

On Regression Analysis of Students' Academic Performance.

R.A. Adeleke*, O.Y. Halid, and O.D. Ogunwale.

Department of Mathematical Sciences, University of Ado-Ekiti, Ado-Ekiti, Ekiti State, Nigeria.

E-mail: waleod@yahoo.com

ABSTRACT

The poor academic performance of students of Nigerian tertiary institution has been a major concern. This paper, however, considers the mode of selection for admission via the University Matriculation Examination (by JAMB), the Post University Matriculation Examination, as well as academic performance after first year in the University. Using the Ordinary Least Square (OLS) method of estimating parameters, we present a comparative assessment of these exogenous and endogenous variables as well as their test hypotheses.

(Keywords: ordinary least square, OLS, regression, variables, parameters, hypotheses, exogenous, endogenous)

INTRODUCTION

The importance of education to national and socio-economic growth and development is immeasurable. Research oriented tertiary education is central to international development since new findings are made, more knowledge are acquired, adopted and further developed to new technologies.

The quest to reduce or estimate illiteracy in developing nations is one of the priorities of their governments. This accounts for why substantial proportions of the annual budget of some countries are being allocated to education.

In many tertiary institutions in Nigeria, the number of admission seekers has exceeded the number of existing resources or facilities, this is however a cause of concern. After the supposed admission war' which includes fulfilling admission requirements (of passing WAEC, JAMB, and POST UME), it is observed that many of this students still have poor academic performance in the universities.

A method of estimating parameters called the OLS in method regressions which enable us to determine the simultaneous or joint effect of several exogenous (independent) variables will be employed.

MODEL SPECIFICATION AND RESEARCH METHOD

We consider the Cumulative Grade Point Average (CGPA) of thirty first year university students as the endogenous variables Y_i , the U.M.E and Post U.M.E performance as the exogenous variables X_{1i} , and X_{2i} , respectively. We therefore represent this as a linear regression model of order (2), that is:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i, \quad i = 1, 2, \dots, 30 \quad (1)$$

where β_1 , β_2 and β_3 are parameters and ε_i are the random error terms.

We seek the estimates of the parameter β_1 , β_2 and β_3 using the OLS method taking the following assumptions:

- (i) The random error terms ε_i are independently and identically normally distributed for each of the exogenous variables X_2 and X_3 ,
- (ii) $E(\varepsilon_i) = 0 \quad \forall \quad X_i$,
- (iii) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall \quad i \neq j \quad \varepsilon_i$ and ε_j , independently,

(iv) $V(\varepsilon_i) = \sigma_i^2$ (i.e., finite and constant variance $\forall i$)

We derive our normal equations by minimizing from (1):

$$Q = \sum \varepsilon_i^2 = \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2 \quad (2)$$

That is, differentiating (2) partially w.r.t the three unknowns and setting the resulting equations to zero, we obtain,

$$\frac{\partial Q}{\partial \beta_1} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-1) = 0 \quad (3)$$

$$\frac{\partial Q}{\partial \beta_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0 \quad (4)$$

$$\frac{\partial Q}{\partial \beta_3} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0 \quad (5)$$

We note that (3), (4), (5) can be written as:

$$\begin{aligned} \sum \varepsilon_i &= 0 \\ \sum \varepsilon_i X_{2i} &= 0 \\ \sum \varepsilon_i X_{3i} &= 0 \end{aligned}$$

which show the properties of the least squares fit, namely that the residual to zero and that they are uncorrelated with the exogenous variables X_2 and X_3 .

From (3), (4), (5) we obtain:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} \quad (3^*)$$

$$\sum X_{2i} Y_i = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} \quad (4^*)$$

$$\sum X_{3i} Y_i = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \quad (5^*)$$

(3*), (4*), and (5*) are called the normal equations.

We see at once that:

$$\hat{\beta}_1 = \hat{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \quad (6)$$

which is the OLS estimator of the population intercept $\hat{\beta}_1$.

