TITLE PAGE

STATISTICAL MODELING OF STUDENTS ACHIEVEMENT

A CASE STUDY OF GOVERNMENT DAY SECONDARY SCHOOL, MINNA

By

MUSA MUHAMMADU ISA

PGD/MCS/2003/2004/1123

A PROJECT SUBMITED TO THE DEPARTMENT OF MATHEMATICS/ COMPUTER SCIENCE.

IN PARTIAL FUL FILMENT OF THE REQIUEMENTS FOR THE AWARD OF A

POSTGRADUATE DIPLOMA IN COMPUTE SCEINCE.

FEDERAL UNIVERSITY OF TECHNOLOGY,

MINNA, NIGER STATE

November 2004

DECLARATION

I hereby declare that this research work (project), a partial fulfillment of requirement for the award of post graduate diploma of computer science of computer science of, Federal University of Technology, Minna is a product of my research effort.

All sources of information collected for the project are clearly acknowledged by means of references.

MUSA MUHAMMADU ISA

SIGNATURE

DATE

CERTIFICATION

I hereby certify that this is the original work of MR MUSA MUHAMMADU ISA PGD/MCS/2003/2004/1123. Submitted for the award of a postgraduate diploma in computer science of federal university of technology, Minna.

MALL. ISA AUDU
PROJECCT SUPERVISOR

DATE

MR. L.N. EZEAKO
HEAD OF DEPARTMENT

DATE

EXTERNAL EXAMINER

DATE

DEDICATION

This project is dedicated to almighty Allah who has granted me the time and energy to go through my studies successful. It is dedicated to my beloved late mother Hajiya Fatima Kaka. May her humble soul rest in perfect peace, Ameen.

ACKNOWLEDGEMENT

MY deep gratitude and appreciation to Almighty Allah who has been the source of my strength and health throughout the pursuit of my PGD. in computer science.

I am full of gratitude to my beloved parents whom their supports have helped me ever before I exist. in this line, I acknowledge the assistance of my brother Dr Mahmud and his family whom their support both morally and financially are immeasurable.

My sincere appreciation goes to my supervisor Mallam ISA AUDU who despite his tight schedule was always patient wit me, reading through the entire manuscript and made many valuable suggestions and corrections which p improved the style and quality of my work. Sincere gratitude is due to the H.O.D Maths/Computer science department MR L.N EZEAKO, DR AKINWANDE, DR YOMI AIYESIMI, PGD CORDINATOR DR U.Y ABUBAKAR, PRINCE BADMUS, MR JIYA MOHAMMED, MR HAKIMI and others lecturer of the department.

I am most grateful to my brothers; Mallam Abdullah Ndagaa, Aliyu Isa, Tanimu Mamman, my friends and class mates; Mohammed Abdulhakeem, Alhaji Lawal Jibrin, Alhaji Sanusi Abubakar, Pharm.Umar Ndagi, Mohammed Abdulkadir, Zak Silas.

Constraints of space and time prevent mention of other friends and classmates whose contribution and assistance I cannot forget but cherish with silent gratitude

May Almighty Allah continue to bless us (Ameen).

ABSTRACT

This project work which is aimed at modeling students achievement was carried out using regression analysis technique. Little introduction concerning case study was made.

Manual computation explanation and computation were carried out and subsequently; computerized method was carried out using statistical package for social sciences (SPSS). Results interpretations and graphs were later made and recommendation and conclusions made.

TABLE OF CONTENT

TITLE PAGE I
DECLARATION II
CERTIFICATION III
DEDICATION IV
ACKNOWLEDGEMENT V
TABLE OF CONTENT VI-VII
ABSTRACT VIII

CHAPTER ONE

GENERAL INTRODUCTION 1-2
1.2 STATISTICS IN EDUCATION 2-3
1.3 CASE STUDY 3-6
1.4 OBJECTIVE OF THE STUDY 7
1.5 SCOPE AND LIMITATION
1.6 DATA 7-8
1.7 LAYOUT OF THE STUDY 8-9

CHAPTER TWO

LITERATURE REVIEW 10
2.1 REGRESSION ANALYSIS 10-11
2.1.1 PROBABILISTIC AND DETERMINISTIC MODELS 12-14
2.1.2 GENERAL FORM OF PROBABILISTIC MODEL 14
2.1.3 A FIRST ORDER (STRAIGHT-LINE) PROBABILISTIC MODEL 14-15
2.1.4 THE LEAST SQUARE APPROACH 15-18
2.2 MULTIPLE REGRESSION 18-19
2.3 ADVANTAGES OF USING REGRESSION ANALYSIS 19
2.4 AREAS OF APPLICATION OF REGRESSION ANAL 19
2.4.1 HEALTH 19
2.4.2 AGRICULTURAL SECTOR 20
2.4.3 BUSINESS SECTOR 20
2.5 RESIDUALS 21
2.6 MODEL TEST OF VARIABILITY 21-22
2.6.1 FITNESS OF MODEL / QUANTITATIVE DATA 22-23
2.6.2 TEST OF UTILITY 24-25

CHAPTER THREE

METHODOLOGY 26
3.1 INTRODUCTION 26
3.2 TECHNIQUES 26-27
3.3 LEAST SQUARE APPROACH 27
3.3.1 LEAST SQUARE COMPUTATION FOR SHEET1 27-28
3.3.2 RESIDUAL COMPUTATION FOR LEAST SQUARE 28-29
3.3.3 LEAST SQUARE COMPUTATION FOR SHEET2 29-30
3.3.4 RESIDUAL COMPUTATION FOR SHEET2 30
3.4 MULTIPLE REGRESSION COMPUTATION 31-33
3.4.1 RESIDUAL FOR MULTIPLE REGRESSION 34
3.4.2 MULTIPLE COEFFICIENT OF DETERMINATION 34
3.4.3 ADJUSTED COEFFICIENT OF DETERMINATION 35
3.4.4 TEST OF HYPOTHESIS 35
3.5 ANALYSIS OF VARIANCE 36-44

CHAPTER FOUR

SOFTWARE COMMAND 45
4.1 HOW TO INPUT DATA INTO THE DATA EDITOR 45-46
4.2 SELECTING A PROCEDURE FROM THE MENU 47-47
4.3SELECTING VARIABLES FOR THE ANALYSIS 47-57

CHAPTER FIVE

SUMMARY AND CONCLUSION 58
5.1 LIMITATION OF THE STUDY 58
5.2 DISCUSSION 58-61
5.3 SUMMARY 61-62
5.4 CONCLUSION 62-63
5.5 RECOMMENDATION 63
REFERENCE 64
APPENDIX

CHAPTER ONE

General Introduction

1.1 Introduction

Just as problems are been expressed in words, figures, etc, so also solutions are expressed. But solutions are generally expressed logically; therefore logical expressions of solutions to problems are called models. Invariable, statistical techniques are unique models applied to problems solving with a view to making meaningful decision. Statistical techniques are applied in making predictions. These predictions though not usually exact but give close to exact solutions to the problems. These techniques can also be used in making estimations and graphical representation of data that eases the interpretation of voluminous set of data at a glance.

This project explains some areas of application of statistical tools employed in this project works that is multiple regression analysis in education. Among these areas is predicting students score based on his/her past performance.

The selected case study, which is Government day Secondary School Minna, has also been explained and hierarchical structure also presented in this chapter.

Objectives of the study have also been high lightened for easy understanding of what the project work is all about. Equally, the scope of the study has been spelt out. The use of data that was used in carrying out the analysis and its type of method of collection was well explained in this chapter. And, lastly a layout of the study was also been presented.

.2 STATISTICS IN EDUCATION

A researcher is asked to decide between two different communities, which should be connected to a national grid. Among factors to be considered is which of the two uses electricity most at night. After several consultations and findings, the researcher comes up with a conclusion that the community with higher number of students should be connected. This judgment may be based on the fact that students make use of light at might fore their assignments and revisions at night.

Through the Federal Office of Statistics, the Federal Government plans for infrastructure to be made available and allocated to each school based on the past population record of the students.

Universities usually have a minimum and maximum number of students it expects to gain admission into the school through JAMB based upon past and present statistics of students in the school. This among other reasons is to avoid over stretching the facilities of the school.

Graphical representation of average performance of various sets of students over a period of time easily conveys the message of the set that performed best and also worst.

Estimates on expected number of graduates for a nation at a particular time can be made based on past record of graduating students. A stable government through its ministry of education determines the work force expected and plans on how to absorb the labor force into its system.

Predictions and forecasting are among the major aspects of statistics study. With particular respect to educational sector, efficiency of different subject teacher can be predicted upon the students' performance. It is possible that the students all performed well in the two different subjects as a result of the two different teachers input. We may still to further to determine whose teaching technique is better. Infact, statistics can be applied in diverse aspects of education apart from the fact that it is a professional discipline of its own

