A COMPUTERISED APPROACH TO REGRESSION AND CORRELATION ANALYSIS:

A Case Study of FCT Agriculture and Rural Development Department, Abuja

BY

BABAGANA ISAH KUPE
PGD/MCS/2003/2004/1144

A Project Presented To:
Department of Mathematics/Computer Science,
School of Science and Science Education,
Federal University of Technology, Minna,
Niger State.

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF POST GRADUATE DIPLOMA (PGD) COMPUTER SCIENCE.

NOVEMBER, 2004

41c

APPROVAL PAGE

We, the undersigned certify that this work was carry out by BABAGANA ISAH KUPE of Department of Mathematics/Computer Science, Federal University of Technology, Minna.

……………………..
Dr. N. I. Akinwande
Project Supervisor

………………………..
Date

………………………..
Mr. L. N. Ezeako
Head of Department

……………………………….
Date

…………………………
External Examiner

………………………………
Date

DEDICATION

With gratitude to Almighty Allah, this project is dedicated to my late father who unfortunately did not live to see the completion of my study. Your contribution towards my knowledge will forever be appreciated.

ACKNOWLEDGEMENTS

In the course of my academic carrier a number of people who make real and solid contribution and make the entire task possible and whose names must not go unmentioned.

First and foremost my gratitude to Almighty Allah for guiding me through the period of writing this project work.

I would like to express my appreciation and vast indebtedness to my my supervisor Dr. N. I. Akinwande for without whom this project work may never see the light of day. My HOD Mr. L. N. Ezeako, my able course co-ordinator Mallam U. Y. Abubakar and to the entire lecturers of Math/Computer Science Department for giving me the necessary insight in my course of study.

My profound gratitude goes to my parent who got me started in live and for letting me know the true value of education, and my lovely wife and darling children for their constant love, encouragement and understanding during the busy period of writing this project work. I cannot find words to describe my heart felt appreciation to my colleagues who contributed in one way or the other and read through the manuscripts. I am delighted to acknowledge my debts to then all.

Finally, my special thanks goes to the entire FCT Executive Management for give me this opportunity to under go this one year computer training, I'm highly grateful. The entire staffs of Department of Planning, Research and Statistics of MFCT, Abuja for taking pains in carry along my schedule in the office during my one years of study leave.

ABSTRACT

This project studies Regression and Correlation analysis with a view to developing a computer program to perform all necessary computation and empirical assessment of the regression model.

The computer program is developed with the uses of Basic programming language. The data is collected on the Rice production, Rainfall and Temperature recorded at FCT Department of Agriculture and Rural Development. Multiple regression models are built to enable the future prediction of the quantity of rice production that may occur as effected by rainfall and temperature.

The Test for the significance of the parameters of the model is done using the analysis of variance. Correlation coefficient is used to test the strength of linear relationship that exists between the Rainfall and quantity of Rice production for the period under study. It is also used to determine the strength of relationship between the quantity of rice and temperature recorded for the period.

The algorithms of the program are the flowchart.

CHAPTER ONE

GENERAL INTRODUCTION

1.1 STATISTICAL DATA ANALYSIS

It is obvious that not every one today is a statistician, but every one, almost without choice, must be a consumer of statistics, modern societies survive on data, every business must know that the characteristics of its client; intelligent consumers needs facts about competing products; in formed voters must have statistics on economic trend voting record and population characteristics, sport fans, every where one master of data consumption.

The scientific community veils on data even more, while the physical scientist is recording the readings of dials and gauges, the social scientist is counting people in the several categories, altering conditions under which people perform various events, and assessing the impact of the alterations. All these activities produce data, which must be put into some kind of order before they yield the information hidden in these masses of numbers.

DATA COLLECTION

In conduction any statistical work, all relevant data must be collected. Obtaining numerical data is as much a part of a statistical analysis itself. Data are required to answer a variety of questions on many spheres of life. The collection of data is a very vital aspect of statistics, for misreporting of fact can be tragic or bias, which arises in data collection, will be reflected in the conclusion drawn.

