East African Journal of Education Studies
eajes.eanso.org
Volume 7, Issue 1, 2024
Print ISSN: 2707-3939 | Online ISSN: 2707-3947
Title DOI: https://doi.org/10.37284/2707-3947

EAST AFRICAN NATURE & SCIENCE ORGANIZATION

Original Article

Assessing the impact of Ordinary Level Grades on the Cumulative Grade Point Average of First Year University Students (A Factorial Design Approach)

Oluwole A. Nuga[1] Tolulope O. Adekola[1] & Abba Zakirai Abdulhamid[1]*

[1] Bells University of Technology, P.M.B. 1015, Ota, Ogun State, Nigeria.
[*] Author for Correspondence ORCID ID: https://orcid.org/0000-0003-4458-5904; Email: oanuga@bellsuniverity.edu.ng

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

Date Published:

05 February 2024

Keywords:

O' Level Grades, Main Effect Model, Interaction Effect Model, Response Surface Model, First-year CGPA.

ABSTRACT

The purpose of this work is to model the Cumulative Grade Point Average (CGPA) of first year engineering students of a private university in southwestern Nigeria using the Ordinary level (O' level) grades in mathematics, physics, and chemistry as factors. The choice of the three subjects was due to the fact that virtually all the courses taken in first year by engineering students require a solid background in these three subjects. Duplicate samples were randomly selected from a population stratified into all possible factor levels The O' level grades were converted to scaled variables (as typically done in factorial design) and used as the model matrix of six levels of a three-factorial design. Three orthogonal statistical models were fitted namely; first order, interaction and response surface models using the Ordinary Least Squares (OLS) Estimators. The model of best fit was identified and used to obtain the combination of ordinary level grades that maximized and minimized first year CGPA. The results showed that the three models were statistically significant with each having p-value < 0.001. Response Surface Model provided a better fit in terms of the R^2=47.0% and the RMSE =0.320. The combination of grades that maximizes and minimizes first year CGPA were A1 in all the three subjects and A2 in mathematics, C6 in physics and chemistry, respectively. The results of this work suggested that a large percentage of extraneous factors is affecting the CGPA of first year engineering students in this university due to the relatively small values of the coefficient of determination returned by the three models.

APA CITATION

Nuga, O. A., Adekola, T. O. & Abdulhamid, A. Z. (2024). Assessing the impact of Ordinary Level Grades on the Cumulative Grade Point Average of First Year University Students (A Factorial Design Approach) *East African Journal of Education Studies*, *7*(1), 144-154. https://doi.org/10.37284/eajes.7.1.1733.

CHICAGO CITATION

Nuga, Oluwole A. Tolulope O. Adekola and Abba Zakirai Abdulhamid. 2024. "Assessing the impact of Ordinary Level Grades on the Cumulative Grade Point Average of First Year University Students (A Factorial Design Approach)". *East African Journal of Education Studies* 7(1), 144-154. https://doi.org/10.37284/eajes.7.1.1733

HARVARD CITATION

Nuga, O. A., Adekola, T. O. & Abdulhamid, A. Z. (2024) "Assessing the impact of Ordinary Level Grades on the Cumulative Grade Point Average of First Year University Students (A Factorial Design Approach)", *East African Journal of Education Studies*, 7(1), pp. 144-154. doi: 10.37284/eajes.7.1.1733.

IEEE CITATION
O. A., Nuga, T. O., Adekola & A. Z., Abdulhamid "Assessing the impact of Ordinary Level Grades on the Cumulative Grade Point Average of First Year University Students (A Factorial Design Approach)" *EAJES*, vol. 7, no. 1, pp. 144-154, Feb. 2024.

MLA CITATION
Nuga, Oluwole A. Tolulope O. Adekola & Abba Zakirai Abdulhamid. "Assessing the impact of Ordinary Level Grades on the Cumulative Grade Point Average of First Year University Students (A Factorial Design Approach)". *East African Journal of Education Studies*, Vol. 7, no. 1, Feb. 2024, pp. 144-154, doi:10.37284/eajes.7.1.1733

INTRODUCTION

The Cumulative Grade Point Average (CGPA) is a comprehensive measure of a student's academic performance in the university system, it is obtained as a function of all the courses registered and offered by the student. The problem of modelling CGPA vis a viz students' academic performance has been extensively discussed in the literature. Attempts have been made by authors to recognize the factors that influence the CGPA of university students.