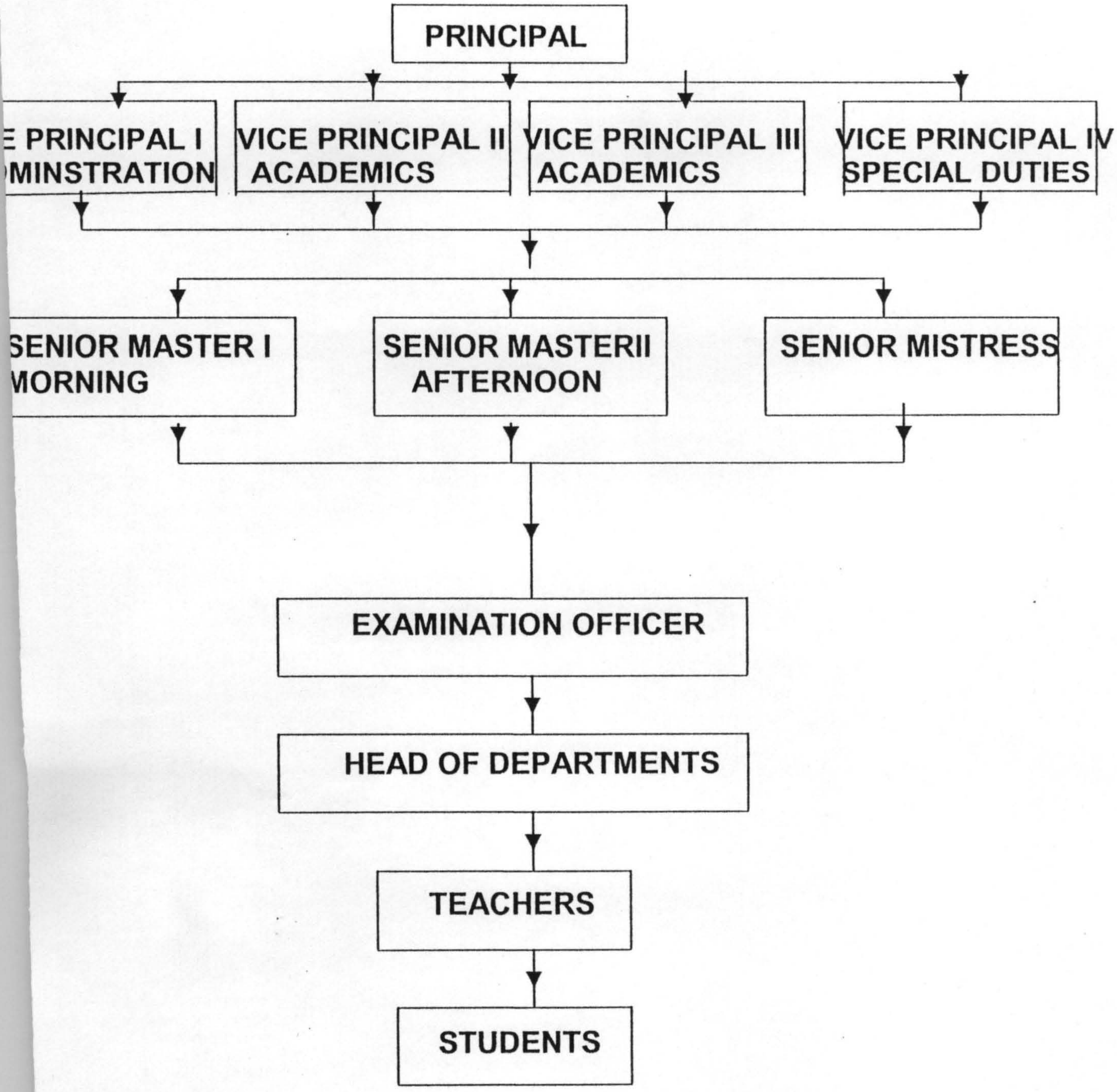
3 CASE STUDY

Government Secondary School Minna located along Bosso Road, Minna is the case study, the school is under the supervision of Niger state ministry of education and been regulated by national policy on education of the Federal Government of Nigeria.

The school was established in November 1979, with Mallam Mammam Batati as its current principal .the principal has the overall say of the school. He ensures that the school is working towards its stated goals. He acts as intermediary between the ministry of education and other staff of the school. Next to the principal are four vice principals, they are vice principal administration I, vice principal academics II, vice principal academics III, and vice principal special duties IV. The vice principal I by name Mallam Mohammad Gbaka receive directives from the principal directly. He is responsible for all administrative activities of the school in conduction with the principal. The two vice principal academics II and III are Mallam Mohammed Gana and Mrs. Eunice Gana respectively. They are in charge of assigning teachers to subjects and classed they are to teach. They make sure that each subject teacher forms his/her lesson note. Lastly, vice principal special duties IV by name Hajiya Aishatu M.Z. Kolo whose duties whose duties are to settle disputes between teachers and students and sometimes teachers and parents of students.

Immediately after the vice-principals are the senior masters and senior mistress. The school has senior master I morning, senior master II, afternoon and senior mistress of the school. The two senior masters I and II are Mallam Sani Mohammed and Mallan Mohammed Adamu respectively. The senior mistress is by name Mrs. Elizabeth Yahya. They act as intermediary between the principal and teachers of the school in dissemination of information. They together prepare timetable of teachers to be on duty for each week.

ORGANOGRAM


PRINCIPAL
E PRINCIPAL I
MINSTRATION
VICE PRINCIPAL II
ACADEMICS
VICE PRINCIPAL III
ACADEMICS
VICE PRINCIPAL IV
SPECIAL DUTIES
SENIOR MASTER I
MORNING
SENIOR MASTERII
AFTERNOON
SENIOR MISTRESS
EXAMINATION OFFICER
HEAD OF DEPARTMENTS
TEACHERS
STUDENTS

OBJECTIVES OF THE STUDY

i) To formulate statistical model of students achievement

ii) Estimate the parameters in models

ii) Fit the model and use the model for prediction

iii) Screen variables to determine which have significant effect on the scores

iv) Arrive at the most effective prediction equation.

.5 SCOPES AND LIMITATION OF THE STUDY

i) The data used are for only two classes which are science class and social science classes

ii) Only simple and multiple regressions are explained and used.

1.6 DATA

The data used for this project work are students' examination scores for Science and social science class in Government Day Secondary School Minna.

Scores of students are entered into the record sheets by each subject teacher teaching that particular subject and these records are collected and stored in the examination office for future uses.

Some students after been issued their results indulge in alterations. Complaints of such cases by the parents are been attended to by the examination officer who makes use of the record sheets to confirm any discrepancies.

Some teachers mistakenly skip entering some student's scores. Such complaints are been dealt with by the clerks in the exams office consulting the record sheet of that particular class.

The school uses the record sheets to evaluate the average performance of a particular class or set. This is used as tool in measuring the performance of the students and also to ascertain whether the aim of the school is achieved.

Record sheets are made available to students of higher learning who make use of the data to carry out their research.

.7 LAYOUT OF THE PROJECT

Chapter one gives an insight on the need for statistics in educational system . a brief introduction was made explaining what the project work is all about. The case study was briefly explained along with organization structure diagram made. Objectives and scope of the study were also clearly spelt out and also the type of data used was explained.

Chapter two is a literature review, which explained some of the challenges faced during the project work. An introduction into simple and multiple regression is also in the chapter. This chapter explained some areas of application of regression analysis, and some advantages of using it. Test to confirm the variability of the tool used, which is regression analysis, is also explained.

Chapter three is concerned with the methodology of the existing system. This implies the manual computation of regression analysis. Some other techniques applied such as matrix model to solution were also explained.

Chapter four explains vividly the syntax used by the application program towards archiving the manually computed results in chapter three were made. Explanations on computed results by the application package were also made clearly.

Lastly, chapter five explains further some of the results in chapter four. In addition, summary of chapter three and four were made and a conclusion reached based on their comparism.

CHAPTER TWO

LITERATURE REVIEW

Isyaku (1997) in his work on effect of rainfall and temperature on rice production employed regression analysis technique to carry out his analysis. He wrote a program in QBasic to compute the regression for effect of annual rainfall on annual yield per bag. He also modeled effect of annual temperature on annual yield. Only simple regression was used in his analysis.

In this research work, program will not be used but a statistical package known as SPSS package, version 10 will be used to compute the regression analysis. Multiple regression will also be explained and employed in modeling linear equation for students performance.

Regression Analysis

Based on the case study (G.D.S.S, MINNA), supervisors from secondary schools board are always interested in whether students who do well in term papers also do well in the final examination and whether students who did poorly in one also did poorly in the other. They are always interested too in knowing whether population of students per class affects the students performance in term papers and examinations or not. It is equally expected that a student who is good in physics should also be good in mathematics. Efficiency of teachers can be determined based on some parameters. Based on past records of students performance, the instructor can estimate the final examination and conclude whether they will do well or not. Proper use of regression technique can give answer to these questions and many more of such related questions.

Regression is the process of fitting a straight line to two or more continuous variables, say x's and 'y'. Regression can be said to as a set of techniques, which utilizes the presence of an association between two variables to predict the value of one (the dependent variable or primary variable) from those of another (the independent or secondary variables).

The process of estimating and using a straight-line relationship is referred to as simple linear regression analysis. Basically, there are two methods of fitting as a straight line to the two variables. There are however other approaches to fitting s straight line but we will be restricted to the two below:

1.1 PROBABILISTIC AND DETERMINSTIC MODEL

If we were to construct a model that suggests an exact relationship between variables, it would be called a deterministic model. For example, if we believe that y, the students' performance in final examination will be exactly 2 times x, his/her performance, then

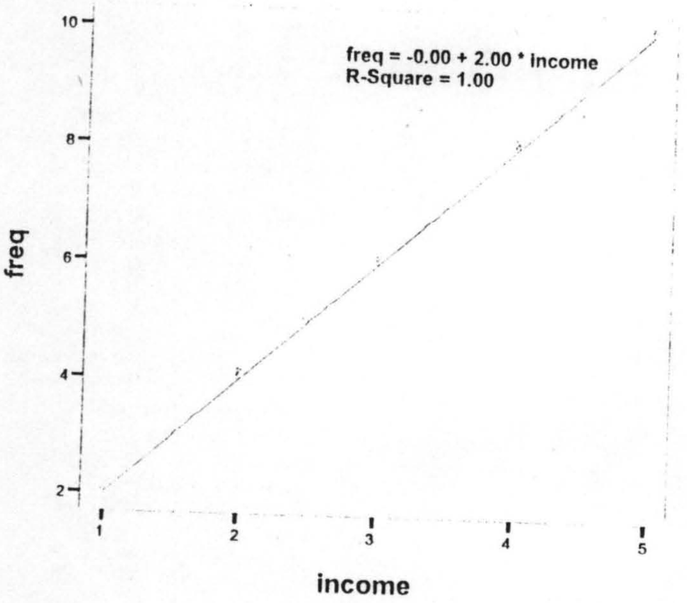
$$Y=2x$$

This represents a deterministic relationship between the variable y and x. and this implies that y cab always be determined exactly when the value of x is known. This method does not give room for error. In other words, its values (y) are exact. If on the other hand, we believe that there will be unexplained variation in the final

examination performance maybe caused by important but unincluded variables or by random phenomenon, we then cannot use the deterministic model but another model. This model accounts for this random error. Therefore, if we assume that the final examination performance is related to the percentage of term paper then

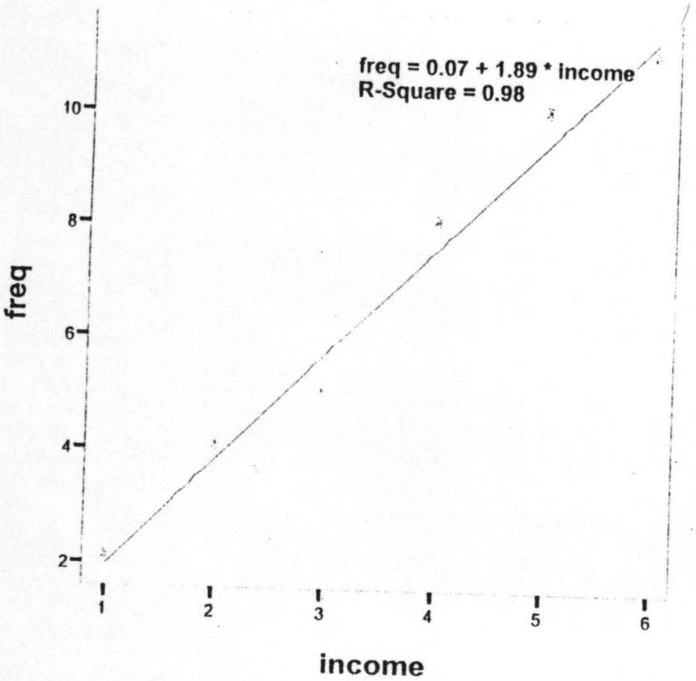
$$Y=2x + \text{Random Error}$$

Fig1



Linear Regression

fig2



Linear Regression

In fig1, all the values most fall exactly on the line because a deterministic model does not give room for error. On the other hand, fig2 shows that some points fall off the straight line implying a probabilistic model, giving room for error.