Data used in statistical study are termed " primary or secondary" according to whether they were collected specifically for the study in question or not. Data use in a statistical study are said to be "primary" when the data are collected under the control and supervision of the person making this particular study, while on the other hand, data not collected under the supervision of the statistical investigator doing the study are termed "secondary data". The data used in this study are secondary in nature.

DATA SUMMARY

The most widely used method of summarising data is organising the data into a frequency table. This is also the process by which statistician present data in an orderly manner so that they are easily comprehended. These reduce and simplify the details into such a form that the important features become vivid and are easily interpreted.

DATA DESCRIPTION

Methods employed to disrobe data are broadly divided into two parts:

(i) Measure of central Tenderly; these include the mean, mode and median.

(ii) Measure of dispersion; These include the range, variance, standard deviation and mean deviation

MATHEMATICAL MODELS

Statistical procedures are derived from mathematical models of what is presumed to be reality, for example grading on a curve. This method is based on the belief that human talents are distributed normally. Many people cluster around average. A few are much above average and a similar sized group are very much below average.

Mathematician then fit to this model of the world, a curve showing how proportion of the population will vary with talent level. Statistician in hurl exploits this curve in describing groups of people and in making references about population.

Statistician use a variety of mathematical models of the world phenomena as foundation for developing their methods and procedures, one of such procedures are correlation and regression and of this project work.

1.2 COMPUTER AND DATA ANALYSIS

The management challenge of huge social data files, especially census data, stimulated the development of several major innovations in computing, automated data processing started from 1890 U.S census, when the punched cards and tabulating machines invented by **HERMAN HOLLEROITH** were first used.

Hollenviths punched card system provided the foundation for electronic data processing **(EDP)** beginning with a unit record, Electro-Mechanical Machines (EMM).

Computers have dramatically affected statistical analysis of data. In fact the impact has been so great that it is difficult to imagine performing an adequate statistical analysis without the uses of the most active areas of continues change and development. Statistical packages are expanding to address more of the task of data manipulation, storage and presentation.

The look and feel of such programs is becoming more sophisticated with greater possibilities for user interaction, impressive graphic display and the beginning of machine intelligence.

1.3 HISTORICAL BACKGROUND

The Federal Capital Territory, Abuja was created in 1976 by Decree No. 6 of that year which also established the Federal Capital Development Authority (FCDA) as the Federal Government Agency responsible for the design, construction and the Administration of the Territory. The Decree also vested ownership and control of all land in the Territory in the Federal Government of Nigeria.

In 1985, by virtue of Decree No. 12 of the same year, the MFCT came into being as the supervising Ministry of the FCDA. The Ministry has a number of departments, which are equivalent to Ministries in the Sates. These departments are Agriculture, Education, Engineering, Finance, Health, Land, Panning and Survey, Planning, Research and Statistics e.t.c.

The Territory, which is located at the centre of the country, covers on area of 8,000 square kilometres. It is, therefore, more than twice the land area of Lagos State (3,535 sq. km) and about two thirds of that of Imo-State (13,032 sq. km).

CLIMATIC CONDITION OF FCT

In the Federal Capital Territory, the duration of sunshine ranges from 6 to 8 hours per day in the south, and 8 to10 hours in the north from January to April/May. There is usually a step down to a mean of about 4 hours per day in the Mouths of July/ August due largely to increase in cloud cover. It start to rise again in September as a result of decrease in cloud cover.

The Federal Capital Territory like most parts of the country records its higher temperatures during the dry season months, which are generally cloudless. During the dry season, the typical months of which is March, temperature could be as high as 37^0c in the southwest, and about 30^0c higher in the Northeast. This period is characterized by high diurnal range when drops of about 170c may be recorded between the highest and lowest temperature in a day. During the highest and lowest temperature drop considerably as a result of dense cloud cover. The diurnal range also drops to about 70c especially July and August.

Rainfall starts from March in the southern parts of the Territory and ends around October in the northern parts and November in the extreme south. The duration of the raining season therefore varies from 240 days in the southern parts to 290 days in the northern parts. The annual total is in the range of 1100mm to 1600mm. It exhibits a spatial parthen, showing an increase from the southwest to the northeast. The mean monthly distribution of rainfall shows a tenderly forwards a high concentration in the three months of July, August and September.