Different factors have been suggested for modelling CGPA, some researchers have used cognitive factors such as pre-admission tests, aptitude tests, previous academic scores. Others have used non-cognitive factors like gender, age, ethnicity/race, etc. Some other authors have used a combination of cognitive and non-cognitive factors. In Nigeria, the main cognitive factors are Ordinary level (O' level) examination results and Unified Tertiary Matriculation Examination (UTME) results. The West African Examination Council (WAEC) as well as the National Examination Council (NECO) are responsible for conducting O' level examinations while the Joint Admission and Matriculation Board (JAMB) is responsible for conducting UTME. Afolabi et al., (2007) used the combination of O' level scores in four subjects and UTME scores of the first-year medical students as a cognitive factor. The relationship of the combined scores with the corresponding CGPA was assessed. Pearson correlation coefficient and student t-test were the statistical techniques used. Their results revealed a positive correlation between the combined scores and CGPA of first-year medical students. Aru et al. (2010) identified both cognitive and non-cognitive factors as key variables for predicting the CGPA of university students from the first year to the fourth year. They transform these variables to forms suitable for an adaptive system coding, applied artificial neural and thereafter suggested an alternative admission criterion that will not only consider the UTME and O' level scores but other factors such as family background, parental income, type of primary and secondary school attended, parents' educational status and family size.

Kolajo and Kolajo (2015) also used several cognitive and non-cognitive factors as an input variable for the Multilayer Perceptron Topology model of the artificial neural network. The model was deployed and train using data of final year University students. Mashael et al. (2016) used previous grades obtained by students in all courses to develop a predictive model for the final GPA of university students. The method applied was the J48 decision tree algorithm. Azeez et al. (2018) predicted students graduating class of degree using data mining techniques such as Classification and Regression Tree (CART). The authors also designed a novel algorithm called Difference Level (DL) which works by adding together the differences in grade point average of each level. This total difference is thereafter subtracted from the result of penultimate semester that gives a predicted graduating CGPA.

Fagoyinbo et al. (2014) classified CGPA of final year Polytechnic students to those less than 2.5 and those more than 2.5, identified the number of female students and male students in each category of CGPA, and thereafter applied the logistic regression to estimate the odds ratio of graduating with a particular category of CGPA for both male and female students. Owolabi et al. (2016) predicted the final CGPA of accounting students in a Nigerian university using the first year CGPA as the predictor variable of a simple linear regression model. The results indicated

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

that 100 level CGPA has a positive effect on final year CGPA; the coefficient of determination also shows that a large proportion of the final Year CGPA is explained by the first-year CGPA. A novel class of multivariate linear regression models was proposed by Huang and Fang (2013) to estimate students' final scores in engineering dynamics using prerequisite courses in calculus Physics and engineering statics. The results show that the models have high average prediction accuracy suggesting that prerequisite courses are good predictors of courses in engineering dynamics.

This present work examined the effect of O' level grades in Mathematics, Physics and Chemistry on the CGPA of first-year engineering students of a private university in southwestern Nigeria using a six level of a three-factorial design (a 6^3 full factorial design. The choice of the three subjects was due to the fact that virtually all the courses taken in first year by engineering students require a solid background in these three subjects.

O' level results are released as categorical variables which are classes of grades even though they are quantitative in their crude form. Students will never know their exact score in these exams but are only aware of the range in which their scores fall into. As an example, a student that got a grade of C4 only knows that he has a score between 60-64 and does not know exactly what his score is. This scenario means that O' level grades cannot be used directly as a quantitative predictor variable for regression modelling. Some authors such as Oyebola (2006) and Adeniyi et.al. (2010) in trying to achieve the objectives of their research converted O' level grades into quantitative variables by assigning A1= 5, B2&B3=4, C4= 3, C5= 2, C6= 1. Kolawole et.al. (2011) used a similar conversion method (A1= 9, B2= 8, B3=7, C4= 6, C5= 5, C6= 4, D7 = 3, D8= 2, F9= 1) in their study of the effect of O' level results on CGPA of chemistry students. These conversion methods if used in regression analysis with more than one O' level grades as predictor variables will generate non-orthogonal models (models in which parameters cannot be estimated independently). In this work, O' level grades were converted using the factorial design approach which allow the scaling of factor levels thereby converting them to continuous predictor variables that can be used for estimating orthogonal models.

