Journal of the Nigerian Association of Mathematical Physics
***Volume 61* (July – September *2021 Issue*), pp163 – 174**

AN OPTIMAL MODEL FOR STUDENT PERFORMANCES OBTAINED BY APPLICATION OF MULTIPLE REGRESSION AND LINEAR DISCRIMINANT MODELS

Achie Ismailu and Abdul Garba Saddiq

Department of Mathematics and Statistics, Kaduna Polytechnic, Kaduna.

Abstract

Kaduna Polytechnic has no predictive model for student performances and administration. This paper considered two separate multivariate models; Multiple Regression and Linear Discriminant Models. Data was collected from Kaduna Polytechnic Records and classified into purposeful and logical categories for analysis. The SPSS package which has been ratified to enhance volume, speed and accuracy was used to derive the Fisher Linear Discriminant Model. The Generalized Multiple Regression Model was derived by solving normalized system of equations. Both models were used to predict the final CGPA of Kaduna Polytechnic students. It was discovered that the Multiple Regression has some significant prediction power while the Linear Discriminant model classified students' into groups. Both the Multiple Regression and Linear Discriminant Models were necessary for comprehensive prediction of students' grades.

Keywords: Regression, Discriminant, Linear, Multiple, Rank Correlation, Tests of Hypothesis, SPSS, Models, ANOVA.

1.0 Introduction

In Kaduna polytechnic, examinations are conducted using the Semester System where students take a certain number of course units per semester. The results of these exams are graded according to the performance of the students as A, B, C, D, E and F and carry points of 5, 4, 3, 2, 1 and 0 respectively, and classified by Distinction, Upper credit, Lower credit, Pass and Fail. These results are of paramount importance and we can use certain models to predict the final outcome or performance of the students in Kaduna polytechnic at any semester and particularly in the Department of Mathematics, Statistics and Computer Science.

Two methods are studied and analyzed for this thesis. They are the Linear Discriminant Model (LDM) and the Multiple Regression Model (MRM).

1. To use a Fisher's Linear Discriminant Model capable of distinguishing and classifying good and average students.
2. To use a Multiple Regression Model capable of predicting the final grades of students.
3. To compare and contrast the powers of the discriminant model relative to the Multiple Regression Model.

Regression analysis, usually termed **regression**, is used to draw the line of 'best fit' through coordinates on a graph [1]. The techniques to be used will enable a mathematical equation of the straight line of the form $y = mx + c$ which will be deduced for a given set of coordinate values, the line being such that the sum of the deviations of the co-ordinate values from the line is a minimum, i.e. it is the line of 'best fit'.

There are different mathematical models in Multiple Regression Analysis. The basic equation relating to the various variables may be written as:

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + E \qquad (2.0)$$

Multiple Regression Analysis was used to derive a Mathematical relationship between the Cumulative Grade Point Average (CGPA) as the dependent variable (Y) and the results of the students as the independent variable x_i and the associated error term *E*. then X_1, X_2 and X_3 were used to denote relevant raw scores in Statistical Theory, Statistical Inference and Applied General Statistics respectively (in percentage).

In this case, the least square method was used to obtain the estimates of the β's written as b (Regression coefficient) while α is the constant value and the E is the error term.

Corresponding Author**:** Achie I., Email: achieismailu@gmail.com, Tel: +2348028566982

The concept of Regression can be traced as far back as to 1889 which was a reported findings about the relationships between Heights of Fathers and Sons and discovered that tall Father always produced children that are tall while short fathers produced children that are also short. His works and findings apply in modern usage to a function that is employed in statistical prediction [2]. Similar findings were done on the proof and measurement of association between two things led to the idea of Rank Correlation and hence the developments of the Spearman's Rank order correlation [3].
In the more general multiple regression model, there are k independent variables:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + E_i$$

The Normal equations are:

$$\sum_{i=1}^{k}\sum_{k=1}^{k} x_{ik} x_{ik} \beta_k = \sum_{i=1}^{k} x_{ik} Y_i, \quad i = 1, 2, \ldots k$$

$$j = 1, 2, \ldots k$$

For N=K=3 in this case we have:

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e_i$$

The parameter estimates $\alpha,\ \hat{\beta}_1,\ \hat{\beta}_2,\ and\ \hat{\beta}_3\ of\ \alpha,\ \beta_1,\ \beta_2,\ \beta_3$ was obtained through the use of Normal Equations and the computations of relevance vector estimates.

$$\sum Y = \hat{\alpha}_n + \hat{\beta}_1 \sum x_1 + \hat{\beta}_2 \sum x_2 + \hat{\beta}_3 \sum x_3$$

$$\sum X_1 Y = \hat{\alpha} \sum x_1 + \hat{\beta}_1 \sum x_1^2 + \hat{\beta}_2 \sum x_1 x_2 + \hat{\beta}_3 \sum x_1 x_3$$

$$\sum X_2 Y = \hat{\alpha} \sum x_2 + \hat{\beta}_1 \sum x_1 x_2 + \hat{\beta}_2 \sum x_2^2 + \hat{\beta}_3 \sum x_2 x_3$$

$$\sum X_3 Y = \hat{\alpha} \sum x_3 + \hat{\beta}_1 \sum x_1 x_3 + \hat{\beta}_2 \sum x_2 x_3 + \hat{\beta}_3 \sum x_3^3 \tag{2.1}$$