FCT AGRICU LTURE AND RURAL DEVELOPMENT.

Agriculture continues to be the dominant economic activity in the FCT despite the environments development of the recent times. Abuja has the advantage of being able to sustain all the crops grown and livestock raised in the southern forest belt and the northern grassland, because of it location.

The centre of country. This faet was not lost on the founding fathers of the city when the master plan allocated over 60 percent of the territorys land area of 8, 000 sq . km for Agriculture and forest reserves.

The major food crops grow in the FCT include yam, maize, rice, guinea corn, beans and millet. Fishing activities are also prominent among the Bassa people and villagers along rivers of Usuma, Jabi and Guara.

The Department of Agriculture and Rural Development is established to stimule agricultural productivity by providing extension services to small scale farmers, the Department of Agriculture and Rural Development is also discharges the following statutory functions.

(i.) Ensuring timely provision of farm imputs such as fertilizer, agrro-chemicals, Veterinary drugs, agricultural machines and equipment.

(ii.) Development and establishing grazing lands to entance ruminant production (cattle, sheep and goats) and folder banks for livestock all the year round.

(iii.) Providing incetives for investors to establish agro-based industries and to enhance backward integration of pioneer industries and the FCT.

(iv.) Development argo-forestry and making it an important component of system mixed farming.

(v.) Evolving a means of promoting and organisation farming.

To be ble to carry out the above functions the department is organised into six distinctive divisions namely; Planning, Research and Statistic, Agricultural serices, livestock services, foresty, fisheries and co-operatives services.

Given the important of weather to agriculture the FCT has established seven agrometeorogical stations for the collection and collation of data on rainfall, Temperature, humidity, wind, direction, sunshine and atmosphere pressure. The stations are located at Abaji,Gwagwalada, Kwali, Yaba, Karshi, Kuje respectively.

The need of adequate and timely information is imperative for policy formulation and strategic planning for sustainable development of a modern society. In the case of the FCT, the fast and ever changing physical, demographic, Social and economic facets of the territory require thorough understanding of interface of economic variables from time to time before any strategy for the territory or any section of it can be realized. In other to achive this in the area of agriculture in FCT, the division of Planning, Research and Statistics of Department of Agriculture and Rural Development collected various data on Agriculture such as Rainfall, Temperature,humidity, production of Rice, beans etc.

FCT Department of Agriculture and Rural Developement has challege of huge data files, these data are usually been analyse manually which cause delay in producing results. This project will be of benifit to the Department, it is necessary to apply computer to analyse this huge data files to achive very high degree of reliability and correctness. Computer packages are expanding to address more of the task of data manipulation, storage and presentation.

1.4 AIM AND OBJECTIVES OF THE PROJECT.

The main aim of this project study is to develop a computer program to perform all necessary computation and the empirical assessment of the Regression and Correlation analysis. The developed program will be used to achieve the following objectives.

(i) To build the multiple regression line for the data which is given by $Y = b_o + b_1 x_1 + b_2 x_2$, where Y represent the Rice production rate, x_2 represent the Tempe ration recorded and x_1 represent the amount of rainfall.

(ii) To Perform the T-test for the multiple regression co-efficient, to test for the significant of bo, b_1, b_2.

(iii) To ascertain the existence of correlation between the production of rice, Amount of Rainfall and Temperature recorded from 1994 – 2003.

(iv) To testing for the significant of the parameter estimates in the multiple regression analysis using the analysis of variance (ANOVA).

1.5 <u>SCOPES AND LIMITATION</u>

The project is based on the application of computer programming to regression and correlation analysis **(A CASE STUDY OF FCT AGRICULTURE AND RURAL DEVELOPMENT DEPARTMENT ABUJA).** It is on the analysis of rice production, Rainfall and Temperature from the qualities of rice produced as it affected by rain and temperature.

CHAPTER TWO

2.1 REGRESSION AND OTHER EXTENSIONS OF THE SIMPLE LINEAR REGRESSION MODEL

There are various econometric methods that can be dervice estimates of the parameters of economic relationship from statistical observation. In this project we shall extend the simple linear regression model to relationship with two explanatory variables. Firstly, we shall develop some pratical rules for the derivaton of the normal equations for models including any number of the variables. Secondly, we shall examine the extension of the two variable model to non-linear relationship.