Researchers such as Willingham (1985) discovered that O' level grades are highly correlated with first-year CGPA. Also, results from the work of Ishitani & Desjardins (2002) showed a negative correlation between first-year CGPA and the dropout rate of undergraduate university students. i.e., the higher the first-year CGPA, the lower the rate of dropout from the university. Adelman (1999) detailed that having first-year grades in the top half increases the chance of degree completion two or three times over students with grades in the bottom half. Hence the importance of a study on factors that influence the first-year CGPA. Data on the CGPA of 432 first-year students in engineering and the corresponding O' level grades in these subjects was obtained from the exams and records unit of the private university using stratified random sampling. The sample selected from a population of students admitted into the College of Engineering between 2016 and 2020. The population was stratified into $6^3 = 216$ possible factor levels combinations of O' level grades and two students were then randomly selected from each category. This approach will help to develop three orthogonal linear models for comparisons in order to obtain the model that best explained the relationship between CGPA of first-year university students and O' level grades. The extent to which O' level grades influence first-year CGPA can be determined as well as the combination of O' level grades that maximizes first-year CGPA of the university students.

Grading Systems

The two examination bodies (WAEC & NECO) use a grading system where the final marks of students are categorized into classes of grades. The scores obtained which is a continuous

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

variable are classified into appropriate grades thereby turning them into a form of categorical variables. The classification of the grading system for O' level examination is displayed on *Table 1*. As stated earlier, CGPA is the standard grading system in the university. Scores obtained by student in each registered course is summarized into different classes of grades as shown on *Table 2*.

Table 1: O' level grading system

Scores (%)	Grades	Interpretation
75-100	A1	Excellent
70-74	B2	Very Good
65-69	B3	Good
60-64	C4	Credit
55-59	C5	Credit
50-54	C6	Credit
45-49	D7	Pass
40-44	E8	Pass
0-39	F9	Fail

Table 2: University CGPA classification

Classes	CGPA
First Class	4.5-5.0
Second Class, (Upper Division)	3.5-4.49
Second Class, (Lower Division)	2.5-3.49
Third Class	1.5-2.49

METHODOLOGY

Factorial Design

Factorial designs are used extensively in experimental / observational research involving so many factors or input variables. It is particularly useful in situation where the aim is to examine the combined effect of these input variables on the dependent variable. Factors in experimental studies are variables that can be controlled by the experimenter (Goos and Jones 2011). In observational study such as the one used in this research, factors are not controlled by the experimenter as they are chosen by the subject or the environment inflict it on them. Factors are categorical or continuous. In continuous factors, only a chosen number of levels within the interval of interest are used to examine the effect on response variable.

Factorial designs are classified by the number of factor levels. The most common classes of factorial designs in experimental study are the 2^k and 3^k factorial design i.e., 2^k indicate 2 levels of a k factorial design while 3^k indicate 3 levels of a k factorial design. However, there are other forms of factorial design. Factorial design can either be full of fractional. Full factorial designs are those that utilize all factor level combination from factorial design and fractional design only utilizes a fraction of the entire factor level combination. As an example, consider a 2^2 factorial design that can only be studied at a maximum of four factor level combination. Assume that the levels of factor A and B are represented as low and high, if the experimenter included all the four-level combination possible in the experiment, then we say that a full factorial design has been used. On the other hand, if less than four level combination is used, then a fractional factorial has been used.

This work used a duplicate of a 6^3 full factorial design (six level of a three-factorial design). The three factors under study are the O' level grades of students in Chemistry, Mathematics and Physics. This approach will convert categorical predictor variables (O' level grades) to continuous predictor variable and use same to estimate three statistical models for comparison.

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

Scaling of the Factor Levels to Continuous Variables

The levels of the three factors chosen for this work are grades between A1 and C6 (the required grades for admission into engineering courses). The idea is to select the interval that can maximize the CGPA of first year students. This work will use the mid-point average of each grade classification as a representation of the continuous variables. Each grade class between A1 and C6 has a class width of 5. It will be assumed in this work that the range of A1 for the categories of student selected is between 75 – 79 in other to have a uniform class width for each class of grades.

The grade classification and their mid-point is shown on *Table 3*.

Table 3: O' level grade classification and mid-point

Grades	Interval	Mid-point	Class width
A1	75 – 79	77	5
B2	70 – 74	72	5
B3	65 – 69	67	5
C4	60 – 64	62	5
C5	55 – 59	57	5
C6	50 – 54	52	5

The common practice is to scale these levels so that they can lie on the interval [-1, +1] Goos and Jones (2011)

The continuous factor levels 52, 57, 62, 67, 72 and 77 are called Unscaled levels and can be scaled.