The solutions are more readily obtained using the matrix approach as:

$$Vector\ \ of\ \ coefficien\ \ t\ \ \beta = \hat{\beta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}\ \ Vector\ \ of\ \ parameters \tag{2.2}$$

$$= (X'X)^{-1}(X'Y)$$

$$(X'X) = \begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \sum x_1 x_3 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \sum x_2 x_3 \\ \sum x_3 & \sum x_1 x_3 & \sum x_2 x_3 & \sum x_3^2 \end{pmatrix}$$

$$(X'Y) = \begin{pmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \sum X_3 Y \end{pmatrix}$$

Where:

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix}\ \ Design\ \ matrix$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}\ \ Vector\ \ of\ \ observed\ \ \ values$$

Then the model derived was of the form:

$$Y = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 \qquad (2.3)$$

Discriminant analysis treats the problems of attempting to differentiate between two or more classes of persons or object [4]. It is the most useful when we have obtained a set of measurement on population which we know are different groups.
The major purpose of discriminant analysis is to predict membership into two or more mutually exclusive groups from a set of predictors, when there is no natural ordering on the groups [5]
Typical application of discriminant analysis is testing for separation and equal means of two multivariate data sets, which assume multivariate normality and test assumes equal covariance [6].
The Linear Discriminant Model used is adopted from the independent works which he developed a method for the solution of the two group case known as Linear Discriminant Analysis and this is analogous to a multiple of regression analysis which the dependent variable Y, assumes only two values, each indicating membership in one of the other two groups [7]. It is similar to the methods of analysis developed by [8].
Similar studies produced similar results of Multiple Regression Analysis and the Linear Discriminant Analysis to support the relationship that leads to the prediction and forecasting to infer causal relationships between the independent and dependent variables [9, 10, 11].
Linear Discriminant Analysis (LDA) and Regression Analysis (RA) are widely used for multivariate statistical methods for analysis of data with categorical outcome variables. Both of them are appropriate for the development of linear classification models, i.e. models associated with linear boundaries between the groups [12].
Nevertheless, the two methods differ in their basic idea. While RA makes no assumptions on the distribution of the explanatory data, LDA has been developed for normally distributed explanatory variables. It is therefore reasonable to expect LDA to give better results in the case when the normality assumptions are fulfilled, but in all other situations RA should be more appropriate. The theoretical properties of RA and LDA are thoroughly dealt with in the literature; however the choice of the method is often more related to the field of statistics than to the actual condition of fulfilled assumptions.
While RA is much more general and has a number of theoretical properties, LDA must be the better choice if we know the population is normally distributed. However, in practice, the assumptions are nearly always violated, and we have therefore tried to check the performance of both methods with simulations. This kind of research demands a careful control, so we have decided to study just a few chosen situations, trying to find a logic in the behaviour and then to think about the expansion onto more general cases. We have confined ourselves to compare only the predictive power of the methods.

2.0 Methodology

After getting the data, transform to tabulate it, certain software was used to enhanced volume, speed and accuracy. The data to be used will be in the form:

Y	X_1	X_2	X_3	E

Where:
Y = dependent variable
X_i = the independent variables
E = Error
Let
Y = Final CGPA
X_1 = Raw scores (%) in Statistical Theory
X_2 = Raw scores (%) in Statistical Inference
X_3 = Raw scores (%) in Applied General Statistics
Data was collected through the documentary method; (i.e. secondary source) data was collected directly from records of examination results from the examination office.

3.0 Analysis

3.1 The Linear Discriminant Model

This is adopted from the independence works of [5, 8].

$$Y = X'S^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

Where :

$$X' = \left(X_1 X_2 X_3\right)$$

$\bar{X} = \frac{1}{n}\sum x_i$ (the mean or average of *x*)

The table was then split into two groups for the purpose of the discriminant analysis as follows:

$$Y = X'S^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

Where:

$$X' = \left(X_1 \ X_2 \ X_3\right)$$

$$\bar{X} = \frac{1}{n}\sum xi$$

The table will then be split into two groups for the purpose of the discriminant Analysis as follows:-

Group 1: This group will consist of only students with lower credit and pass (Low performance group)

Group 2: This group will consist of only students with distinction and upper credit (high performance students).

$$\bar{X}_1 = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \end{bmatrix}$$

$n_1 =$ Number of observations in group 1

$\bar{X}_{11}$ = Mean scores in Statistical Theory for Group 1

$\bar{X}_{12}$ = Mean scores in Statistical Inference for Group 1

$\bar{X}_{13}$ = Mean scores in Applied General Statistics for Group 1

$$\bar{X}_2 = \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \end{bmatrix}$$