Following the convention of letting the lowercase letter denote deviations from sample mean values, one can derive the following formulas from normal equations (4*) and (5*), so that:

$$\hat{\beta}_2 = \frac{(\sum Y_i X_{2i})(\sum X_{3i}^2) - (\sum Y_i X_{3i})(\sum X_{2i} X_{3i})}{(\sum X_{2i}^2)(\sum X_{3i}^2) - (\sum X_{2i} X_{3i})^2} \quad (7)$$

$$\hat{\beta}_3 = \frac{(\sum Y_i X_{3i})(\sum X_{2i}^2) - (\sum Y_i X_{2i})(\sum X_{2i} X_{3i})}{(\sum X_{2i}^2)(\sum X_{3i}^2) - (\sum X_{2i} X_{3i})^2} \quad (8)$$

which gives the OLS estimates of the population partial regression coefficient β_2 and β_3 , respectively.

Fitting our data (See Appendix I) to this model (1), using results (6), (7) and (8). We have:

$$\hat{\beta}_1 = 3.062, \quad \hat{\beta}_2 = -0.005, \quad \hat{\beta}_3 = -0.054$$

so that,

$$Y_i = 3.062 - 0.005 X_2 - 0.054 X_3 \quad (9)$$

We also carried test hypothesis by Analysis of Variance where:

$H_0 : \exists$ no significant relationship between the exogenous and endogenous variables.

$H_1 : \exists$ a significant relationship between the exogenous and endogenous variables.

This results in the acceptance of the null hypothesis since:

$$F_{cal} = 0.563 < 3.37 = F(0.05, 3.27)$$

(See Appendix II)

That is, no regression (significant relationship) between the variables (even at all levels of significance). We also consider the partial correlation coefficients between the variables, that is :

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad (10)$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}} \quad (11)$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}} \quad (12)$$

where $r_{12.3}$ = Partial correlation coefficient between Y and X_2 holding X_3 constant;

$r_{13.2}$ = Partial correlation coefficient between Y and X_3 holding X_2 constant;

$r_{23.1}$ = Partial correlation coefficient between X_2 and X_3 holding Y constant;

and r_{12} , r_{13} and r_{23} are simple correlation coefficients.

From (9), (10), and (11) we have that

$r_{12.3}$, $r_{13.2}$ and $r_{23.1}$ are 0.064, -0.209 and 0.134, respectively.

INTERPRETATION AND DISCUSSION OF RESULTS

From (9), $Y_i = 3.062 - 0.005X_{1i} - 0.054X_{2i}$

3.062 is the intercept which is constant;

-0.005 is the regression coefficient β_2 on X_{2i} ;

-0.054 is the regression coefficient β_3 on X_{2i} .

It then follows that a unit increase in the U.M.E. and Post U.M.E scores will cause 0.005 and 0.054 decrease in CGPA. Again, the Analysis of Variance shows that there exist no significant relationships between all the variables at all levels of significance.

Lastly, the values of the partial correlation coefficients show that there exist very weak positive linear relationship between CGPA and U.M.E score ($0 < r < 0.5$).

CONCLUSION AND RECOMMENDATION

Based on the three important results, we conclude that there exist no relationship between first year university students' performance and performance in U.M.E and Post U.M.E. Again, a weak relationship between U.M.E and Post U.M.E results clearly indicates that an excellent performance in U.M.E may not necessarily mean an excellent performance in Post U.M.E.

More importantly, an excellent performance in U.M.E may not clearly mean an excellence performance in 100 level University examinations. We therefore recommend that the Post U.M.E examination should be employed by all universities as one of the tools for screening admission seekers.

REFERENCES

1. Gujarati, D.N. 2003. *Basic Econometrics*. McGraw Hill: London, UK.
2. Koutsoyiannis, A. 2001. *Theory of Econometrics*. McMillan Press: London, UK.

3. Ryan, T.P. 2003. *Modern Regression Analysis for Scientists and Engineers*. NIST: Washington, DC.
4. Adeleke, R.A., O.Y. Halid, and B.J. Adegboyegun. 2008. "Analysis of Extreme-Value Distribution with Applications to Maximum Temperature Measurements" (Unpublished Article).
5. Smith, R.L. 1985. "Maximum Likelihood Estimation in a Class of Non-Regular Cases". *Biometrika*. 72:67-90

SUGGESTED CITATION

Adeleke, R.A., O.Y. Halid, and O.D. Ogunwale 2009. "On Regression Analysis of Students' Academic Performance". *Pacific Journal of Science and Technology*. 10(2):207-210.