.1.2 GENERAL FORM OF PROBABILISTIC MODEL

Y= Deterministic component + Random Error

Where y is the variable of interest. We always assume that the mean value of the random error equals '0'. This implies that the mean value of y, E (y) equals the deterministic component of the model. That is

E (y)= Deterministic component.

.1.3 A FIRST ORDER (STRAIGHT LINE) PROBABILISTIC MODEL

$$Y = \beta_o + \beta_1 x + \epsilon$$

Where y =Dependent or response variable (variable to be modeled)

X= Independent or predictor variable (variable used as predictor of y)

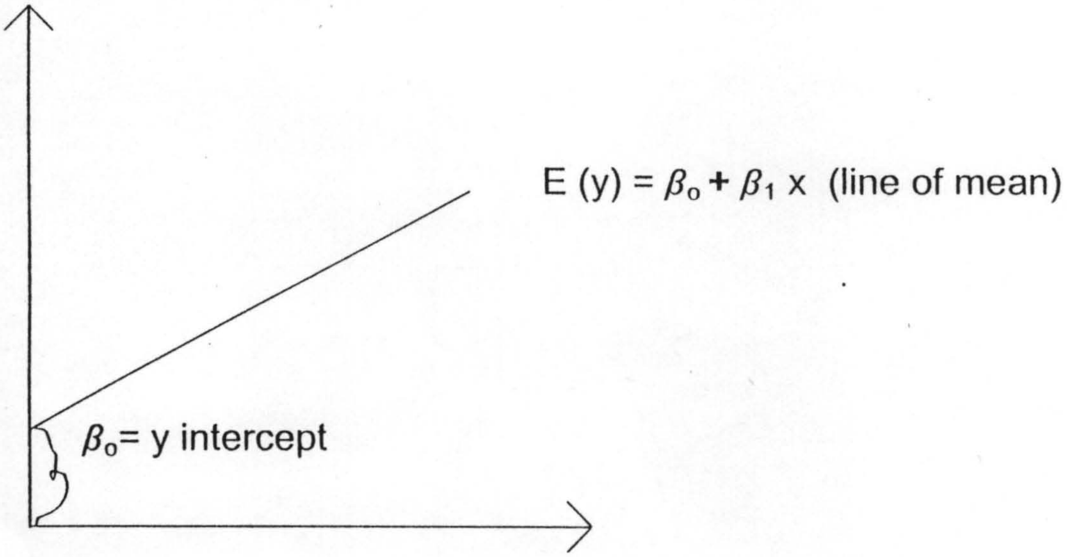
$E (y) = \beta_o + \beta_1 x$ = deterministic component

ϵ = Random error component or residual

β_o = y intercept of the line, the point at which the line cuts through the Y-axis

β_1 = Slope of the intercept, that is the amount of increase (or decrease in the deterministic component of y for every 1 unit in x)

graph



Note that in the probabilistic model, the deterministic component is referred to as the line of mean, because the mean of y, E (y) is equal to the straight-line component of the model.

That is

$$E(y) = \beta_o + \beta_1 x$$

1.4 THE LEAST SQUARE APPROACH

This method is used to determine the regression line using mathematical method called the least square principle. It determines a regression equation by minimizing the sum of squares of the vertical distance between the actua

Y values and the predicted values $\hat{y}$.

$\hat{Y} = a + bx$--------------------(1)

$\hat{Y}$ is the fitted estimate of Y based on a, b, and x

'a' is the Y-intercept. It is the estimated value of Y when $x = 0$. it can also be said to as the estimated values of Y where the regression line crosses the Y axis when x is equal to zero

Mathematically,

$$\hat{a} = \overline{y} - bX \quad \text{or}$$

b is the slope (that is, the average change n y per unit change in x).

Another name for b is the regression coefficient

Mathematically,

Where

x_i = value of the independent variables

y_i = values of the dependent variables

n= Number of items in the sample

The above equation (1) is the same for equation of a straight-line graph in general and $\hat{a}$ is the intercept (the average value of y when $x=0$). The said equation is also referred to as linear equation because when the values are plotted and joined together, it gives us a straight-line graph. The equation is a simple, two variable regressions; this is because only two variables can be analyzed with this method.

The accuracy of an estimate solemnly depends on the extent to which the regression equation and its corresponding graph actually fit the data. The regression line is always the line of best fit. One way of obtaining the line of best fit is to plot the values and ion as many points as possible with a single straight line. We then locate where this line cuts the Y-axis to give us the constant 'a'

Line of best fit is a situation of drawing the graph going by some criteria. Among, we could adopt would be to minimize the magnitude of the error involved, that is the difference between the values of the points on the diagram (Y) and the values of the points on the regression line we have drawn (y) and the values of point on the regression line we have drawn. From that, we can see that line of best fit is that line which minimizes the sum of square s of the errors.

The values of â and b will be unknown in almost all practical applications of regression analysis. The process of developing a model, estimating the unknown parameters and using the model can be viewed as the five step procedures stated below

1) Postulate the deterministic component of the model that related the mean E(y), to the independent variable x.
2) Use the sample data to estimate unknown parameters in the model

3) Specify the probabilistic distribution of the random error term and estimate the standard deviation of this distribution.
4) Statistically, evaluate the usefulness of the model.
5) When satisfied that the model is useful, use it for prediction, estimation and other purpose.

The above-mentioned steps will be applied in a later chapter of this project.

2.2 MULTIPLE REGRESSION

A Multiple regression analysis involves fitting the model to data set, testing the utility of the model and using it for estimation and prediction. It is a technique developed for modeling the mean of y as a function of two or more independent variables. As such, probabilistic models that include more than one independent variable are called multiple regression models.

The general form of this model is

$$Y= \beta_o + \beta_1 x_1+\beta_2 x_2+ ---+\beta_k x_k$$

Which minimizes the sum of square error (SSE)

$$SSE =\Sigma (y-\acute{y})^2$$

The primary difference between fitting the simple and multiple regression models is computational difficulty. The (k+1) simultaneous linear equations that must be solved to find the estimated coefficients of

$\beta_o, \beta_1, \beta_2, \text{----------}, \beta_k$ are difficult and sometimes nearly impossible to solve with manual computation. However, with computer and computer software such as SAS, MINITAB and SPSS, the computation becomes easier.

2.3 Advantages Of Using Regression Analysis

1) In the first place, we square the errors, which takes care of the negative signs and as such avoids errors canceling out the positive ones.
2) Squaring tends to emphasis the larger errors and minimizing criterion means that we are trying to avoid in the larger errors.
3) All points (variables) are taken into consideration, hence equal representation.

2.4 AREAS OF APPLICATION OF REGRRESSION ANALYSIS

2.4.1 HEALTH: - Health researcher may want to find out the relationship between hemoglobin and creatinine to see if hemoglobin and creatinine depends on renal function (as measured by the serum creatinine Concentration). Can patients with renal failure be easily exposed to anemia? Does the amount of drugs sold in a pharmaceutical locality depend on? the number of patients who visit the local hospital? The researcher is trying to find out the rat of self-medication.

The regression technique can be used in diverse areas of health.

2.4.2 **AGRICULTURAL SECTOR: -** in a highly mechanized farm, projection on yield can be made based on the quality of seeds planted, number of bags of fertilizer applied, mean annual rainfall, number of hectares covered, e.t.c. Different brands of fertilizer applied on a piece of farmland at different time can he analyzed to know the better of the fertilizer brands that yield better and quality produce.

2.4.3 **BUSINESS SECTOR: -**The manager of an industry may want to know which of the factors affects the gross profit of the firm among the following variables; the number of employees, consecutive dividends, beginning inventory, etc.

Circulation manger of a newspaper may want to confirm what variables relate to the number of subscriptions to the paper. Variables such as metropolitan population, the advertising budget of the paper and median family income in the metropolitan area. With proper regression analysis, solution will be obtained.

Regression analysis techniques can be employed for analysis in diverse area of human endeavor and areas of study (discipline).

2.5 RESIDUALS

When regression equation is used to determine the value of a variable 'y' from those one or more independent variable(s) 'x', the estimates ý will usually fall short of complete accuracy, geometrically, speaking, the data points will not all fall precisely upon the straight line, plane or hyper plane specified by the regression. The discrepancies (y- ý) on the predicted variable are known as residual. When using regression methods, the study of the residuals is of great importance, because they form the basis for measures of the accuracy of the estimates and of the extent to which the regression model gives a good account of the data in question.

Mathematically, Residual

$$Sy = \sqrt{\frac{\Sigma (y-\acute{y})^2}{n-1}}$$

Model Test Of Variability

Confirmatory tests are carried out to ascertain the overall utility of a model. Despite the fact that a t-test can be conducted in checking the overall utility of a model, it has some lapsed. Fitting a model to two or more quantitative variability and using t-test to check its variability may include a large number of insignificant variables and exclude some useful ones. Hence, if we want to test the utility of a multiple regression model, we will need a more comprehensive test that combines all the B_i's. Also we may like to find some statistical quantities that measure how well the model fits the data.