$n_2 =$ Number of observations in group 2

$\bar{X}_{21}$ = Mean scores in Statistical Theory for Group 2

$\bar{X}_{22}$ = Mean scores in Statistical Inference for Group 2

$\bar{X}_{23}$ = Mean scores in Applied General Statistics for Group 2

Where:

X_1, X_2, and X_3 are courses taken in the first semester and Y is the Final CGPA at graduation.

$$S_1 = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix}$$

$$S_2 = \begin{bmatrix} S_{21} & S_{22} & S_{23} \\ - & - & - \\ - & - & - \end{bmatrix}$$

$$S_{ij} = \frac{\sum \left(x_i - \bar{x}_i\right)\left(x_j - \bar{x}_j\right)}{n}$$

S_1 is variance/covariance matrix of sample of size n_1
S_2 is variance/covariance matrix of sample of size n_2
S = Pooled variance - covariance matrix (since they are from the same population)

$$S = \frac{n_1 s_1 + n_2 s_2}{n_1 + n_2}$$

The Mahalanobis square distance is given by:

$D^2 = \frac{1}{2}\left[\bar{X}_2 + \bar{X}_1\right] S^{-1}\left[\bar{X}_2 - \bar{X}_1\right]^1$ and was used as the test statistic in the analysis.

Decision Rule

$H_O : Classify \quad as \ \pi_2 \ if \ Y \geq D^2$

$H_1 : Classify \quad as \ \pi_1 \ if \ Y < D^2$

Where:

π_2 = Population of exceptionally good students (upper and distinction)

π_1 = Population of low students (low and pass)

Regression models involve the following variables:

(i) The unknown parameters denoted as β; this may be a scalar or a vector of length k.
(ii) The independent variables, x
(iii) The dependent variable, Y

A regression model relates Y to a function of X and β.

$Y \simeq f(X, \beta)$

The approximation is usually formalized as E(Y/X) = f(X,β). To carry out regression analysis, the form of the function f must be specified. Sometimes the form of this function is based on knowledge about the relationship between Y and X that does not rely on the data. If no such knowledge is available, a flexible or convenient form for f is chosen.

Now that the vector of unknown parameters β is of length k, (k=3), it means that N=k data points are observed, and the function f is linear, the equations Y=f(X,β) can be solved exactly rather than approximately. This reduces to solving a set of N equations with N unknowns (the elements of β), which has a unique solution as long as the x are linearly independent. If f is non-linear, a solution may not exist, or many solutions may exist.

In the process of analysis, data was classified into purposeful and logical categories or groups. The first group consisted of only students with lower credit and pass (i.e. low performance students) while the second group consisted of only students with distinction and upper credit (high performance students). The possible categories were considered when plans were made for collecting the data to facilitate analysis. Therefore, the process of analysis was partially concurrent with collection and presentation. Hence, there is the need to, first and foremost, present the data in their original form before extracting the desired analytical tables for the actual data analysis. This will actually reflect the originality of the data collected by the documentary method i.e. from examination results from the examination office.

Table 1: Kaduna Polytechnic Students Examination Records

S/No	Scores (%) in Statistical theory (X_1)	Scores (%) in Statistical inference (X_2)	Scores (%) in Applied general Statistics (X_3)	Final CGPA (Y)	Remark	Performance Group
1	71	40	40	2.66	Lower credit	Group 1
2	50	40	40	2.47	Pass	Group 1
3	57	58	45	3.08	Upper credit	Group 2
4	50	50	40	2.51	Lower credit	Group 1
5	46	44	40	2.61	Lower credit	Group 1
6	50	45	53	2.58	Lower credit	Group 1
7	57	55	40	3.03	Upper credit	Group 2
8	70	86	75	3.72	Distinction	Group 2
9	50	53	55	3.04	Upper credit	Group 2
10	55	67	65	3.01	Upper credit	Group 2
11	43	42	40	2.58	Lower credit	Group 1
12	40	40	45	2.78	Lower credit	Group 1
13	61	42	40	2.52	Lower credit	Group 1
14	40	40	40	2.58	Lower credit	Group 1
15	40	45	40	2.55	Lower credit	Group 1
16	56	75	82	3.53	Distinction	Group 2
17	50	74	78	3.06	Upper credit	Group 2
18	63	45	56	2.79	Lower credit	Group 1
19	40	49	54	2.61	Lower credit	Group 1
20	55	54	40	2.99	Lower credit	Group 1
21	55	58	40	2.77	Lower credit	Group 1
22	60	78	68	3.42	Upper credit	Group 2
23	60	76	88	3.52	Distinction	Group 2
24	50	59	45	2.71	Lower credit	Group 1
25	71	85	76	3.89	Distinction	Group 2
26	50	59	55	2.66	Lower credit	Group 1
27	70	80	74	3.58	Distinction	Group 2
28	69	61	70	3.55	Distinction	Group 2
29	58	56	45	3.06	Upper credit	Group 2
30	71	85	78	3.73	Distinction	Group 2

Results of Regression Analysis

Table 2: SPSS Output Regression Analysis

	Unstandardized Coefficients				Collinearity Statistics	
	B	Std. Error	t	Sig.	Tolerance	VIF
(Constant)	1.111	.174	6.368	.000		
Scores (%) in Statistical theory (X1)	.012	.004	2.995	.006	.633	1.581
Scores (%) in Statistical inference (X2)	.015	.004	3.611	.001	.225	4.440
Scores (%) in Applied general Statistics (X3)	.006	.004	1.688	.103	.262	3.817

a. Dependent Variable: Final CGPA (Y)

Hence, the required regression model is as follows:

$$\hat{Y}_i = 1.111 + 0.012X_{1i} + 0.015X_{2i} + 0.006X_{3i}$$

From all the p-values in Table 2 above, all the coefficients are relevant to the model.