MODEL WITH TWO EXPLANATORY VARIABLES

We shall ilustrate the three variable model with given important to quantity of rice produced. The theory that the quantity of rice yeild (Y) depends on the rainfall (X_1) and Temperature (X_2).

$$\text{i.e.}\quad Y = F(X_1, X_2)$$

given that the theory does not specified there mathematical form of the important function, we strat our investigation by assuming that the relationship between Y, X_1,X_2 is a linear.

$$Y_i = b_o + b_1X_{1i} + b_2X_{2i} + U_i$$

where $_i$ = 1, 2, 3, - - - - - - - - - - - - -,n.

This is an exact relationship whose meaning is that the variation in the role in which the two variables X_1 and X_2 are fully explained by changes in amount of rainfall and temperature, if this form were true observation on Y_1, X_1, X_2 would determine a which would lie on a plane.

However, if we gather observation on these variables during a certian period of time and plot them on a diagram we will observed that not all of them lie on plane, some lies on it, but others lies above or below it. This dispersion is due to various factors these are:-

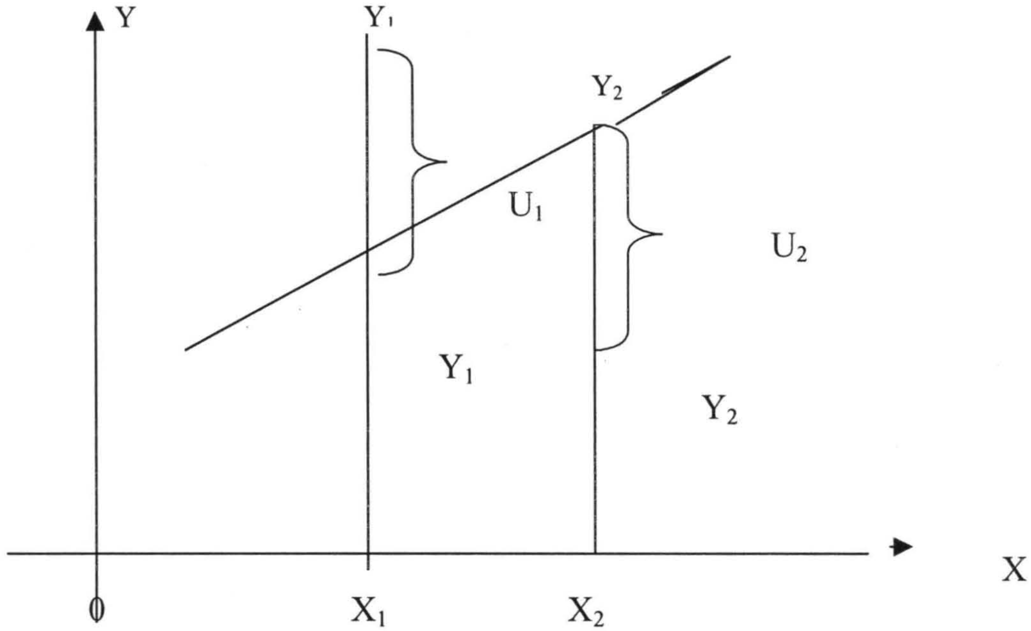
(i.) Omitted from the function and other types of error which are misinterpreted.

(ii.) Random bahviour of beings. The dispersion of points arounds the line may be attributed to an erratic element which inherent in human behaviour. This can be related methods of planting and weeding.

(iii.) Imperfect specification of the mathematical form of the model. We may have linearised a possible non-linear relationship.

(iv.) Errors of aggression: We often use aggregate data in which we add magnitudes referring to each months whose behaviour is dissimilar.

(v.) Systematic Error. The deviation of points from the lines may be due to errors of measurement of the varibles, which are inevitable due to the methods collections and processing statistical information



The true relatiopnship which connects the variables is split into two parts; a part represented by line and a part represented by random term (μ). To copmplete the specification our simple model we need some assumptions about the random variable (Y). The assumption include,

(i.) The randomness of μ. That is the variable μ is a real random variable.