Let U represent the upper end-point of the unscaled levels and L the lower end-point of the unscaled level. The midpoint M of the unscaled interval [L, U] is

$$M = \frac{L + U}{2}, \qquad M = \frac{52 + 77}{2} = 64.5$$

The half of the range of the interval is

$$\Delta = \frac{\mathrm{U} - \mathrm{L}}{2}, \qquad \Delta = \frac{77 - 52}{2} = 12.5$$

The scale level X_k of a factor with unscaled level L_k is

$$X_k = \frac{L_K - M}{\Delta}$$

Therefore, the lower end point L= C_6 =52 and the upper end point U = A1 = 77 can be scaled as

$$C_6 = \frac{52 - 64.5}{12.5} = -1 \qquad A_1 = \frac{77 - 64.5}{12.5} = +1$$

The four other levels are scaled in the same manner. The scaled values are presented on *Table 4.*

Table 4: Scaled and unscaled level for o' level grades

Grades	Scaled Level	Unscaled Level
A1	+1	77
B2	0.6	72
B3	0.2	67
C4	-0.2	62
C5	-0.6	57
C6	-1	52

There are a number of advantages for using the scaling convention in modelling. First, it allows for directs comparison of the sizes of effects i.e., assuming the effects of a factor 1 (β_1) is thrice as large in magnitude as the effect of factor 2 (β_2), then the first factor effect of the first factor on

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

the response is thrice as big as the effect of the second factor. The second benefit is that it allows for the development of models containing both main effects and two factor interaction effects without introducing the challenge of correlation (multicollinearity) between these two effects. Specifically stating, for a full factorial design, this scaling convention generate an orthogonal design that guarantees that the effect of all factors in the models (main, interactions, quadratic) can be estimated independently of each other.

Model Selection

To ascertain the most precise model that can best explain the relationship between O' level grades and CGPA. This work will fit three different linear models namely, Main Effect Model (MEM), Interaction Model (IM) and the Response Surface Model (RSM).

Main Effect Model

The MEM contains first-order terms in each factor, for k = 3 factors, the model equation is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i \qquad (1)$$

$\beta_0, \beta_{1,}\ \beta_2$ and β_3 are the parameters of the model and the average effect of each factor on the response, e_i represents the error term., Y_i is the response corresponding to the i^{th} subject, $X_{1,}\ X_2$ and X_3 are the factors.

A significant main factor effect shows that the factor has a linear effect on the response variables i.e., the effect of a particular factor on the response variable is independent of all other factors in the model.

Interaction Effect Model

The IM contains terms in the MEM and an interaction term of the form $X_i X_j$, for a model with
$k = 3$ factors and 2 factor interaction terms, the model equation can be written as;

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_3 X_{3i} + \beta_{12} X_{1i} X_{2i} + \ldots + \beta_{23} X_{2i} X_{3i} + e_i \qquad (2)$$

A significant interaction effect indicates that the impact of a factor on the response depends on the level of the other factor. However, if the result indicates a non-significant interaction effect, then there would be no need to include this term in the model and the researcher can fit the main effect model instead.

Response Surface Models

These models contain quadratic terms of the form X_i^2. The inclusion of a quadratic term is to find optimal settings or factors to maximize the response. It is used to model 'pure quadratic curvature" in the response variable. In essence, the response at the centre level of the factor is not equal to the average of the responses at the extreme levels. The model equation for the quadratic model for $k = 2$ factors is written as;

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_{2i} + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{12} X_2^2 + e_i \qquad (3)$$

A significant quadratic effect reveals that there is "pure quadratic curvature" in the response of interest. That is, the response at the centre level is not equal to the average of the responses at the extreme levels. However, if the results of the analysis reveal a non-significant quadratic effect there would be no need to include these terms in the model as the centre level response is approximately equal to the average at the extreme levels.

General Linear Model (GLM)

In general, the three models described above can be represented in matrix form called the GLM as follows;

$$Y_i = X\beta + e \qquad (4)$$

Y is $n \times 1$ vector of observations, X is $n \times p$ model matrix of constant terms., e is $n \times 1$ vector of error terms, β is $(k + 1 = p) \times 1$ vector of parameters.