Table 3: The Regression Model Summary

	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
	.933[a]	*.870*	*.855*	*.16557*	*2.609*

a. Predictors: (Constant), Scores (%) in Applied general Statistics (X3), Scores (%) in Statistical theory (X1), Scores (%) in Statistical inference (X2)

b. Dependent Variable: Final CGPA (Y)

From Table 3, Since R is large; this confirms the suitability as well as the goodness of fit of the regression model obtained.

3.3 Collinearity Diagnostics

Table 4: The Regression Collinearity Diagnostics.

	Eigenvalue	Condition Index	Variance Proportions			
			(Constant)	Scores (%) in Statistical theory (X1)	Scores (%) in Statistical inference (X2)	Scores (%) in Applied general Statistics (X3)
1	*3.927*	*1.000*	*.00*	*.00*	*.00*	*.00*
2	*.050*	*8.890*	*.24*	*.04*	*.04*	*.13*
3	*.014*	*16.512*	*.68*	*.70*	*.03*	*.16*
4	*.009*	*21.337*	*.07*	*.26*	*.92*	*.71*

a. Dependent Variable: Final CGPA (Y)

<u>Eigen Values and Condition Index Multi-Collinearity Test</u>

As particularly used for this study, from the eigenvalues of the $(X'X)$ matrix, we can respectively derive what is known as the condition number K and condition index $CI = \sqrt{K}$ as follows:

$$K = \frac{\max imum \quad eigenvalue}{\min imum \quad eigenvalue} \quad \text{and} \quad CI = \sqrt{K} = \sqrt{\frac{\max imum \quad eigenvalue}{\min imum \quad eigenvalue}}$$

If K is between 100 and 100, there is a moderate to strong multicollinearity and if it exceeds 1000, there is severe multicollinearity. Alternatively, if the $CI = \sqrt{K}$ is between 10 and 30 there is moderate to strong multicollinearity and if it exceeds 30, there is severe multicollinearity. Some statisticians believe that the condition index $CI = \sqrt{K}$ is the best available multicollinearity diagnostics for all kinds of regression models.

From Table 4 above, since the entire condition index values are less than 30, there is no evidence of severe multicollinearity among the explanatory variables used for the model.

3.4 Spearman's Rank Correlation

Table 5: The Regression Correlation Matrix

Correlations		Scores (%) in Statistical theory (X1)	Scores (%) in Statistical inference (X2)	in Applied general Statistics (X3)	Final CGPA (Y)
Scores (%) in Statistical theory (X1)		*1*	*.606***	*.514***	*.713***
Scores (%) in Statistical		*.606***	*1*	*.859***	*.901***
Scores (%) in Applied		*.514***	*.859***	*1*	*.833***
Final CGPA (Y)		*.713***	*.901***	*.833***	*1*

**. Correlation is significant at the 0.01 level (2-tailed).

From Table 5 above, the pairwise correlations are all positive and strong indicating that each of the three selected courses is positively and strongly contributing to the students' CGPA.

Table 6 : The Spearman's Correlation Coefficients.

Spearman's rho			Absolute Residual
	Scores (%) in Statistical theory (X1)	Correlation Coefficient	-.182
		Sig. (2-tailed)	.335
	Scores (%) in Statistical inference (X2)	Correlation Coefficient	-.205
		Sig. (2-tailed)	.276
	Scores (%) in Applied general Statistics (X3)	Correlation Coefficient	-.162
		Sig. (2-tailed)	.392
			30

From Table 6 above, the pairwise correlations in table 5 are all significant indicating that each of the three selected courses is significantly contributing to the students' CGPA.