2.6.1 Fitness Of A Model On Quantitative Data

The method for finding the fitness of a model on quantitative data can be measured using a statistical technique called multiple coefficient of determination, R^2 and it is expressed as

$$R^2 = 1 - \frac{SSE}{SS_{yy}}$$

Where $SSE = \Sigma\,(y-\acute{y})^2$

$$SS_{yy} = \Sigma\, y^2 - \frac{(\Sigma y)^2}{n} \qquad = \frac{\textbf{Explained variability}}{\textbf{Total variability}}$$

R^2 is interpreted as the fraction of the sample variation of the y values (measured by SS_{yy}) that is explained by the least square prediction equation. Thus, R^0**=0** implies a complete lack of fit of the model to the data and $R^2 = 1$ implies a complete perfect with the model passing through every data point. However, we should note that $0< R^2$ **<1** .R^2 is a sample statistics that tells us how well the model fits that the data and therefore represents a measure of the usefulness of the entire model.

A large value of R^2 computed form the sample data does not necessarily tell us that the model provides a good fit to all of the data points in the population. Cases of models that comprise of more than two parameters will always provide a perfect fit. Also, R^2 is always equal to 1 for model with more than two parameters. As such, when dealing with more than two parameters β_i, a model adequacy; the adjusted multiple coefficient of determination denoted by R_a^2 Is preferred.

Mathematically,

$$R_a^2 = 1 - \left[\frac{n-1}{n-(k+1)}\right]\left(\frac{SSE}{SS_{yy}}\right)$$

$$= 1 - \frac{n-1}{n-(k+1)} * (1 - R^2)$$

Note that R_a^2 is always less than or equal to R^2, though they have similar interpretation. R^2 and R_a^2 are only sample statistics and as such are not reliable in judging usefulness of the model. Therefore, a test of utility to the multiple regression models is required.

6.2 Test Of Utility

Test of utility involves a test of hypothesis encompassing all the β parameters (except β_o) in a model.

Hypotheses can be said to as statement about a population developed for the purpose of testing. Hypothesis testing; a procedure based on a sample evidence and probability theory to determine whether the hypothesis is reasonable statement.

Null hypothesis, denoted as Ho; is a statement about the values of a population parameter, the alternative hypothesis (H_1) describes what you will conclude if you reject the null hypothesis.

Alternative hypothesis is a statement that is accepted if the sample data provide enough evidence that the null is false.

In making hypothesis formulation for multiple regressions, we would test:

$$Ho:\ \beta_1 = \beta_2 = \beta_3 = 0$$

$H_{1:}$ at least one of the coefficients is non-zero

To confirm the utility of the null hypothesis, an F-statistics test can be computed and used

Mathematically

$$\text{F-statistics} = \frac{SS_{yy} - SSE/k}{SSE/[\ n-(k+1)]}$$

$$= \frac{R^2 / k}{(1- R^2) / [n-(k+1)]}$$

We then compare the calculated F-statistic to a tabulated F-value with k degree of freedom (df) in the numerator and [n-(k+1)] df in the denumerator. Of the calculated value exceeds the tabulated value, we deject the null hypothesis and conclude that al least one of the model coefficients $\beta_{1,} \beta_{2,} \beta_{3,\text{------}} \beta_k$ is non zero. Therefore, this global F-test indicates that the first order model

$$Y= \beta_o + \beta_1 x_1+\beta_2x_2+ \text{---}+\beta_kx_k$$

Is usually for predicting the dependent variable for example say final examination score.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.1 Introduction

As earlier explained, this project work will be based on regression analysis technique. The set of data to be used are exams scores of students in the senior classes of science and social science classes. The data variables used are mathematics, physics and chemistry scores from the science class. While data variables used for social science class are Government, Economics and commerce subject.

3.2 Techniques-:

In carrying out the regression analysis, certain variables (subjects) are made dependent while some independent variables.

For the science class, the criteria for selecting the dependent variable is that a student who is good in physics must be very good in mathematics which was used for simple regression. This implies also that for a student to be good in physics, he/she must be well knowledgeable in mathematics subject. Equally, for the multiple regressions, physics and chemistry were selected as independent variables since a student who is good in those subjects is also expected to be very good in mathematics. The same reason goes for the social science class where Economics was selected as dependent variable wit Government and Commerce subjects as independent variables.

.3 Least square computation

3.3.1 Least-Square Computations for sheet1

Computation using the least square approach as earlier explained involves a dependent variable with only one independent variable. Form score Sheet1, assuming we use Maths score as dependent variable and Physics scores as independent variable. Maths performance score can to some extent be estimate using the least square approach as thus-:

Let Maths score $= \hat{y}$

Physics score $= x_i$

Then $\hat{y} = \hat{a} + b\,x_i$

$\hat{a} = -\,b\,x_i$

$$b = \frac{n\,\Sigma xy - (\Sigma x)(\,\Sigma y)}{n\,(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{44*80189 - 1781*1923}{44*78463 - (1781)^2}$$

$$b = 0.3689 \approx 0.37$$

$$\hat{a} = \frac{1923}{44} - 0.3689 * \frac{1781}{44}$$

$$\hat{a} = 28.77$$

$$\hat{y} = 28.77 + 0.37\, x_i$$

Accordingly, a linear equation can be deduced for mathematics score using chemistry scores

$$b = \frac{44 * 72820 - 1595 * 1923}{44*63635 - (1595)^2}$$

$$= \frac{13689}{255915}$$

$$= 0.535 \qquad \approx 0.54$$

$$\hat{a} = \frac{1923}{44} - 0.535 * \frac{1595}{44}$$

$$\hat{y} = 24.31 + 0.54 * x_i$$

3.3.2 Residual Computation For Least Square

The residual values are used to estimate the random error and to check the regression assumptions. The residual values are plotted on the vertical axis against the variable (dependent) y_i, on the horizontal axis

$$Sy = \sqrt{\frac{\Sigma\,(y-\acute{y})^2}{n-1}}$$

$$= \sqrt{\frac{1580.624}{44 - 1}} \qquad = \sqrt{36.76} \qquad = 6.06$$

The above computed residual is for mathematics with Physics as independent variable.

We then computed residual for mathematics with chemistry as independent.

$$Sy = \sqrt{\frac{\Sigma (y-\acute{y})^2}{n-1}}$$

$$= \sqrt{\frac{1930.94}{4}} \qquad = \sqrt{44.9} \qquad = 6.70$$

3.3 Least Square Computations For Sheet2

We know consider sheet2 by changing the variables that is making use of economics marks as dependent variable y and commerce independent

$$\hat{y} = \hat{a} + b\, x_i$$

$$b = \frac{n\, \Sigma xy - (\Sigma x)(\Sigma y)}{n\, (\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{46 * 93177 - 2205 * 1909}{46 * 120127 - 2205^2}$$

$$= \frac{76797}{663817} \qquad = 0.116$$

$$\hat{a} = y - b\, x_i$$

$$= \frac{1909}{46} - 0.166 * \frac{2205}{46}$$

$$= 41.50 - 5.56 \qquad \hat{a} = 39.9$$

We can equally derive a linear equation for eons score using government score as independent marks

$$\hat{y} = \hat{a} + b\, x_i$$

$$b = \frac{n\, \Sigma xy - (\Sigma x)(\Sigma y)}{n\, (\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{46 * 88050 - (1932)(1909)}{46 * 99436 - 1932^2}$$

$$= \frac{362112}{841432}$$

$$= 0.43$$

$$\hat{a} = y - b\, x_i$$

$$= \frac{1909}{46} - 0.43 * \frac{1932}{46} \quad = 41.5 - 18.06$$

$$\hat{a} = 23.44$$

$$\hat{y} = 23.44 + 0.43 * \text{Gov't score}$$

.4 Residual Computation For Least Square Of Sheet2

$$Sy = \sqrt{\frac{\Sigma (y-\acute{y})^2}{n-1}}$$

$$= \sqrt{\frac{12868}{45}}$$

$$= \sqrt{285.96} \quad = 14.10$$

3.4 Multiple Regression Computation

From record sheet1, we will have more than one independent variable since we are computing a multiple regression. Mathematics score will be our dependent variable with physics and chemistry scores as independent variables denoted by x_1 and x_2 respectively. Note that more than two independent variables can be used but the computation becomes more complex and in some cased impossible.

To compute and obtain the linear equation for three variables, we generate the following three equations using

$\hat{y} = \hat{a} + b_1x_1 + b_2x_2$ giving rise to the three equations as follows

$$\Sigma y = an + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = a \Sigma x_1 + b_1 \Sigma x^2_1 + b_2 \Sigma X_1 X_2$$

$$\Sigma x_2 y = a \Sigma x_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma x^2_2$$

The unknown variables are computed using any convenient method to solving matrix problem. We then substitute the various values from the table given.

