraged, while reading daily regardless of the study hours seems to produce better result and should be encouraged among students.

Table 1: Significance of the Regression Coefficients

Variable	$RH_1(1)$	$RH_2(2)$	$SH_1(3)$	$SH_2(4)$	$ALW_1(5)$	$ALW_2(6)$	$ME_1(7)$	$ME_2(8)$	$ECA(9)$
Coefficient	0.176	0.307	0.205	0.277	-0.133	0.141	0.269	0.507	0.138
P-value	0.141	0.001	0.111	0.069	0.279	0.202	0.237	0.015	0.078

Comparison of means to indicate the extent of the contribution of each significant level over others.

The least squares estimates of β_7 and β_8 are 0.269 & 0.507 respectively. The equations above implies that the mean cgpa of students whose mother's education is primary/secondary is 2.6 greater than students whose mother's are not educated; the mean cgpa of students whose mother's are educated above secondary school level is 5.0 greater than students whose

mother's are not educated; the mean cgpa of students whose mother's are educated above secondary school level is $0.507 - 0.269 = 0.238$ i.e. 2.4 greater than the mean cgpa of students whose mother's are educated at the primary/secondary school level. These means show that students whose mother's education is above secondary school level have more positive contribution to the student's academic performance than a lower level of education.

The comparison for the reading habit also shows that reading daily has greater contribution than the other reading habit has shown in table 2.

Table 2: Comparison of means for Reading habit and Mother's education

Reading Habit	$\mu_{rh,ed} - \mu_{rh,nt} = B_1$	$B_1 = 0.176$
	$\mu_{rh,d} - \mu_{rh,nt} = B_2$	$B_2 = 0.307$
	$\mu_{rh,d} - \mu_{rh,ed} = B_2 - B_1$	$B_2 - B_1 = 0.131$
Mothers Education	$\mu_{p/s} - \mu_{ne} = B_7$	$B_7 = 0.269$
	$\mu_{ed} - \mu_{ne} = B_8$	$B_8 = 0.507$
	$\mu_{ed} - \mu_{p/s} = B_8 - B_7$	$B_8 - B_7 = 0.238$

Note the interpretation of the notations in table 2:

$\mu_{rh,d}$ = the mean CGPA for reading every day, $\mu_{rh,ed}$ = the mean CGPA for reading every other day

$\mu_{rh,nt}$ = the mean CGPA for reading near test or exam, $\mu_{p/s}$ = the mean CGPA for mother's education at primary or secondary school level

μ_{ed} = the mean CGPA for mother's education above secondary school level, μ_{ne} = the mean CGPA for mother not educated.

Recommendations and Conclusion

Since the study shows that the reading habit of students has significant relationship with the student's performance, more frequent reading habit should be encouraged among students. Regular test & assignment can motivate a more frequent reading habit and thereby better performance. Also, since mother's education was seen to have significant relationship

with student's performance, mothers are encouraged to pay more attention to their children's education and keep themselves educated enough to be of positive influence on their children's academic performance. This will in turn bring about highly qualified personnel that will work to improve the Economy of the nation.

References:

- Agus, A and Makhbul, Z.K. (2002). An empirical study on academic achievement of business students in pursuing higher education: An emphasis on the influence of family backgrounds. *Paper presented at International Conference on the Challenges of Learning and Teaching in a Brave New World: Issues and Opportunities in Borderless Education. Hatyai Thailand*
- Checchi, D. (2000). University education in Italy. *International Journal of Manpower*, Vol 21 No 3, pages. 177-205.
- Ermisch, J and Francesconi, M (2001). Family matter: Impacts of family background on educational attainment. *Economica* Vol. 68. pages 137-156
- Guest, Andrew and Schneider, Barbara. (2003). Adolescents' extracurricular participation in context: The mediating effects of school, communities and identities. *Sociology of Education*, Vol. 76, No 2.
- Jeffrey M. Wooldridge. (2009). Introductory Econometrics. *Cengage learning*. Pages 225-260
- Marsh, H. W. & Kleitman, S. (2002). Extracurricular school activities. The good, the bad, and the nonlinear. *Harvard Educational Review* vol.72.
- Syed Tahir Hijazi and S.M.M. Raza Naqvi. (2006). Factors affecting students' performance. A case of private colleges, Bangladesh. e- *Journal of Sociology*, Vol. 3. No.1.