The Regression Models

From Table 1 above, the matrix of original exam scores that was used to obtain the regression coefficients and thereafter the models for predicting students' final CGPA is as follows:

$$\mathbf{X} = \begin{pmatrix} 1 & 71 & 40 & 40 \\ 1 & 50 & 40 & 40 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & 71 & 85 & 78 \end{pmatrix}$$

$$\mathbf{Y} = \begin{pmatrix} 2.66 \\ 2.47 \\ . \\ . \\ . \\ 3.73 \end{pmatrix}$$

$$\mathbf{X'X} = \begin{pmatrix} n & \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{2i} & \sum_{i=1}^{n} X_{3i} \\ \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{1i}^2 & \sum_{i=1}^{n} X_{1i}X_{2i} & \sum_{i=1}^{n} X_{1i}X_{3i} \\ \sum_{i=1}^{n} X_{2i} & \sum_{i=1}^{n} X_{1i}X_{2i} & \sum_{i=1}^{n} X_{2i}^2 & \sum_{i=1}^{n} X_{2i}X_{3i} \\ \sum_{i=1}^{n} X_{3i} & \sum_{i=1}^{n} X_{1i}X_{3i} & \sum_{i=1}^{n} X_{2i}X_{3i} & \sum_{i=1}^{n} X_{3i}^2 \end{pmatrix} = \begin{pmatrix} 30 & 1658 & 1741 & 1647 \\ 1658 & 94512 & 98900 & 93404 \\ 1741 & 98900 & 107833 & 101697 \\ 1647 & 93404 & 101695 & 97873 \end{pmatrix}$$

$$(\mathbf{X'X})^{-1} = \begin{pmatrix} 1.109 & -0.018 & -2.262\times10^{-5} & -1.665\times10^{-3} \\ -0.018 & 5.448\times10^{-4} & -2.246\times10^{-4} & 9.09\times10^{-6} \\ -2.262\times10^{-5} & -2.246\times10^{-4} & 6.528\times10^{-4} & -4.637\times10^{-4} \\ -1.666\times10^{-3} & 9.09\times10^{-6} & -4.637\times10^{-4} & -2.262\times10^{-5} \end{pmatrix}$$

$$\mathbf{X'Y} = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{1i}Y_i \\ \sum_{i=1}^{n} X_{2i}Y_i \\ \sum_{i=1}^{n} X_{3i}Y_i \end{pmatrix} = \begin{pmatrix} 89.59 \\ 5040.80 \\ 5372.86 \\ 5086.68 \end{pmatrix}$$

The summary of the computations were obtained through the SPSS tables; thus

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Y}\right) = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 1.111 \\ 0.012 \\ 0.015 \\ 0.006 \end{pmatrix}$$

Therefore, the required regression model is as follows:

$$\hat{\mathbf{Y}}_{\mathbf{i}} = \mathbf{1.111} + \mathbf{0.012X}_{\mathbf{1i}} + \mathbf{0.015X}_{\mathbf{2i}} + \mathbf{0.006X}_{\mathbf{3i}}$$

3.5 Predicting Students' CGPA Using the Regression Model

Consider our regression model as given below:

$$\hat{\mathbf{Y}}_{\mathbf{i}} = \mathbf{1.111} + \mathbf{0.012X}_{\mathbf{1i}} + \mathbf{0.015X}_{\mathbf{2i}} + \mathbf{0.006X}_{\mathbf{3i}}$$

It is pertinent to use the regression model to predict the CGPA of students using the regression model and the following table containing the scores of ten students as follows:

$$\hat{\mathbf{Y}}_{\mathbf{1}} = \mathbf{1.111} + \mathbf{0.012(50)} + \mathbf{0.015(65)} + \mathbf{0.006(55)} = \mathbf{3.02}$$

$$\hat{\mathbf{Y}}_{\mathbf{2}} = \mathbf{1.111} + \mathbf{0.012(50)} + \mathbf{0.015(49)} + \mathbf{0.006(45)} = \mathbf{2.72}$$

$$\hat{\mathbf{Y}}_{\mathbf{3}} = \mathbf{1.111} + \mathbf{0.012(46)} + \mathbf{0.015(40)} + \mathbf{0.006(40)} = \mathbf{2.52}$$

$$\hat{\mathbf{Y}}_{\mathbf{4}} = \mathbf{1.111} + \mathbf{0.012(50)} + \mathbf{0.015(53)} + \mathbf{0.006(55)} = \mathbf{2.84}$$

$$\hat{\mathbf{Y}}_{\mathbf{5}} = \mathbf{1.111} + \mathbf{0.012(47)} + \mathbf{0.015(45)} + \mathbf{0.006(45)} = \mathbf{2.62}$$

$$\hat{\mathbf{Y}}_{\mathbf{6}} = \mathbf{1.111} + \mathbf{0.012(42)} + \mathbf{0.015(49)} + \mathbf{0.006(40)} = \mathbf{2.59}$$

$$\hat{\mathbf{Y}}_{\mathbf{7}} = \mathbf{1.111} + \mathbf{0.012(55)} + \mathbf{0.015(67)} + \mathbf{0.006(65)} = \mathbf{3.17}$$

$$\hat{\mathbf{Y}}_{\mathbf{8}} = \mathbf{1.111} + \mathbf{0.012(56)} + \mathbf{0.015(75)} + \mathbf{0.006(82)} = \mathbf{3.40}$$

$$\hat{\mathbf{Y}}_{\mathbf{9}} = \mathbf{1.111} + \mathbf{0.012(40)} + \mathbf{0.015(40)} + \mathbf{0.006(40)} = \mathbf{2.43}$$

$$\hat{\mathbf{Y}}_{\mathbf{10}} = \mathbf{1.111} + \mathbf{0.012(50)} + \mathbf{0.015(74)} + \mathbf{0.006(78)} = \mathbf{3.29}$$