From record sheet1

$$1923 = 44a + 1781b_1 + 1595 b_2$$

$$80189 = 1781a + 78463 b_1 + 67747 b_2$$

$$72820 = 1595a + 67747 b_1 + 63635 b_2$$

Extracting out the coefficients to form a corresponding matrix as A, to use the inverse method

$$A = \begin{pmatrix} 44 & 1781 & 1595 \\ 1781 & 78463 & 67747 \\ 1595 & 67747 & 63635 \end{pmatrix}$$

$AX = b$

But $b = A^{-1}X$

To find A^{-1},

$$|A| = \begin{vmatrix} 44 & 1781 & 1595 \\ 1781 & 78463 & 67747 \\ 1595 & 67747 & 63635 \end{vmatrix}$$

$|A| = 44(403336996) - 1781(5277470) + (-4491078)$

$= 1184384344$

We then find and form the matrix of cofactors 'N'

$$\begin{pmatrix} 403336996 & -5277470 & -4491078 \\ -5277470 & 255915 & 191093 \\ -4491078 & -140173 & 280411 \end{pmatrix} = N$$

There after, we transpose the matrix of cofactors

$$N^T = \begin{pmatrix} 403336996 & -5277470 & -4491078 \\ 5277470 & 255915 & -140173 \\ -4491078 & 191093 & 280411 \end{pmatrix}$$

$$A^{-1} = \frac{1}{|A|} * N^T$$

$$= \frac{1}{1184384344} * \begin{pmatrix} 403336996 & -5277470 & -4491078 \\ 5277470 & 255915 & -140173 \\ -4491078 & 191093 & 280411 \end{pmatrix} = Adj.A$$

$$\frac{1}{A} * \begin{pmatrix} 403336996 & -5277470 & -4491078 \\ 5277470 & 255915 & -140173 \\ -4491078 & 191093 & 280411 \end{pmatrix} \begin{pmatrix} 1923 \\ 80189 \\ 72820 \end{pmatrix} = \begin{pmatrix} a \\ b_1 \\ b_2 \end{pmatrix}$$

$\hat{a} = 654.8693819 + (-357.3122559) + (-276.1268347)$

$\hat{a} = 21.43$

$b_1 = -8.568649916 + 17.32678082 + (-8.618315424)$

$b_1 = 0.140$

$b_2 = 0.46$

Therefore,

$\hat{y} = 21.43 + 0.14b_1 + 0.46b_2$

3.4.1 Residual For Multiple Regression

$$Sy = \sqrt{\frac{\Sigma\,(y-\acute{y})^2}{n-1}}$$

$$= \sqrt{\frac{1580.624}{44-1}}$$

$$= \sqrt{36.754} \qquad = 6.06$$

3.4.2 Multiple Coefficient Of Determination

$$R^2 = 1- \frac{SSE}{SS_{yy}}$$

Where $SSE = \Sigma\,(y-\acute{y})^2$

$$=3335.159$$

$$SS_{yy} = \Sigma\, y^2 - \frac{(\Sigma y)^2}{n}$$

$$= 87379 - \frac{1923^2}{44}$$

$$= 3335.159$$

$$= 1-\frac{1580.624}{3335.159}$$

$$= 1- 0.4739$$

Therefore,

$$R^2 = 0.53$$

3.3.4 Computation For Adjusted Coefficient Of Determination

$$R_a^2 = 1 - \left[\frac{n-1}{n-(k+1)}\right]\left(\frac{SSE}{SS_{yy}}\right)$$

$$= 1 - \frac{n-1}{n-(k+1)} * (1 - R^2)$$

Where k is the number of independent variables

$$R_a^2 = 1 - \left[\frac{44-1}{44-(2+1)}\right] (1-0.53)$$

$$= 1 - \left[\frac{(43)}{41}\right] (0.47)$$

$$= 1 - 0.493 \qquad = 0.507$$

3.3.5 Test Of Hypothesis

$$H_o: b_1 = b_2$$

$$H_1: b_1 \neq b_2$$

$$\text{F-statistics} = \frac{SS_{yy} - SSE/k}{SSE/[\,n-(k+1)]}$$

$$= \frac{R^2 / k}{(1-R^2)/[n-(k+1)]}$$

$$= \frac{0.53/2}{(1-0.53)/[44-(2+1)]}$$

$$= \frac{0.265}{0.01146} \qquad = 23$$

3.5 Analysis Of Variance

An experiment is a planned procedure that gathers comparative data under controlled conditions. When experiments are performed, conclusions are reached after analyzing the results, based on the data with two or more independent variables; ANOVA is used to determine the difference between the averages.

The research questions are

1) How strongly are variables associated?

2) Can score on a target variable (or category membership, if the variables are qualitative) be predicted from data on other variables?

3) Is a difference (between averages) significant?

The analysis of variance (ANOVA) is actually a whole set of technique, each based upon a model of how the data were generated and culminated in tests that are appropriate for that are appropriate for that particular model only. As such, it is important to identify ANOVA correctly, in order to choose the right test.

In ANOVA, a factor is a set of related conditions or categories.

The conditions or categories making up a factor are known as levels.

Some factors are said to be between, this implies that participants are treated under the same conditions. Other factor can be said to as within subjects, that is, participant is treated under all the conditions (levels) making up the factor.

In the course of this project work, I am trying to analyze the relationship between μ_1, μ_2, μ_3,----, μ_k . And since we are making use of sample, the data will be in a tabular form as shown below

			Total
	X_{11}	X_{12}---------X_{1j} ------X_{1c}	$\bar{X}_{1.}$
	X_{21}	X_{22}---------X_{2j}-------X_{2c}	$\bar{X}_{2.}$
	------------------	--------------------------------------	--
	X_{r1}	X_{r2}---------X_{rj}--------X_{rc}	$\bar{X}_{r.}$
Total	$\bar{X}_{.1}$	$\bar{X}_{.2}$---------$\bar{X}_{.j}$------- $\bar{X}_{.c}$	$\bar{X}_{..}$

One way classification model assumes a hypothesis $\mu_{.1} = \mu_{.2} =$ ------- $= \mu_{.k}$. The model is then applied to actual data to find out whether the hypothesis is true. An example is that given below

X_{i1}	X_{i2}	X_{i3}
x_{11}	x_{12}	x_{13}
--------	--------	--------
x_{41}	x_{42}	x_{43}
Σx_{i1}	Σx_{i2}	Σx_{i3}

Example,

$$\mu_{.2} = \frac{\Sigma x_{i2}}{N2} \quad = \frac{\Sigma X_{i2}}{4}$$

$$\mu_{..} = \frac{\Sigma\, \Sigma x_{ij}}{N1+N2+N3} \quad = \quad \frac{\Sigma X_{ij}}{12}$$

The interpretation is the same as before. If the hypothesis is rejected, we assume that the µ's are independent. We can only say that they are not significantly different keeping in mind that, generally small samples cannot discover small differences. Lets now see how a researcher can reach conclusion about µ's from the comparism of variances of we have k sub-population whose sizes are N_1, N_2, N_3, -------,N_k and whose means are µ.1, µ.2,---------- µ.k, the pooled variance is found by

$$\delta^2_p = \frac{\Sigma(X_{i1} - \mu_{.1})^2 + \Sigma\,(Xi2 - \mu_{.2})^2 + \text{----} + \Sigma(X_{ik} - \mu_{.k})}{N_1+N_2+N_3+\text{-----------}+N_k} \qquad \text{------}1$$

Where $\mu_{.j}$ are column means, the total variance then will be computed by the formula

$$\delta^2_t = \frac{\Sigma\, \Sigma(XIJ - \mu_{..})^2}{N_1+N_2+\text{---}N_k} \qquad \text{---}2$$

Between μ_{ij} values

$$\delta^2 \mu_{.j} = \frac{\Sigma(\mu_{.j} - \mu_{..})^2}{k} \qquad \text{--}3$$

Example1

X_{i1}	X_{i2}	X_{i3}
1	1	1
5	5	5
3	3	3
7	7	7

Hypothesis

Ho: $ɥ_{ij} = ɥ$ and $\delta^2_t = \delta 2p$

H_1: $ɥ_{.1} \neq ɥ_{.2} \neq ɥ_{.3} \neq ɥ$ and $\delta^2_t > \delta^2_p$

Recall that the sample mean is

$$\delta^2_{\bar{X}} = \frac{\delta^2}{n} \quad \text{---4}$$

From which the population variance is obtained by writing

$$\delta^2 = n\,\delta^2_{\bar{X}} \quad \text{---5}$$

And the value of the pooled variance is according to equation (1)

$$\delta^2_p = \frac{20+20+20}{4+4+4} = 5$$

We can also have another example as below

Example (2)

X_{i1}	X_{i2}	X_{i3}
1	11	21
5	15	25
3	13	23
7	17	27

The total variance using equation (2) given ɥ..= 4 is

$$\delta^2_t = \frac{\Sigma\, \Sigma\, (x_i-4)}{12} = 5$$

Thus, it is obvious that these two variances are equal because in the problem at hand, the variance for each column are the same and the mean $\mu_{.j}$ for each column are also the same. In this case, $\delta^2_{\mu.j}$ given in equation (3) is equal to zero. Therefore, comparism of these two variance (that is δ^2_t and δ^2_p) will certainly tell us that the column means $ɥ_{ij}$ are the same.

The relation between the same variance is different in example (2). The variance for individual columns remain the same so that $\delta^2_1 = \delta^2_2 = \delta^2_3 = 5$. Therefore, the value of pooled variance δ^2_p computed by equation (1) remains the same. However, the grand mean now is

$$\mu_{..} = \frac{\Sigma\, \Sigma X_{ij}}{N_1+N_2+N_3} = 14$$

Therefore, the total variance is

$$\delta^2_t = 860 / 12$$

$$= 71.67$$

which is much greater than the pooled variance $\delta^2_p = 5$ in example1.

Again in example 2, the pooled variance is equal to the column variance. However, the total variance is greater than the pooled variance and this is an indication that column means must be different. Thus, the knowledge and the comparism of these two variances are enough to tell us that the column means are different. We come to this conclusion without looking at the values of column means.

Now, assuming that the data in example 2 are not population but sample data. If so, then lieu to comparing δ^2_t with δ^2_p, we will compare (with the same result)

$$S^2_p = \frac{\Sigma (X_{i1} - \bar{X}_{.1})^2 + \Sigma (X_{i2} - \bar{X}_{.2})^2 + \text{------} + \Sigma (X_{ik} - \bar{X}_{.k})^2}{n_1 + n_2 + n_3 + \text{----------} + n_k - k} \text{ --------------------6}$$

with

$$S_b^2 = \frac{n \, \Sigma (\bar{X}_{.j} - \bar{X}_{..})^2}{k-1} \text{ ---7}$$

where $\bar{X}_{..}$ is the grand mean of sample data, computed in the same manner as $\mu_{..}$ above. Equation 7 assumes that sample size n are equal. If they are not, the formula becomes

$$S_b^2 = \frac{\Sigma n_j (\bar{X}_{.j} - \bar{X}_{..})^2}{k-1} \text{ --8}$$

Without n and n_j in equation 7and 8 respectively, the equation resembles the formula in equation 3. It actually becomes

$$\delta^2_{\bar{x}} = \frac{\Sigma (\bar{X}_{.j} - \bar{X}_{..})^2}{k-1} \text{ --9}$$

The formula in equation 9 measures the dispersion of sample means $\bar{X}_{.j}$

and it is an estimate of δ^2_x given in equation 4 above. If both δ^2_x and the sample size n are known, δ^2_x can be used as an estimate of population variance

$\delta^2 = n \delta^2_x$.. If such as estimate has to be made from sample data, the variance of the population is estimated by equation 7,which is obtained by multiplying equation 9 by n. This variance, that is S_b^2 in equation 7 is generally an estimate of $\delta^2 + n\delta^2\mu_{.j}$ .However, it has already been noted that if the column means $\mu_{.j}$ are equal to zero, then $\delta^2\mu_{.j}$ is equal to zero. In such a case, S_b^2 in equation 7 (or equation 8) becomes an estimator of δ^2 in equation 6. if on the other hand; the column means are not the same, then S_b^2 becomes an estimator of $\delta^2 + n\delta^2_{\mu.j}$. it will therefore be significantly greater than the pooled variance S_b^2 as a result, we have the following fundamental rules in the analysis of variance: if S_b^2 is significantly greater (and not significantly different) than S^2_p, column μ's are significantly different. A significantly greater S^2_p indicates that the variance $\delta^2_{\mu.j}$ is significantly different from zero, which means that the $\mu_{.j}$ are significantly different. If on the other hand, S_b^2 is not significantly greater than S^2_p, the column means are not significantly different.