To fit the general linear model in equation (3.4) the Ordinary Least Square (OLS) estimation procedure is normal choice if the errors and the

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

independent variables are uncorrelated, with an expectation of zero and equal variances. The OLS estimators are obtained by the minimizing the sum of squares of observation from the expected value. $(e'e = (Y - X\beta)'(Y - X\beta))$. The resulting estimator is

$$(\hat{\beta}) = (X'X)^{-1}X'Y \qquad (5)$$

It is to be noted that the OLS estimators require no assumption about the distributional form of the error term.

The covariance matrix of the estimators is

$$\text{Var}\,(\hat{\beta}) = (X'X)^{-1}\sigma^{2'} \qquad (6)$$

The error variance is estimated using

$$\sigma^2 = \frac{1}{n-p}(Y - X\beta)'(Y - X\beta) \qquad 7)$$

The root mean square error is the square roots of the error variance

The coefficient of determination (R^2) is given by

$$R^2 = \frac{\beta' X' Y}{Y'Y} \qquad (8)$$

Significance Test for Model Parameters

To ascertain whether there is an effect for each factor. The hypothesis of no effect is tested for each of the k parameters in the model.

$H_o: \beta_i = 0$

$H_1: \beta_i \neq 0$ for i = 1 ……k

The test statistic which is a t-distributed random variable is given as

$$t = \frac{\hat{\beta}}{\sqrt{var(\hat{\beta})}} \qquad (9)$$

A rejection of H_0 indicates that there is an effect of the factor otherwise, there is no effect.

RESULT AND DISCUSSION

The results of the main effect, interaction and response surface models are presented on *Table 5, 6 and 7*. The parameter estimates, standard error, student t- statistic and probability value of the t-statistic are presented for each of this model.

The parameter estimates of the main effect model fitted for the first year CGPA presented on *Table 5* show values of 0.20, 0.19 and 0.25 for O' level Mathematics, Physics and Chemistry respectively and a corresponding p-value less than 0.0001 for the three O' level subjects. This indicates that the three subjects have linear effect on first year CGPA. For every one-mark increase in O' level mathematics, first year CGPA is expected to increase by the value of its parameter estimates (0.20) while keeping the factor level of physics and chemistry constant. The same interpretation holds for the effect of physics and chemistry too. The coefficient of determination (R^2) for the main effect model is 0.364, signifying that 36.4% of first year CGPA variation is accounted for by the main effect model. The Root Mean Square Error (RMSE) value of 0.348 is an estimate of the average differences between the actual or observed and the predicted CGPA.

The parameter estimates of the interaction effect model fitted for the first year CGPA presented on *Table 6* show values of 0.20, 0.19 and 0.25 for the main effect of O' level Mathematics, Physics and Chemistry respectively; values of 0.23, 0.12, 0.04 and 0.17 for the interaction effect of Mathematics*Physics, Mathematics*Chemistry, Physics*Chemistry and Mathematics*Physics*Chemistry. All the two factor interaction effects and three factor interaction effect are significant except Physics*Chemistry. The p-values of all significant effects are not more than 0.0006. Significant interaction effect indicates that the impact of a factor on the response depend on the level of one or more factor. A significant interaction effect of Mathematics*Physics indicates that the impact of mathematics on first year CGPA depends on the level of physics and vice versa. The average effect of mathematics on first year CGPA (0.20) is expected to increase by 0.23(the value of the interaction effect Mathematics*Physics). Likewise, the average effect of physics on first year CGPA (0.20) is expected to increase by 0.23(the value of the interaction effect Mathematics*Physics). The

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

same interpretation holds for all other significant two factor interaction effect. The coefficient of determination (R^2) for the interaction effect model is 0.458, signifying that 45.8% of first year CGPA variation is accounted for by the interaction effect model. The RMSE value of 0.323 is an estimate of the average differences between the actual or observed and the predicted CGPA.

The parameter estimates of the response surface model fitted for the first year CGPA presented on *Table 6* show the same estimated values as earlier presented for the main effect terms and interaction terms. The quadratic effects are the additional terms introduced into the model (Mathematics× Mathematics, Physics × Physics, Chemistry × Chemistry). The results show that the quadratic effects of mathematics and physics are significant with values 0.12 and 0.13 respectively, the p-values of significant effect are less than 0.002. However, the quadratic effect of chemistry is not significant as it has a p-value of 0.65. A significant quadratic effect for mathematics reveals that the effect of O' level mathematics on first year CGPA is not linear. The same interpretation holds for the quadratic effect of physics. The coefficient of determination (R^2) for the response surface model is 0.469, signifying that 46.9% of the first year CGPA variation is accounted for by the response surface model. The RMSE value of 0.320 is an estimate of the average differences between the actual or observed and the predicted CGPA.