Table 7: The Scores of Ten Students

S/No	Statistical Theory (X_1) %	Statistical Inference (X_2) %	Applied General Statistics (X_3) %	Final CGPA	Actual Grade	Predicted CGPA	Predicted Grade
1	50	65	55	2.98	LC	3.02	UC
2	50	49	45	2.59	LC	2.72	LC
3	46	40	40	2.52	LC	2.52	LC
4	50	53	55	3.04	UC	2.84	LC
5	47	45	45	2.41	PASS	2.62	LC
6	42	49	40	2.73	LC	2.59	LC
7	55	67	65	3.01	UC	3.17	UC
8	56	75	82	3.53	DIST	3.40	UC
9	40	40	40	2.50	LC	2.43	PASS
10	50	74	78	3.06	UC	3.29	UC

The mean absolute deviation MAD for the actual and predicted CGPA is obtained as follows;

$$MAD = \frac{\sum_{i=1}^{n}\left|e_i\right|}{n} = \frac{1.31}{10} = 0.13$$

Similarly, the coefficient of mean absolute deviation MAD for the actual and predicted CGPA is obtained as follows;

$$\therefore Coefficien\ t\ \ of\ MAD\ = \frac{MAD}{\bar{Y}} \times 100\ = \frac{0.13}{2.84} \times 100\ = 4.58\,\%$$

3.6 The (Fishers') Linear Discriminant Model

<u>Group 1</u>

The mean vector and dispersion matrix for group 1 are as follows:

$$\bar{\mathbf{X}}_1 = \begin{pmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \end{pmatrix} = \begin{pmatrix} 50.25 \\ 47.00 \\ 44.25 \end{pmatrix}$$

$$\mathbf{S}_1 = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{22} \end{pmatrix} = \begin{pmatrix} 77.81 & 3.06 & 0.063 \\ 3.06 & 46.13 & 10.94 \\ 0.063 & 10.94 & 37.94 \end{pmatrix}$$

<u>Group 2</u>

The mean $\bar{\mathbf{X}}_2 = \begin{pmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \end{pmatrix} = \begin{pmatrix} 61.00 \\ 70.64 \\ 67.07 \end{pmatrix}$ vector and dispersion matrix for group 2 are as follows:

$$\mathbf{S}_2 = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{22} \end{pmatrix} = \begin{pmatrix} 55.14 & 52.43 & 39.07 \\ 52.43 & 134.66 & 136.45 \\ 39.07 & 136.45 & 211.21 \end{pmatrix}$$

$$\bar{\mathbf{X}}_2 + \bar{\mathbf{X}}_1 = \begin{pmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \end{pmatrix} + \begin{pmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \end{pmatrix} = \begin{pmatrix} 61.00 \\ 70.64 \\ 67.07 \end{pmatrix} + \begin{pmatrix} 50.25 \\ 47.00 \\ 44.25 \end{pmatrix} = \begin{pmatrix} 111.25 \\ 117.64 \\ 111.32 \end{pmatrix}$$

$$n_1 = 16$$

$$n_2 = 14$$

The pooled variance-covariance matrix is as follows:

$$\mathbf{S} = \frac{n_1\mathbf{S}_1 + n_2\mathbf{S}_2}{n_1 + n_2} = \begin{pmatrix} 67.23 & 26.10 & 18.27 \\ 26.10 & 87.44 & 69.51 \\ 18.27 & 69.51 & 118.80 \end{pmatrix}$$

$$\therefore \mathbf{S}^{-1} = \begin{pmatrix} 67.23 & 26.10 & 18.27 \\ 26.10 & 87.44 & 69.51 \\ 18.27 & 69.51 & 118.80 \end{pmatrix}^{-1} = \begin{pmatrix} 0.0169 & -0.0056 & 0.00066 \\ -0.0056 & 0.023 & -0.0127 \\ 0.00066 & -0.0127 & 0.0158 \end{pmatrix}$$

The Mahalanobis square distance is computed as follows:

$$\mathbf{D}^2 = \tfrac{1}{2}\left(\underline{\mathbf{X}}_2 + \underline{\mathbf{X}}_1\right)' \mathbf{S}^{-1}\left(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1\right)$$

$$\therefore \mathbf{D}^2 = \tfrac{1}{2}\begin{pmatrix} 111.25 & 117.64 & 111.32 \end{pmatrix} \begin{pmatrix} 0.0169 & -0.0056 & 0.00066 \\ -0.0056 & 0.023 & -0.0127 \\ 0.00066 & -0.0127 & 0.0158 \end{pmatrix} \begin{pmatrix} 10.75 \\ 23.64 \\ 22.82 \end{pmatrix}$$

$$\therefore \mathbf{D}^2 = \tfrac{1}{2}(37.92) = 18.96$$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