It is useful to not that there is additive relationship between the numerators of equation 6,7 and 8. It can be shown as

$$\sum_i \sum_j (X_{.j} - X_{..})^2 = \Sigma\, \Sigma\, (X_{ij} - X_{.j})^2 + n\, \Sigma\, \Sigma\, (X_{.j} - X_{..})^2$$

In example2, we have

$$860= 4(200)$$

And in example1

$$60=60+0$$

These numerators are called sum of squares. The first is "Total Sum of Squares" or SS_t, the second," Sum of Square Within" or $SS_{w,}$ and the third, the "Sum of Square Between" or SS_b and the Mean Square Between $MS_{b;}$ MS_b can be computed as

$$MS_b = SS_b / v_1$$

Where v_1 is the number of degree of freedom. it is taken from the denumerator in equation 7 that is v_i =k-1. The variance $S^2{}_p$ is called mean square within or MS_w and can be computed as

$$MS_w = SS_w / v_2$$

The value of v_2 comes from the denumerator in equation 6. It is

$$v_2= \Sigma n_j -k$$

Using the students score sheets however, computing these various sums of square becomes very tedious, cumbersome and time consuming. For a voluminous data, it becomes impossible manually. This is why I have decided t only explain it computational procedure with example.

The analysis of variance in this classification consists of comparing MS_b

With SS_w. If MS_b Is significantly greater than SS_w, the investigator concludes that mean $\mu_{.j}$ are significantly different. The compares itself is done with the help of F-test. The computed F-value is obtained from

$F^c = SS_b / MS_w$ and it is compared with $F_{(\alpha,v1,v2)}$ which can be found.

These basic relationships are usually put together in tabular form

ANOVA TABLE

	SS	V_i	Ms	F
Between	SS_b	V_1	MS_b	SS_b / MS_w
Within	SS_w	V_2	MS_w	
Total	SS_t	V_3		

$SS_b = n \Sigma \Sigma (X_{.j} - X_{..})^2$

$SS_w = \Sigma \Sigma (X_{ij} - X_{.j})^2$

$SS_t = \Sigma \Sigma (X_{.j} - X_{..})^2$

$V_3 = V_1 + V_2$

CHAPTER FOUR

DESIGN AND IMPLIMENTATION

Since a program was note developed, but rather a standard application package known as Statistical Package for Social Sciences (SPSS) was used in modeling the students' achievement in their examination. Syntax towards achieving this objective is as below:

HOW TO INPUT DATA INTO THE DATA EDITOR

An operating system such as Windows creates a computing environment, or interface, within which the user accesses or runs applications such as SPSS and word processor. Windows creates a graphical interface, in which applications and other user-accessible software appear on the screen as small graphical symbols known as icons. When icon is activated and opened, commands can be given by using a device such as a mouse from lists of choices known as menus. When windows are first accessed, the screen shows an array in which several icons appear against a homogenous colored background known as the desktop. SPSS is opened by double clicking on the spss icon, a dialogs box with the question:

"What would you like to do?" appears.

For more experienced users, this dialog box can be stopped from appearing automatically on opening in a non-network computer by clicking on the don't show this dialog box in the future. If the user is entering mew data, click on the Type in data radio button and then on **OK** to bring the Variable View version of the SPSS Data Editor to the screen as shown below. If the **Data View** appears first, simply click on the tab labeled **Variable View** at the bottom left hand side of the window to open the Variable View

To name the variables English, Maths, Physics, type into the first cell of the first row in the column labeled Name of the Variable View and then press the down arrow key to move the highlighting down to the cell below. The type Maths, continue as explained earlier to type in Physics. The Variable View will appear as in below.

Once the user has completed the details of the variable in Variable View, then clicking on the Data View tab at the bottom of the window will bring on to the screen the Data View version.

In an SPSS data set, each row represents a single case, or subject and each column represents a variable. The cell to the left is highlighted and typing goes into the cell. Use the directional keys to navigate through the work sheet to enter values for each variable.

.3 SELECTING A PROCEDURE FROM THE MENU

After entering the data into the data view, clicking **Analyze** initiates a pull down menu. Pointing the mouse on **REGRESSION** brings a sub-menu, then click on **Linear.**

SELECTING A VARIABLE FOR THE ANALYSIS

To select variables to regress after clicking on Linear, dialog box appears The three variable names English, Maths and Physics will appear in the left-hand box. It is important to be clear about which variable is the dependent variable and which is the independent variable. Transfer these variable names into the appropriate boxes in the dialog box by clicking on the variable name and then on ► .This is a case of one independent variable which is an example of leas square method. It is recommended to request additional descriptive Statistics and a residual analysis. To obtain descriptive statistics, click on the **Statistics** button on the dialog box to open the **Linear Regression:** statistics dialog box Click on the **Descriptive** checkbox. A residual is the difference between the actual value of the dependent variable and its predicted value using the regression equation. Analysis of the residuals gives a measure of how good the prediction is and whether there are any cases, which are so discrepant that, they might be considered as outliers and so dropped form the from the analysis. Click on the Casewise diagnostics check box to obtain a listing of any exceptionally large residuals. Click on **Continue** to return to the Linear Regression dialog box

Information about residuals is obtained by clicking on the Plot button in the dialog box to open the **Linear Regression:** Plots dialog box

Since systematic patterns between the predicted values and the residual can indicate possible violations of the assumption of linearity, it is recommended that a plot of the standardized residual **(*ZRESID)** against the standardized predicted values ***ZPRED** is also requested by transferring ***ZRESID** into the Y: box and ***ZPRED** into the box X: box.

Click on **Continue** to return to the Linear Regression dialog box

Click on **OK** to return the regression command.

OUTPUT FOR SIMPLE REGRESSION

The various tables and charts in the output are listed in the left-hand pane of SPSS Viewer. Since the information about outliers in the Casewise Diagnostics table may affect whether the regression analysis is aborted and rerun after such outliers have been removed from the data set; it is recommended that this table is scrutinized first. It can be selected directly by moving the cursor to Casewise Diagnostics in the left-hand pane and clicking the left-hand mouse button.

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	PHYSICS	.	Enter

a All requested variables entered.

b Dependent Variable: MATHS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.510	.260	.242	7.67

a Predictors: (Constant), PHYSICS

b Dependent Variable: MATHS

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	867.438	1	867.438	14.764	.000
	Residual	2467.721	42	58.755		
	Total	3335.159	43			

a Predictors: (Constant), PHYSICS

b Dependent Variable: MATHS

Coefficients

Model	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1 (Constant)	28.771	4.055		7.096	.000
PHYSICS	.369	.096	.510	3.842	.000

a Dependent Variable: MATHS

Casewise Diagnostics

Case Number	Std. Residual	MATHS
4	3.446	74

a Dependent Variable: MATHS

Residuals Statistics

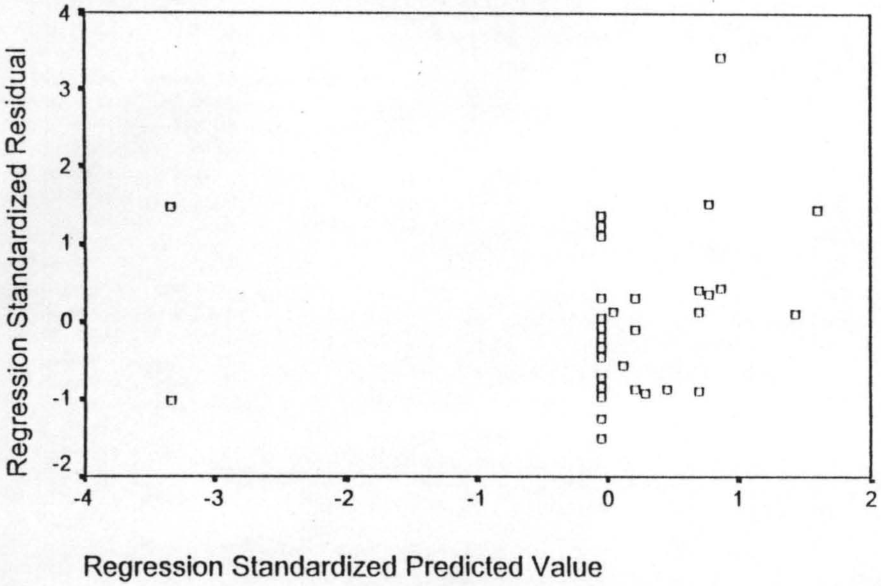
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	28.77	50.91	43.70	4.49	44
Residual	-11.53	26.41	3.23E-16	7.58	44
Std. Predicted Value	-3.325	1.604	.000	1.000	44
Std. Residual	-1.504	3.446	.000	.988	44

a Dependent Variable: MATHS

Charts

Scatterplot

Dependent Variable: MATHS



Result Interpretation Of Simple Regression

The table of casewise diagnostic in the output shows that case 4 with a score of 74 for physics score is the only outlier with an absolute value greater than 3. This outlier has to be eliminated and the regression analyses run again.