The graph of the residual against the predicted CGPA for the three models are displayed on *Figures 1, 2 and 3*. It can be observed from the three plots that some points are far from the line of zero residual. A plot with all points on the zero residual indicates that the model estimated is 100% precise and the farther points are from this line, the higher prediction variance and hence lower the prediction capability of the models. In general, the response surface model provided a better fit for the first year CGPA than the other two models since the model have a higher R^2 value and a slightly smaller RMSE value. Equation (10) shows the estimated model that best explain the relationship between O' level grades and CGPA.

Using the model, the O' level grades that maximizes CGPA are A1 in Mathematics, Physics and Chemistry while the grades that minimizes CGPA are A2 in Mathematics, C6 in Physics and C6 in Chemistry. The summary is presented in *Table 8*.

$$Y_i = 3.11 + 0.20X_1 + 0.20X_2 + 0.26X_3 + 0.23X_1X_2$$

$$+0.12X_1X_3 + 0.12X_1^2 + 0.13X_2^2 \qquad (10)$$

where X_1: Mathematics, X_2: Physics, X_3: Chemistry

Table 5: Parameter estimates of the main effect model (First Year CGPA)

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.2390278	0.016749	193.39	<.0001*
Maths	0.2038889	0.024518	8.32	<.0001*
Physics	0.198631	0.024518	8.10	<.0001*
Chemistry	0.2578175	0.024518	10.52	<.0001*
$R^2 = 36.4\%$, $Adj\ R^2 = 35.9\%$				

Table 6: Parameter estimates of the interaction effect model (First Year CGPA)

Term	Estimate	Std Error	t Ratio	Prob>|t|

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

Intercept	3.2390278	0.015542	208.40	<.0001*
Maths	0.2038889	0.022752	8.96	<.0001*
Physics	0.198631	0.022752	8.73	<.0001*
Chem	0.2578175	0.022752	11.33	<.0001*
Maths*Physics	0.2261565	0.033305	6.79	<.0001*
Physics*Chem	0.035085	0.033305	1.05	0.2927
Maths* Chem	0.1239541	0.033305	3.72	0.0002*
Maths*Phy*Chem	0.1685897	0.048754	3.46	0.0006*
$R^2 = 45.8\%, \quad Adj\ R^2 = 44.8\%$				

Table 7: Parameter estimates of the response surface model (First Year CGPA)

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.113138	0.034798	89.46	<.0001*
Maths	0.2038889	0.022553	9.04	<.0001*
Physics	0.198631	0.022553	8.81	<.0001*
Chem	0.2578175	0.022553	11.43	<.0001*
Maths*Maths	0.1184431	0.038602	3.07	0.0023*
Maths*Physics	0.2261565	0.033014	6.85	<.0001*
Physics*Physics	0.1341611	0.038602	3.48	0.0006*
Maths*Chem	0.1239541	0.033014	3.75	0.0002*
Physics*Chem	0.035085	0.033014	1.06	0.2885
Chem * Chem	0.0171596	0.038602	0.44	0.6569
$R^2 = 47.0\%, \quad Adj\ R^2 = 45.8\%$				

Table 8: Summary of fit for whole model

Model	P-Value	R square	R square (Adjusted)	RMSE
Main Effect	< 0.0001	0.364	0.359	0.348
Interaction Effect	< 0.0001	0.458	0.449	0.323
Response Surface	< 0.0001	0.47	0.458	0.320

Table 9: Maximization and minimization of first year CGPA using the estimated model

	Mathematics	Physics	Chemistry	Predicted CGPA	Actual CGPA
Max	A1	A1	A1	4.37	4.60
Min	A2	C6	C6	2.73	3.3