(ii.) Zero mean value for each $X_{i.}$

(iii.) Normality of μ. The values of each (μi) are normally distributed.

μi ------------------N(Ui Xi) = 0

Having specified our model we next use sample observation on Y_1, X_1, and X_2 and obtain estimates of the true parameters $b_{0,}$ b_1, and b_2.

i.e. $Yi = b_0 + b_1X_1 + b_2X_2$.

Where b0, b_1, and b_2 are the estimates of the true parameters b0, b_1, and b_2 of the relationship. The estimate will be obtained by minimizing the sum of squared residuals.

$$\sum_{i=1}^{n} e^2 = \sum^{n}(y_i - y_i)^2 = \sum_{i}^{n}(Y_i - bo - b_1x1_i - b_2x2_i)^2$$

A necessary condition for this expression to assume a minimum value is that its partial derivatives with respect to b_0, b_1 and b_2 be equal to zero.

$$\delta(Y_i - b_0 - b_1X_{1i} - b_2X_{2i})^2/\delta b_0 = 0$$

$$\delta\Sigma(Yi - b_0 - b_1X_{1i} - b_2X_{2i})^2/\delta b_i = 0$$

$$\delta\Sigma(Yi - b_0 - b_1X_{1i} - b_2X_{2i})^2/\delta b_2 = 0$$

Performing the partial differentiations we get the following system of the normal equations in the three unknown parameters b_0, b_1 and b_2

$$\Sigma Y_i = nb_0 + b_1\Sigma X_{1i} + b_2\Sigma X_{2i}$$

$$\Sigma X_{1i}Y_i = b_0\Sigma X_{1i} + b_1\Sigma X_{1i} + b_2\Sigma X_{1i} X_{2i}$$

$$\Sigma X_{2i} Y_i = b_0\Sigma X_{2i} + b_1\Sigma X_{1i} X_{2i} + b_2\Sigma X_{2i}$$

From the above equations we determine the valves for b_0, b_1, and b_2. Which can be estimated as

$b_0 = Y - b_1X_1 - b_2 X_2$

$b_1 = \Sigma(X_{1i}Y_i)(X_{2i}) - (\Sigma X_{2i}Y_i)(\Sigma X_{1i}X_{2i}) / (\Sigma X_{1i}\Sigma X_{2i}) - (\Sigma X_{1i}X_{2i})$

$b_2 = (\Sigma X_{2i}Y_i)(\Sigma X_{1i}) - (\Sigma X_{1i}Y_i)(\Sigma X_{1i}X_{2i}) / (\Sigma X_{1i})(\Sigma X_{2i}) - (\Sigma X_{1i}X_{2i})^2$

2.2 SQUARED MULTIPLE CORRELATION COEFFICIENT.

When the explanatory are more than one we talk of multiple correlation. The square of the correlation coefficient is called the coefficient of multiple determinations or squared multiple correlation coefficients, it is denoted as r^2 this shows the percentage of the total variation of rice yield (Y) explained by the regression plane that is by changes in the rainfall (X_1) and temperature (X_2).

$r^2y. X_1,X_2 = \Sigma Y^2 = \Sigma(Y-Y)^2 \Sigma Y^2 = \Sigma(Y-Y)^2$

$= 1 - \Sigma e^2/\Sigma Y^2$

$$= \frac{\Sigma Y^2 - \Sigma e^2}{\Sigma Y^2}$$

we established that $e_i = Y_i - Y_i$ and $Y_i = b1X_{1i} + b_2X_{2i}$. The squared residuals are $\Sigma e^{2i} = \Sigma e_i (Y_i - Y_i)$. By substituting in the formula of $r^2y.X_1X_2$ we get

$r^2y. X_1,X_2 = \Sigma Yi^2 - b_1\Sigma Yi X_{1i} - b_2\Sigma Yi X_{2i} / \Sigma Yi^2$

$= b_1\Sigma Yi X_{1i} + b_2\Sigma Yi X_{2i} / \Sigma Yi^2$

The value of r^2 lies between 0 and1, the higher r^2 the percentage of the variation of Y explained by regression plane, that is the better the goodness of the regression plane to the sample observation the closer r2 does not take into account the loss of freedom of the introduction of additional explanatory variables in the function.