The Fishers linear discriminant model is computed as follows:

$$\mathbf{Y} = \mathbf{X}'\mathbf{S}^{-1}(\overline{\mathbf{X}}_2 - \overline{\mathbf{X}}_1)$$

$$\therefore \mathbf{Y} = (X_1 \quad X_2 \quad X_3)\begin{pmatrix} 0.0169 & -0.0056 & 0.00066 \\ -0.0056 & 0.023 & -0.0127 \\ 0.00066 & -0.0127 & 0.0158 \end{pmatrix}\begin{pmatrix} 10.75 \\ 23.64 \\ 22.82 \end{pmatrix}$$

Finally we obtain the following linear discriminant model:

$$\mathbf{Y} = \mathbf{0.065X}_{\mathbf{1i}} + \mathbf{0.199X}_{\mathbf{2i}} + \mathbf{0.066X}_{\mathbf{3i}}$$

The classification rule that was adopted is as follows:

Classify as Group 2 if $\mathbf{Y} \geq \mathbf{18.96}$

Classify as Group 1 if $\mathbf{Y} < \mathbf{18.96}$

3.7 Classifying Students' CGPA Using the Discriminant Model

It is pertinent to use the discriminant model to classify the CGPA of students. Using the discriminant model:

$$\mathbf{Y} = \mathbf{0.065X}_{\mathbf{1i}} + \mathbf{0.199X}_{\mathbf{2i}} + \mathbf{0.066X}_{\mathbf{3i}}$$

And the classification rule is as follows:

Classify as Group 2 if $\mathbf{Y} \geq \mathbf{18.96}$

Classify as Group 1 if $\mathbf{Y} < \mathbf{18.96}$

To compute estimates or forecasts, consider the discriminant model as given below:

$$\mathbf{Y} = \mathbf{0.065X}_{\mathbf{1i}} + \mathbf{0.199X}_{\mathbf{2i}} + \mathbf{0.066X}_{\mathbf{3i}}$$

That will be used to predict the final grades of students.

Using the discriminant model and the following table containing the scores of ten students as follows:

$\mathbf{Y_1} = \mathbf{0.065(50)} + \mathbf{0.199(65)} + \mathbf{0.066(55)} = \mathbf{19.82} > \mathbf{18.98}$ Classified as group 2

$\mathbf{Y_2} = \mathbf{0.065(50)} + \mathbf{0.199(49)} + \mathbf{0.066(45)} = \mathbf{15.97} < \mathbf{18.98}$ Classified as group 1

$\mathbf{Y_3} = \mathbf{0.065(46)} + \mathbf{0.199(40)} + \mathbf{0.066(40)} = \mathbf{13.59} < \mathbf{18.98}$ Classified as group 1

$\mathbf{Y_4} = \mathbf{0.065(50)} + \mathbf{0.199(53)} + \mathbf{0.066(55)} = \mathbf{17.43} < \mathbf{18.98}$ Classified as group 1

$\mathbf{Y_5} = \mathbf{0.065(47)} + \mathbf{0.199(45)} + \mathbf{0.066(45)} = \mathbf{14.98} < \mathbf{18.98}$ Classified as group 1

$\mathbf{Y_6} = \mathbf{0.065(42)} + \mathbf{0.199(49)} + \mathbf{0.066(40)} = \mathbf{15.12} < \mathbf{18.98}$ Classified as group 1

$\mathbf{Y_7} = \mathbf{0.065(55)} + \mathbf{0.199(67)} + \mathbf{0.066(65)} = \mathbf{21.20} > \mathbf{18.98}$ Classified as group 2

$\mathbf{Y_8} = \mathbf{0.065(56)} + \mathbf{0.199(75)} + \mathbf{0.066(82)} = \mathbf{23.98} > \mathbf{18.98}$ Classified as group 2

$\mathbf{Y_9} = \mathbf{0.065(40)} + \mathbf{0.199(40)} + \mathbf{0.066(40)} = \mathbf{13.20} < \mathbf{18.98}$ Classified as group 1

$\mathbf{Y_{10}} = \mathbf{0.065(50)} + \mathbf{0.199(74)} + \mathbf{0.066(78)} = \mathbf{23.22} > \mathbf{18.98}$ Classified as group

Table 8: Table for final CGPA

S/No	Statistical Theory (X_1) %	Statistical Inference (X_2) %	Applied General Statistics (X_3) %	Final CGPA	Actual Group	Predicted Group
1	50	65	55	2.98	1	2
2	50	49	45	2.59	1	1
3	46	40	40	2.52	1	1
4	50	53	55	3.04	2	1
5	47	45	45	2.41	1	1
6	42	49	40	2.73	1	1
7	55	67	65	3.01	2	2
8	56	75	82	3.53	2	2
9	40	40	40	2.50	1	1
10	50	74	78	3.06	2	2

Hence, the probabilities of misclassifications are obtained as follows:

$$\mathbf{p(group\ 1/group\ 2)} = \tfrac{1}{10}$$

$$\mathbf{p(group\ 2/group\ 1)} = \tfrac{1}{10}$$

4.0 Discussion and Conclusion

4.1 Discussion

This work has achieved three main objectives among others.