Figure 1: Residual by predicted plot for main effect model

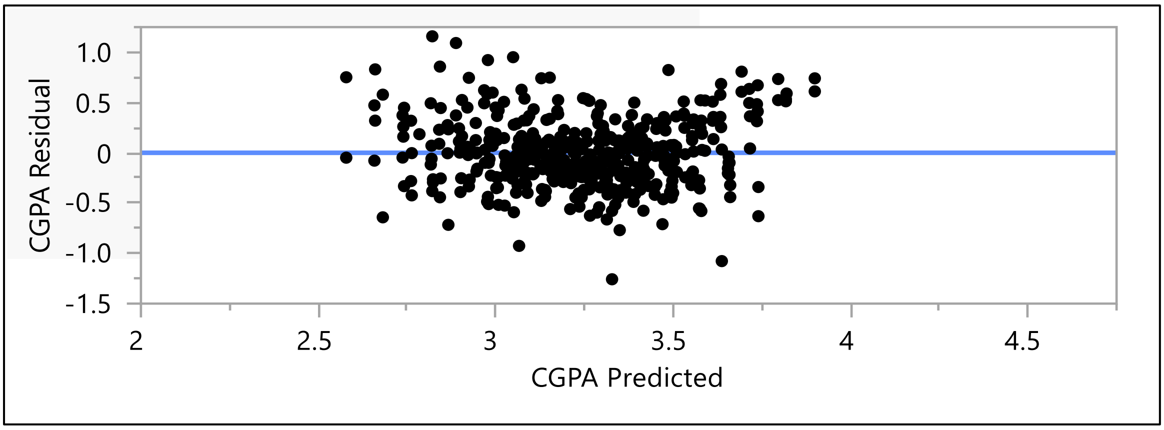


Figure 2: Residual by predicted plot for interaction effect model

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

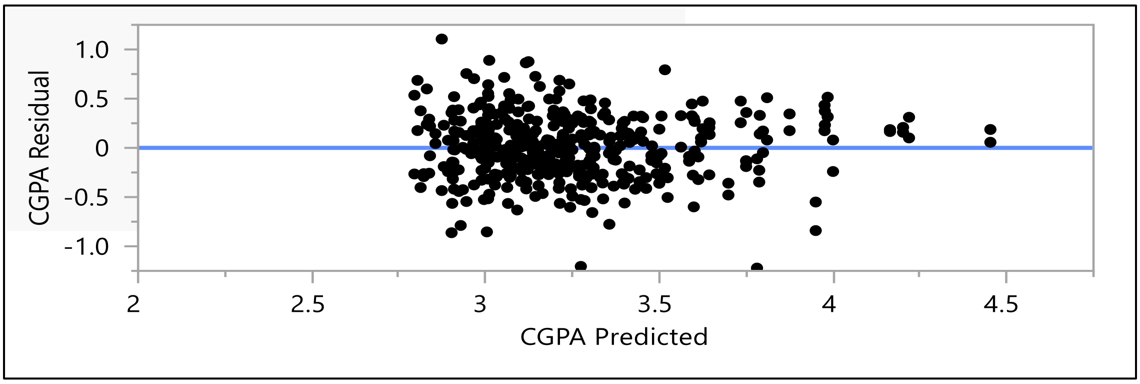
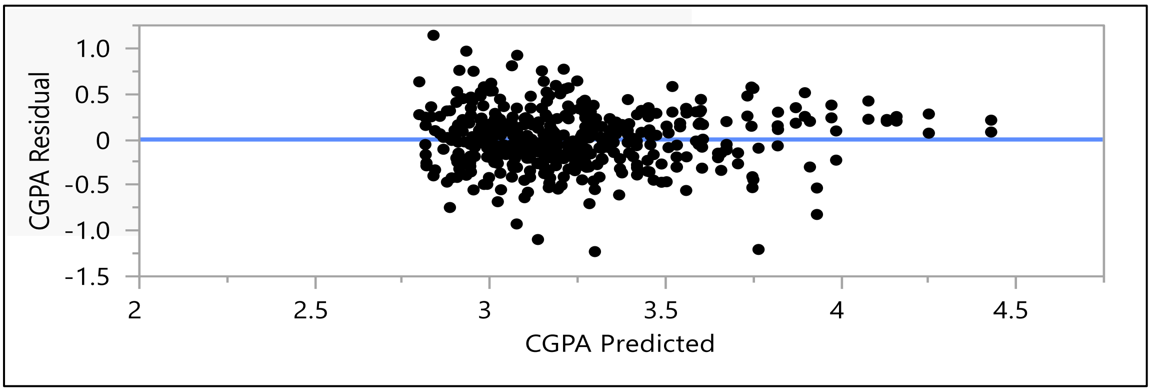


Figure 3: Residual by predicted plot for response surface model



CONCLUSION

The work used the factorial design approach to model first year CGPA of selected private university students. O' level results are released as categorical variables which are classes of grades even though they are continuous in their crude form. Students will never know their exact score in these exams but are only aware of the range in which their scores fall into. This scenario means that O' level grades cannot be used directly as a continuous predictor variable for regression modelling. Available methods in literature for converting grades to continuous variables could lead to estimation of non-orthogonal models thereby making model selection difficult. This challenge was solved by using the factorial design approach that allow the scaling of factor levels thereby converting them to predictor continuous variables that can be used for estimating orthogonal statistical model.