2.2 PARTIAL CORRELATION COEFFICIENT

A partial correlation coefficient measures the relationship existing between any two variables when all other variables connected with those two are kept constant. The partial correlation coefficient is determined in terms of simple correlation coefficient among the variable involved in a multiple relationship. Partial correlation coefficient is obtained by using the formula rzy.x = rzy –rzx ryx /√(1-r2zx)(1-r2yx) where z, y, x are variables and (r) is the computed values of simple correlation coefficient.

2.3 VARIANCE OF THE PARAMETER ESTIMATES

The estimate b0, b1 and b2 are unbiased estimates of the true parameters of the relationship between Y, X1, and X2.

1. **The standard error test: -** we print the standard errors (sb1=√vari (bi)) underneath the respective estimate and compare them with the numerical values of the estimates.

(a.) If S (bi) ≥ ½bi we accept the hypothesis, that is we accept that the estimate bi is not statistically significant at 5% level of significant.

(b.) If S (bi) < ½bi we reject the null hypothesis in other words we accept that our parameter estimate is statistically significant at 5% level of significant for a two tail test.

2. **The student's test** of the null hypothesis we compute the t-ratio for each bi.

$$t^* = bi/s(bi)$$

With n-k degree of freedom the theoretical values of t at the chosen level of significance are the critical values that define the critical region in a two-tail test with n-k degree of freedom.

(a.) If t* falls in the acceptance region, we accept null hypothesis, that is we accept that bi is not significant at 5% level of significance and hence the corresponding regress or does not appear to contribute to the explanation of the variation in Y.

(b.) If t* falls in the critical region, we reject the null hypothesis and we accept that alternative one; bi is statistically significant clearly the greater the value of t* the stronger is the evidence that bi is significant.

2.4 TESTING THE OVERALL SIGNIFICANCE OF A REGRESSION

This test has been explained in the proceeding section for the multiple regressions. The test aims at finding out whether the explanatory variables (X_1,

X_2 - - - - - - -, X_k) do actually have any significant of the regressions implies testing the null hypothesis.

$$H_0 : b_1 = b_2 = b_3 = 0$$

against the alternative hypothesis.

H1: not all bi's are zero

If the null-hypothesis is true, that is if all the true parameters are zero, there is no linear relationship between Y and the regressors. The test of the over all significance may be carried out with the table of analysis of variance. We compute the regression of Y on X_1, and X_2 and we estimate.

(a.) The total sum of square deviation of the Y's ΣY_2

(b.) The sum of square deviation explained by all the regressors (X_1, X_2) ΣY_2

(c.) The sum of residual deviation Σe^2 from these terms we can evaluate the expression $\Sigma Y^2 = \Sigma Y^2 + \Sigma e^2$ we next find the degree of freedom for each of the terms of the identity. The degrees of freedom for Σy^2 is K-1 where K=(K+1) is the total number of $b^{'s}$ including the constant intercept. The degree of freedom for Σe^2 is N-K.Where N is the sample size.

The degrees of freedom of the total sum of squares is (k-1)+(N-K)=N-1 with this information we may computes F* ratio as

$$F^* \frac{\Sigma Y^2/(K-1)}{\Sigma e^2/(N-K)}$$

Which is composed with the theoretical F (at the chosen level of significance) with V1=N-K degrees of freedom if F*<f we accept the over all regression is not significant.

The above information is summarized in the table of analysis of variance (ANOVA) below.

ANOVA TABLE

Source of Variation	Degree of Freedom	Sum of Squares	Mean Sum of Squares	Fc
Replicate sum of squares ($SSRx_1$)	b – 1	SSR	MSR	MSR/MSE
Treatment Sum of Squares ($SSTx_2$)	t – 1	SST	MST	MST/MSE
Residual Sum of Squares (SSE)	(b - 1)(t - 1)	SSE	MSSE	
TOTAL	bt – 1	TSS		

CHAPTER THREE

PROGRAM DESIGN

3.1 DESIGN ELEMENT

The design of a new system can be conveniently divided into the following elements.