1. A linear discriminant model capable of distinguishing and classifying students with good and average (mediocre) grades were objectively built.
2. A multiple linear regression model capable of predicting the final grades of students via their GPAs was also built.
3. The prediction powers of the discriminant model and that of the regression model were empirically compared.

The data collected and used for the study were initially displayed on a summary table for easy access and to facilitate analysis. Hence, the data analysis was conducted in an objective and factual manner using the summary table. Therefore, the multiple linear regression model built was:

$$\hat{\mathbf{Y}}_i = \mathbf{1.111} + \mathbf{0.012X}_{1i} + \mathbf{0.015X}_{2i} + \mathbf{0.006X}_{3i}$$

Moreover, this model has satisfied all the validation and diagnostic tests of goodness of fit, autocorrelation, homoscedasticity and multicollinearity. Consequently, the model was used to predict the CGPA and by extension the final grades of ten students.

Similarly, the Fisher's linear discriminant model built herein was:

$$\mathbf{Y} = \mathbf{0.065X}_{1i} + \mathbf{0.199X}_{2i} + \mathbf{0.066X}_{3i}$$

with a dichotomized classification threshold of $\mathbf{M} = \mathbf{18.96}$ which forms the decision criteria for the model. This model was also used to predict and classify the final grades of ten students.

4.2 Conclusion

Objectively, this paper has applied all the laid down procedures to collect, analyze and interpret educational data for the purpose of evaluating students' performance. The outcome of the analysis has produced two separate mathematical models each capable of independently predicting accurate students' performance through their grades.

First, the multiple linear regression with three predictor variables has a coefficient of determination of $R^2 = 0.870$ or simply 87.0%. This implies that the three predictor variables included in the model can explain at least 87.0% of the changes in the CGPA of students which is good enough for the model.

Secondly, on the other hand, the Fisher's linear discriminant model also has the same three predictor variables with a probability of misclassification from group 1 into group 2 of 0.1and vice versa the probability of misclassification from group 2 into group 1 is also 0.1. Hence, the combined misclassification (error) probability is therefore (0.1)(0.1)=0.01. That is, the linear discriminant model has a 0.99 classification power; in other words, it has a 99% prediction power into either of the two categories of students – those with good grades or average (mediocre) grades.

Conclusively, the Fisher's linear discriminant model should be preferred and used for direct classification into group membership (good or mediocre grades) as it has a higher precision in that regards. On the other hand, there is no direct alternative to multiple regression when the actual CGPAs of the students are desired. Hence, each of the models has a peculiar advantage over the other. Both the linear discriminant and multiple regression models could be used together for effective results.

4.3 Recommendations

In view of the applicability of these findings, the following are therefore recommended:

1. The management of Kaduna Polytechnic and by extension other tertiary institutions of learning should utilize these models for predicting their students' grades to enhance educational planning.
2. Guidance and counseling practitioners should also utilize these models for counseling, evaluation and placement of students.
3. For the predictions of CGPAs only, the multiple regression model with some good predictor variables should always be preferred and used in this kind of situation.
4. For the classification of students' groups (good or mediocre grades) only, the linear discriminant model be preferred and used in this kind of situation.
5. Both the multiple regression and linear discriminant models should be used together when a more comprehensive predictions of students' grades is required.

References

[1] Morison, D. F (1990) *"Multivariate Statistical method".* New York, MC Graw Hill

[2] Galton, F. (1889), Natural inheritance New York: Macmillan, 1889 Hotelling, H. (1931), *The generalization of Students Ratio. Annals of mathematical statistics*, 2m 360-378

[3] Spearman, C. (1904), *The Proof and measurement of association between two things.* American Journal of Psychology, 15, 72-101 (a)

[4] Bolch, B.M. and Huang, C. J. (1974) *Multivariate Statistical Method* Prentice Hall

[5] Press, S. J. and Wilson, S. (1978), "Choosing Between Logistic Regression and Discriminant Analysis"; *Journal of American Statistical Association;* 58, pp 236-244.

[6] Hotelling, H. (1935), *The most predictable criterion*, journal of educational psychology, 26, 139-142

[7] Fisher, R. A. (1936), *The use of Multiple measurements in taxonomic problems.* Annals of Eugenics, 7, 179-188.

[8] Mahalonobis, P. C.(1936), *On the generalized distance in statistics.* Proceedings of the national institute of science, India, 12, 219, 49-55

[9] Lindley, D.V. (1987), "Regression and Correlation Analysis", New Palmgrave: A Dictionary of Economics, V. 4, pp. 120-23

[10] Drapper N, R. and Smith, H. (1998). *Applied Regression Analysis* Wiley Series in Probability and Statistics.

[11] McLachlan, G.J. (2004), *Discriminant Analysis and Statistical Pattern Recognition.* Wiley Interscience. MR1190469. ISBN 0471691155

[12] Pohar, M, Mateja, B, and Sandra, T (2004), *"Comparison of Regression Analysis and Linear Discriminant Analysis: A Simulation Study".* Metodološki zvezki, Vol. 1, No. 1, 2004, 143-161