A more reliable regression analysis can be obtained by eliminating any outliers using the select case procedure described below

Choose Data,

Select cases…

To open the select dialog box. Click on the **if** condition is satisfied radio button and define the condition as Case ~=4 (the symbol ~= means ' not equal to')

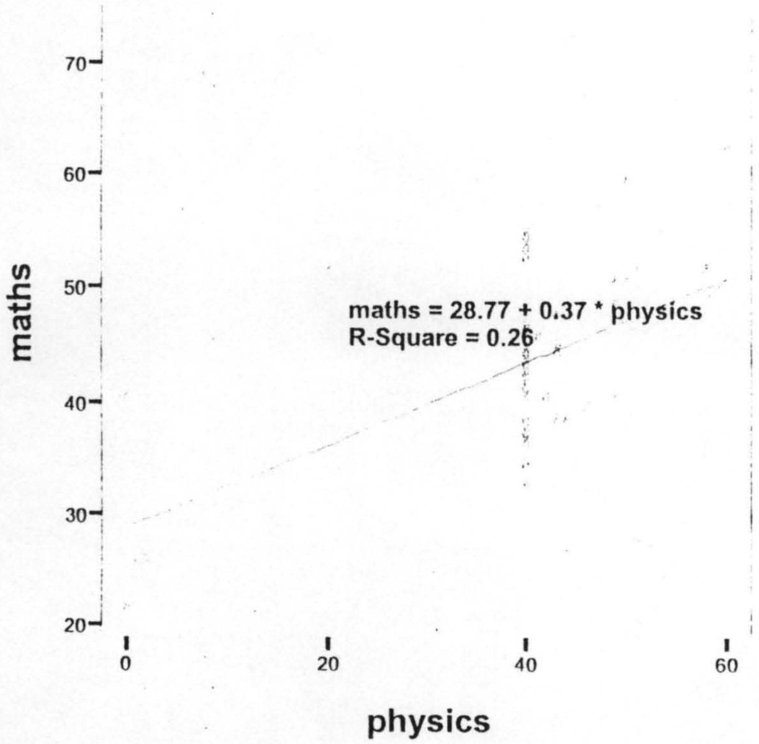
Clicks on continue and then **OK** to deselect this case to remove this case.

When the regression analysis is run again after deleting the original output to avoid confusion, there is no table of Casewise Diagnoses since no case have outlying residuals.

FITTING A SIMPLE REGRESSION LINE

To fit a simple regression line, click on **Graph** to obtain a pull down menu, point to **Interactive** and a sub menu appears. Click on **Scatterplot** to obtain a dialog box as. Click on the **3D-cordinate** at the top right side of the dialog box, a pull sub menu appears, then click on **2D-corddinate**. Click and drag on the intended dependent and independent variables into Y-axis and X-axis respectively. Click **OK** to run the graph.

Linear Regression



PROCEDURE FOR MULTIPLE REGRESSION

In the linear Regression dialog box, transfer the variable name Maths into the **Dependent Variable:** and Physics and chemistry into the **Independent Variables:** box by highlighting them and clicking on the appropriate button. To obtain descriptive statistics, click on the Statistics button as shown before to open the **Linear Regression**: Statistics dialog box. Click on the **Descriptive** che box. Click on **OK** to run the regression.]

Output For Multiple Regressions

Variables Entered/Removed

```
Model          Variables Entered Variables Removed   Method

    1      CHEMISTY, PHYSICS                    .      Enter

a  All requested variables entered.

b  Dependent Variable: MATHS
```

Model Summary

```
Model       R R Square      Adjusted R        Std. Error of the

                                Square                 Estimate

    1   .725     .526             .503                     6.21

a  Predictors: (Constant), CHEMISTY, PHYSICS

b  Dependent Variable: MATHS
```

ANOVA

```
Model             Sum of        df      Mean          F        Sig.

                  Squares              Square

  1  Regressio   1754.752       2     877.376    22.761       .000

          n

      Residual   1580.407      41      38.547

         Total   3335.159      43

a  Predictors: (Constant), CHEMISTY, PHYSICS

b  Dependent Variable: MATHS
```

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	21.430	3.623		5.915	.000
	PHYSICS	.140	.091	.193	1.532	.133
	CHEMISTY	.458	.096	.605	4.798	.000

a Dependent Variable: MATHS

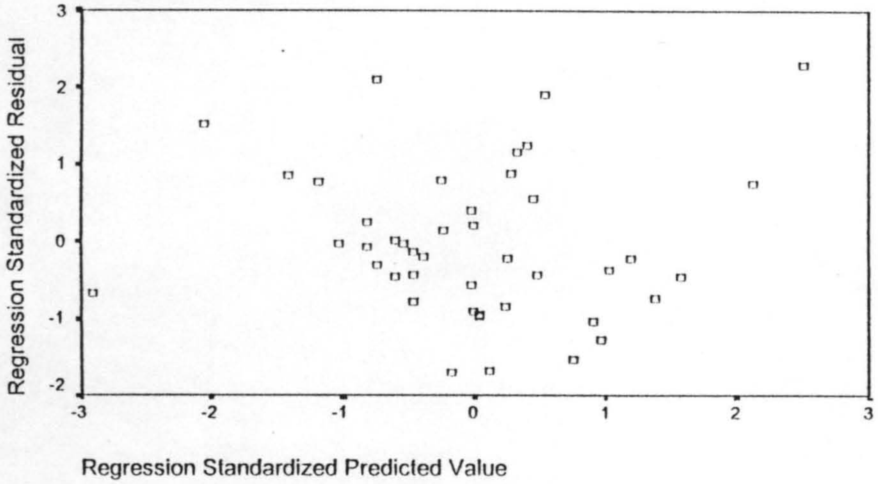
Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	25.10	59.73	43.70	6.39	44
Residual	-10.61	14.27	4.52E-15	6.06	44
Std. Predicted Value	-2.913	2.508	.000	1.000	44
Std. Residual	-1.708	2.299	.000	.976	44

a Dependent Variable: MATHS

Scatterplot

Dependent Variable: MATHS



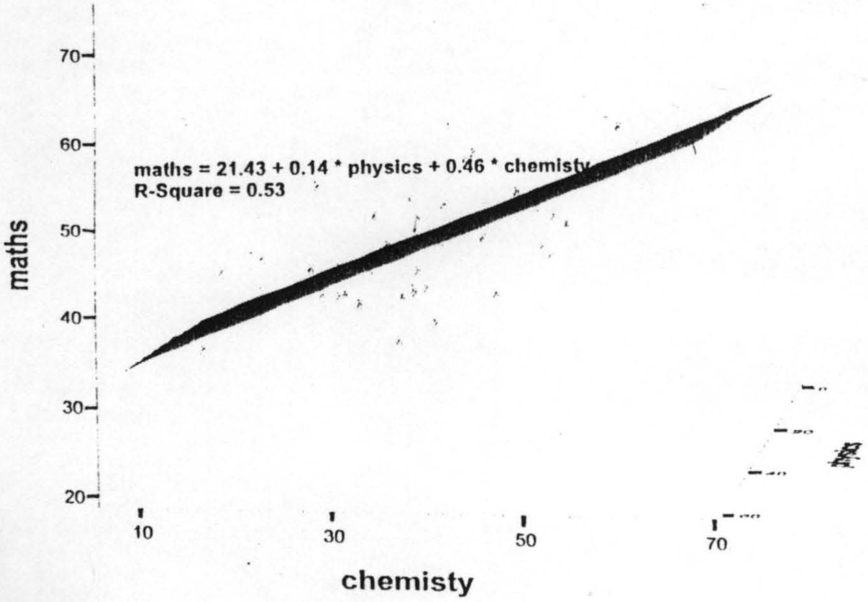
FITTING A MULTIPLE REGRESSION LINE MODEL

The syntax is as thus:

Click on **Graph** to obtain a pull down menu, point to **Interactive** and a sub menu appear. Click on **Scatterplot** to obtain a dialog box**.** Click and drag on the intended dependent and independent variables into the coordinates. Click **OK** to run the graph.

MULTIPLE REGRESSION GRAPH

Linear Regression



CHAPTER FIVE

SUMMARY AND CONCLUSION

.1 Limitation Of The Study

Derivations of the three linear equation that resulted into the matrix formulation was not explained.

This project work is geared towards attempting modeling students scores for only two independent variables.

5.2 Discussion

For the simple regression result, the first tables are the requested and correlation coefficients for the remaining 43 cases in output 2.

Output model summary shows the value for MULTIPLE **R** which in the case of just one independent variable has the same absolute value as the correlation coefficient listed in output of descriptive statistics. The other statistics listed are **R square** (a positively biased estimate of the proportion of the variance of the dependent variable accounted for by regression). **Adjusted R Square** (corrects this bias and therefore has a lower value), and **Standard Error** (the standard deviation of the residuals).

Output ANOVA contains the regression **ANOVA**, which is intended to test whether there really is a linear relationship between the variables by forming an F-ratio of the mean square for regression to the residual mean square.

In this example, the value of F in the ANOVA is highly significant. It should be noted, however, that only an examination of the scatterplot of the variables could confirm that the relationship between two variables is genuinely linear.

Output coefficient is the kernel of the regression analysis, because it contains the regression equation. The value of the **regression coefficient** and constant are given in column **B** of the table. The equation is therefore,

Predicted mathematics score = 28.771 + 0.369 * (physics score)

Thus person with Physics score of 40 would be predicted to score

$$28.771 + 0.369 * 40 = 43.531$$

Notice from the data that the person who did score 40 in Physics actually scored 42 in mathematics .the residual is, therefore,

$$43.531 - 40 = +3.531$$

Other statistics listed are Std. Error, which is the standard error of the regression coefficient B, Beta which is the beta weight showing the change in the dependent variable (expressed in standard deviating units) that would be produced by a positive increment of one standard deviation in the independent variable, t which is a t-test for testing the regression coefficient for significance and sig. which is the P-value of t (here .00means <0.005,that is , t is significant well beyond the 0.0. level for the variable Physics).

Output residual is a table of statistics relating to the residuals. The variables predicted value comprises the Unstandardized predicted values, the variable Residual is the set of Unstandardized residuals, the variable std. predicted value (identified as *ZPRED in the plots dialog box in table model summary) contains the standardized predicted values (that is predicted values transformed to a scale with mean 0 and SD 1). The scarterplot shows no obvious pattern, thereby confirming hat the assumptions of linearity and homogeneity of variance have been met. If the cloud of points were crescent-shaped or funnel-shaped, further screening of the data (or abandoned of the analysis) would be necessary.