INPUT:- Considering of input will be influenced greatly by the needs of output e.g the necessity for quick response from the system would determines the need for an online type of input. In this project consideration would be given to design of input layouts.

OUTPUTS:- It is necessary to consider what is required from the system before deciding how to set about producing it. These requirements will have become clear as this project progressed.

FILES:- This element is very much linked to input and output. Input is processed against the files to produce the necessary output.

PROCEDURES:- They are the steps, which unify the whole process, which link everything together to produce the desired output. These will involve both computer and clerical procedure. They will start with the origination with the source document and end with the output document being distributed.

3.2 INPUT DESCRIPTION

A data file is a separate stored in Secondary. It is not attached to any program, but may be used by any number of programs. To avoid confusion each buffer is numbered (often referred to as a file number) and assigning that number is also part of the opening process.

The only time a file number appears in a program is during the opening process. Once the file number is assigned, only that number refers to the file. The opening statement will require you to state the file name, any allowable options and the buffer number (file number) assigned to it.

After finishing using a file at the end of a program the file must be close. The closing process writes the current block of the file permanently into secondary storage, and releases the buffer area for other uses. Failing to close a file will result in the buffered data being lost.

In this project four data files namerly **YRSFILE, RICFILE, RAIFILE AND TEMFILE** were created in order to provide a convenient means of storing data sets. Since data file can easily be read and updated by a basic program. The **YRSFILE** data was created for the actual yearly production of rice in the Federal Capital Territory.

3.3 ALGORITHM REPRESENTATION.

In order to find computer solution to a problem, one must decide steps to be taken by the computer. This involves working algorithm for the problem. The algorithm is them presented in any of the acceptable forms.

Although there exist other forms (e.g pseudocodes, N-S diagrams etc) of presenting algorithms. In this project flowchart is employed as it provides good visual representation and easy appreciation of the logic of the algorithm its representing. After an understanding of the input and output, a rigorous description of these which covers almost all cases were drawn. The flowchart for data entry procedure and Co-efficient of regression estimate procedure were drawn as shown in Appendix 'A'.

3.4 CHOICE OF PROGRAMMING LANGUAGE

Basic is defined as know as the beginner all purpose symbolic instruction code. Those who are writting a program for the first or those who are new into the program writting mostly use it. It is choosing as the programming language for the following reasons.

i) It has only view concepts to learn and digest

ii) It has a design that facilitates the writing of programs in a style that is accepted as a good programming pratice.

iii) It is very easy to implement and compile.

iv) It has a good file management facility.

v) Is the most commonly used programming language for mini computers and microcomputers system because of it small interpreter and compiler.

3.5 PROGRAM COMMANDS AND STATEMENTS

In BASIC, you make three types of entries from the keybord these are COMMANDS, DATA and PROGRAM STATEMENTS. A Command is an immediate order for the computer to do something. Once the command is given it is executed by the computer. Command do not constitute part of a program rather, they process programs by performing various tasks related the program. Some commands in Basic program are:

LOADING QBASIC: Before you can enter, change or execute a BASIC program, you must load the Basic intepreter/compiler program into your computer memory. When you have booted the system and see C:/>(DOS PROMPT) on the screen then qbasic is typed and press the Enter key. Basic statement or system command from the keyboard con only be accepted if the Basic interpreter/compiler has been loaded into the memory.

LIST COMMAND:- To see all of the program statements that have been entered, we type the LIST COMMAND. The list command has the following variations

LIST	list the entire program
LIST 10	list statements with statement number 10
LIST 10 – 200	list from line 10 – 200
LIST –200	list from the beginning to line to line 200
LIST 200 -	list from 200 to the end of the program

The LIST direct the listing to a computer printer instead of the screen.

DELETE COMMAND: The delete command erases whole sections of a program from main memory. It has the same variation as for LIST command. DELETE alone erases the entire program from the main memory.

SAVE: Main memory is volatile, i.e as soon as you get out of BASIC enviroment or when the computer is turned off, the program in main memory disappears. To save the current program to a disk (secondary storage media) we uses the save command.

RUN COMMAND: Once a program has been enter into the memory of the computer for us to follow the instruction the program statement will have to be executed. The command for program execution is RUN.