Data on the CGPA of 432 first-year engineering students and the corresponding O' level grades in these subjects was obtained from the exams and records unit of a private University in South-western Nigeria. The data formed a duplicate of 6^3 full factorial design. The Main Effect Model Interaction Effect Model and Response Surface Model were fitted using the OLS Estimators. The results revealed that the three models were statistically significant with all parameters having a positive effect on CGPA. That is, higher O' level grades in the three subjects are expected to cause an increase in Cumulative Grade Point Average. Response Surface Methodology provided a better fit with regards to the Coefficient of Determination and the Root Mean Square Error. However, the Coefficient of Determination value is relatively small suggesting that the three O' level subjects are explaining a relatively small variation in first year CGPA. The prediction variance is also high i.e., the prediction capability of the estimated model is low. Ordinary level grades in the three subjects are not sufficient to precisely model first year CGPA for this category of students.

Article DOI: https://doi.org/10.37284/eajes.7.1.1733

REFERENCES

Adeniyi, O. S., Araoye M. A., Amali E. O., Eru E. U., Ojabo C. O. & Alao O.O (2010) Afr. J. Biomed. Res. 13;189 -195

Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor’s degree attainment. Washington, DC: U.S. Department of Education, Office of Educational Research, and Improvement.

Afolabi, A. O., Mabayoje V. O., Togun V. A. & Oyadeyi A. S. (2007) Selection Criteria for entry into the Medical Programme of Nigerian Universities: African Journal of Biomedical Research, Vol. 10, 203 – 209.

Aru, O. E., Achumba I. E. & Opara F. K. (2010). Assessment of The Admission Criteria that Students’ Academic Performance in Undergraduate Years in a Nigerian University. *Int'l Conf. Artificial Intelligence. CSREA Press. ISBN: 1-60132-438-3.*

Azeez, T.O, Awe, A.C. & Omosebi, P.A. (2018). Predicting Students' Graduating Cumulative Grade Point Average Using Difference Level, Classification and Regression Tree and Linear Regression Algorithm. Journal of Science and Logics in ICT Research. Vol.2, No. 1. 40-47.

Fagoyinbo, I. S., Ajibode I. A., & Olaniran Y.O.A (2014). The Application of Logistic Regression Analysis to the Cumulative Grade Point Average of Graduating Students: A Case Study of Students of Applied Science, Federal Polytechnic, Ilaro. Developing Country Studies. Vol.4, No. 23.

Goos, P. & Jones, B. (2011). Optimal Design of Experiments: A Case-Study Approach. John Wiley and Sons.

Huang, Shaobo & Fang, Ning (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education. 61. 133–145.

Ishitani, Terry & DesJardins, Stephen. (2002). A Longitudinal Investigation of Dropout from College in the United States. Journal of College Student Retention: Research, Theory and Practice. Vol., 4. No 2. 173-201.

Kolajo, T. & Kolajo J. O. (2015). Career Guidance through Admission Procedures in Nigerian Universities Using Artificial Neural Networks. International Journal of Advanced Research in Computer Science and Software Engineering. 5: 9.

Kolawole, E. B., Oginni, O. I. & Fayomi, E. O. (2011). Ordinary Level as Results Predictors of Students' Academic Performance in Chemistry in Nigerian Universities. Educational Research and Reviews Vol. 6(16), pp. 889-892

Mashael, A. Al-Barrak & Muna Al-Razgan (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. International Journal of Information and Education Technology, Vol. 6, No. 7.

Owolabi, F, Oguntunde P., Adetula, D & Fakile, S. (2016). Learning Analytics: Datasets on the academic record of accounting students in a Nigerian University. Data in Brief. 19. 10

Oyebola, D. D. O. (2006) The Importance of „O" level grades in Medical School Admission. The University of Ado-Ekiti Experience. Afr. J. Biomed. Res 9: 15-21.

Willingham, W. W. (1985). Success in college: The role of personal qualities and academic ability. New York: College Entrance Examination Board.