For the multiple regression result, recall that when one independent variable physics was used to predict maths score, the value of R was 0.510. With R=0.725, we see hat the answer to the question of whether adding more independent variables mproves the predictive power of the regression equation is certainly 'yes', infact the mprovement is remarkable. The ANOVA output shows that the regression is still ighly significant (p<0.01). But what about the second question? Do both new ariables contribute substantially to the predictive power of the regression equation? From column B in the section headed unstandardized coefficients, we see that the multiple regression equation of maths score upon physics and chemistry score is

Y=21.430 + 0.14*physics +0.458*chemistry score

The beta weight tell us about the relative importance of the independent variables, because each gives the number of standard deviations change on the dependent variable that will be produced by a change of one standard deviation on the independent variable concerned.

Chemistry makes by far the highest contribution, because a change of one standard deviation on that variable produces a change of 0.608 standard deviation on maths score, whereas a change of one standard deviation in physics produces an increase of only 0.193 of a standard deviation in maths score. This equally signifies that the variable (predictor) with the largest beta weight also ha s the largest correlation wit the dependent variable.

.3 Summary

Computed results in chapter three can be used to form the models. These models formulated in turn are used to estimate students' scores. Using the linear equation in page 24

Maths score = 28.77 + 0.37 *Physics score

Substituting the various values of physics score into the linear equation gives an estimated corresponding Mathematics score for that particular student.

Example, for the student who scored 58 marks in Physics will have an estimated Mathematics score as

Maths score = 28.77 + 0.37 * 58

= 28.77 + 21.46

= 50.23

≈50

Substituting higher values of physics score results to a better mark or pass score equivalent to maths. Likewise, substituting a lower marks of physics results to a failure of mathematics.

Also from page 25

Maths score =24.31 + 0.54 * x_i

Example, a student who scored marks in chemistry will have an estimated or predicted maths score s thus,

Maths score = 24.31 + 0.54*53
= 24.31 + 28.62
= 52.93
≈53

5.4 Conclusion

The various results computed manually to obtain the linear regressions are almost as those of the computed using SPSS. Only slit variations occur. The estimate obtained using SPSS results give a better estimate to the corresponding dependent variable.

In the multiple regression models, better estimates are made of the dependent variable compared to those of simple regression.

In general, we are able to identify that Chemistry makes by far the highest contribution.

5.5 Recommendation

Manual computation in anyway is not as easy as the computer-applied method. Once data has been imputed into the Data Editor of the SPSS, results are computed for analysis in matter of fractions after issuing the necessary command without mistakes.

Multiple regression is preferred to simple regression, this is because estimates made using the multiple regression are closer to the dependent variables compared to their corresponding linear regression.

Reference

damu S.O. (1985) : STATISTICS FOR BEGINNERS
Second Edition
Evans Brother (Nigeria Publisher)
LTD

attely I.G Brokes C.J, And Loxton .S.M (1979): FUNDAMENTAL OF MATHEMATICS AND STATISTICS
Published By Bro
(Norwich) Ltd.

enjamin Fruchter (1987):
uilford J.P
FUNDAMENTAL STATISTICS IN PSYCHOLOGY AND EDUCATION
Third Edition
Printed By B& Jo Enterprises Pte
Ltd Spore

ank J.Kohout (1974) STATISTICS FOR SOCIAL SCIENTISTS
Published By John Wiley & Sons, Inc

arry Franck & (1995):
teven C.Althoen
STATISTICS
Concept And Application
Cambridge University Press

onald J, Guilford (1986):
omas H.Wonnacott
INTRODUCTORY STATISTIC
Third Edition

rry, Sinich (1990): STATISTICS BY EXAMPLE
4th Edition
Dellan Publishing Company,
Sanfranscisco.

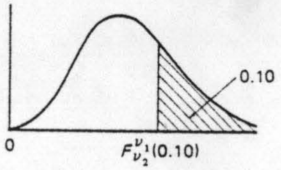
chel S.Kramer (1984): CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS
2nd Edition
Springer-Verlag Publication

	case	english	maths	physics	chemisty	biology	geograpy
1	1	72	51	58	53	48	68
2	2	68	62	60	60	50	59
3	3	58	48	49	48	51	56
4	4	75	74	51	68	56	70
5	5	71	42	40	50	47	62
6	6	61	47	43	32	47	72
7	7	67	43	40	49	48	77
8	8	62	54	40	42	52	62
9	9	71	40	49	37	54	51
10	10	57	39	46	45	48	49
11	11	64	51	51	37	0	62
12	12	66	50	50	50	27	51
13	13	56	48	49	53	48	65
14	14	53	59	50	41	31	42
15	15	69	53	40	41	40	52
16	16	51	50	49	40	30	49
17	17	68	44	40	43	60	76
18	18	68	44	43	39	44	75
19	19	57	53	40	41	46	39
20	20	61	46	40	36	29	63
21	21	52	34	40	38	50	0
22	22	60	38	40	37	57	63
23	23	51	45	41	36	43	39
24	24	66	38	43	29	28	43
25	25	48	40	40	31	45	34
26	26	63	52	40	26	41	37
27	27	68	40	0	41	44	56
28	28	56	38	43	36	32	31
29	29	49	37	40	28	52	42
30	30	68	38	44	35	37	45
31	31	49	40	40	25	21	44
32	32	48	36	40	30	40	39
33	33	50	40	40	36	25	49
34	34	46	40	40	30	25	28
35	35	56	43	40	33	43	50
36	36	52	32	40	34	34	45
37	37	57	37	40	26	48	33
38	38	49	40	40	28	22	51
39	39	45	41	40	20	20	34

	case	english	maths	agric	econs	biology	chemist
1	1	74	56	84	90	63	85
2	2	71	64	76	88	66	64
3	3	72	57	61	82	54	59
4	4	65	44	48	54	49	48
5	5	62	57	52	63	34	69
6	6	59	37	69	89	43	56
7	7	39	41	43	57	46	70
8	8	73	34	55	63	42	43
9	9	61	43	63	49	29	41
10	10	73	30	45	60	32	43
11	11	58	28	55	56	29	52
12	12	57	48	41	70	38	72
13	13	63	34	50	53	27	54
14	14	48	37	50	70	29	45
15	15	71	29	72	53	21	48
16	16	74	48	65	61	21	39
17	17	51	27	26	59	25	46
18	18	44	28	40	48	5	46
19	19	52	38	30	52	32	61
20	20	59	28	36	51	34	50
21	21	61	40	68	62	12	39
22	22	40	23	0	62	42	30
23	23	68	35	41	44	30	44
24	24	52	39	54	44	16	84
25	25	63	30	61	45	8	36
26	26	40	24	45	40	34	34
27	27	42	33	41	58	28	48
28	28	45	25	44	35	31	44
29	29	49	24	40	42	17	30
30	30	54	21	0	46	16	30
31	31	44	29	42	64	31	38
32	32	30	13	41	65	13	16
33	33	44	30	60	42	18	25
34	34	50	26	35	55	31	32
35	35	40	36	35	63	28	33
36	36	33	32	41	44	17	32
37	37	44	27	40	47	21	15
38	38	41	24	36	42	18	28
39	39	59	26	0	27	0	15

es tabulated are $F^{\nu_1}_{\nu_2}(0.10)$, such that
$> F^{\nu_1}_{\nu_2}(0.10)) = 0.10$, where ν_1 is the degrees
om in the numerator and ν_2 is the degrees
om in the denominator.
ower percentage points of the distribution are
d using the relationship

$$F^{\nu_1}_{\nu_2}(0.90) = 1/F^{\nu_2}_{\nu_1}(0.10)$$



✓

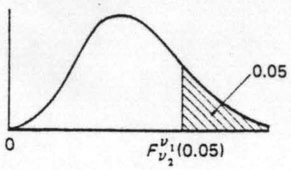
×

1	2	3	4	5	6	7	8	9	10	15	20	40	60	120	∞
86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	61.22	61.74	62.53	62.79	63.06	63.33
53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.42	9.44	9.47	9.47	9.48	9.49
54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.20	5.18	5.16	5.15	5.14	5.13
54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.87	3.84	3.80	3.79	3.78	3.76
06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.24	3.21	3.16	3.14	3.12	3.10
78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.87	2.84	2.78	2.76	2.74	2.72
59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.63	2.59	2.54	2.51	2.49	2.47
46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.46	2.42	2.36	2.34	2.32	2.29
36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.34	2.30	2.23	2.21	2.18	2.16
29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.24	2.20	2.13	2.11	2.08	2.06
23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.17	2.12	2.05	2.03	2.00	1.97
18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.10	2.06	1.99	1.96	1.93	1.90
14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.05	2.01	1.93	1.90	1.88	1.85
10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.01	1.96	1.89	1.86	1.83	1.80
07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.97	1.92	1.85	1.82	1.79	1.76
05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.94	1.89	1.81	1.78	1.75	1.72
03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.91	1.86	1.78	1.75	1.72	1.69
01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.89	1.84	1.75	1.72	1.69	1.66
99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.86	1.81	1.73	1.70	1.67	1.63
97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.84	1.79	1.71	1.68	1.64	1.61
96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.83	1.78	1.69	1.66	1.62	1.59
95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.81	1.76	1.67	1.64	1.60	1.57
94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.80	1.74	1.66	1.62	1.59	1.55
93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.78	1.73	1.64	1.61	1.57	1.53
92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.77	1.72	1.63	1.59	1.56	1.52
91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.76	1.71	1.61	1.58	1.54	1.50
90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.75	1.70	1.60	1.57	1.53	1.49
89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.74	1.69	1.59	1.56	1.52	1.48
89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.73	1.68	1.58	1.55	1.51	1.47
88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.72	1.67	1.57	1.54	1.50	1.46
84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.66	1.61	1.51	1.47	1.42	1.38
79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.60	1.54	1.44	1.40	1.35	1.29
75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.55	1.48	1.37	1.32	1.26	1.19
71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.49	1.42	1.30	1.24	1.17	1.00

The values tabulated are $F^{\nu_1}_{\nu_2}(0.05)$ such that
$\Pr(F^{\nu_1}_{\nu_2} > F^{\nu_1}_{\nu_2}(0.05)) = 0.05$, where ν_1 is the degrees
of freedom in the numerator and ν_2 is the degrees of
freedom in the denominator.

The lower percentage points of the distribution are
obtained using the relationship

$$F^{\nu_1}_{\nu_2}(0.95) = 1/F^{\nu_2}_{\nu_1}(0.05)$$



✓

×

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	15	20	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	245.9	248.0	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.18	2.10	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.06	1.97	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.03	1.94	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.75	1.66	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.39	1.32	1.22	1.00