BASIC STATEMENTS

A BASIC program centres on statements, each statement is written in conformity with the language syntax the pattern or structure of the word order. The order in which the statements are written is of vital importance, since the computer executes a program by embodying the ingredients of sequence. i.e. one statement after the other starting from the first selection, i.e, it skips a sequence of statements; and iteration which implies that it is directed by a loop to repeat a sequence of instructions.

BASIC statements are classified into two namely; executable and non executable statements. An executable statements command the computer to perform some actions or simply to take note of asomething. A non executable statement on the other hand, does not require the computer to take any action. Below is the function of Qbasic statement use in the program.

REM statement (short of Remark) are comment or remark statement that help you to understand your program. You can use Rem at any point in time inside your program to explain what you are doing inside the program or when you feel like giving your program a tittle. Its use is not compulsory but it is advisable to use it in order to describe what you are doing.

INPUT statement enable computer to transfer data or values to variable names. The values could be assigned to variables within the program or inputted while the program is running.

PRINT statements enable computer to display or print a constant or the content of a variable name. In basic the print statement is in the form of using output statements.

LOCATE functions allow the movement of a cursor row and column wise and

DIMENSSION statements is used to inform the computer to reserve cells for the element of list in the program.

CHAPTER FOUR

PROGRAM IMPLEMENTATION

4.1 PROBLEM SPECIFICATION

The task of this design is to develop a system or program that could accept regression analysis data and generate result on such things as the equation of the model. Correlation matrix and analyse the variance of the regression co-efficient as well. The system should also be able to estimate the correlation co-efficient and test the data

4.2 THE PRATICAL OPERATION OF THE PROGRAM

The program was developed in BASIC programming language to compute the regression and correlation analysis to know if rainfall and temperature has any effect on the production of rice in the FCT. The entire program is divided into various subprograms and this can be illustrates in the following steps below.

STEP 1 CREATING AND STORING OF INPUT DATA FILES

The two lines (10 – 20) stated the title of the program **"USING THE REGRESSION AND CORRELATION ANALYSIS TO ANALYSE A SET OF DATA WRITTEN BY ISAH BABAGANA".** The next two lines (30 –40) dimensioned the variables into their appropriates sizes. These were followed by twenty-one lines (50 – 250) for creating and storing the four input data files.

YRSFILE for the years from 1994 – 2003, RICFILE for the total production of rice for years, RAIFILE for the amount of rainfall and TEMFILE for Temperature recorded for the years.

STEP 2 THE MULTIPLE REGRESSION LINE FOR THE DATA

The multiple regression line for the data is given by $Y = bo + b_1x_1 + b_2x_2$. The line numbers (260 – 650) were meant to compute the multiple regression line for the data. In achieving this, the original data was sorted out and their means were obtained and the values (constant bo, b_1, b_2) were equally computed and regression equation is printed using the appropriate formulae.

STEP 3 TESTING FOR THE SIGNIFICANCE OF THE PARAMETER ESTIMATES USING ANOVA.

The test aims at finding out whether the explanatory variables (X_1, X_2 i.e Rainfall and Temperature) do actually have any significant influence on the dependent variable Y (production of rice). Line numbers 660 – 1240 were meant to test if there is any significance different between the data. To achieved these , the data obtained were sorted using the binary sorting method. We calculate the sum of squares and mean sum of squares using the appropriate formulae. Lastly, a decision was taken at a 5% level of significance.

STEP 4 PERFORMING THE T – TEST FOR THE MULTIPLE REGRESSION COEFFICIENT.

The line numbers 1250 – 1760 is set to test for the significance of the parameters used bo, b_1 and b_2. We estimates the co-efficient of multiple determination R^2. We also compute the sum of squares and means of squares test for the parameter individaully by comparing the computed result with the tabulated value at 5% level of significance, Lastly a decision is taken.

STEP 5 THE CERRELATION CO-EFFICIENTS BETWEEN Y, X_1 AND X_2

The line numbers 1770 – 1960 were meant to find the partial correlation co-efficients between Y, X_1 and X_2. We first compute the r for Yo and X1 (r_{12}), Yo abd X_2 (r_{13}) and X_1 and X_2 (r_{23}) and there relationship were obtained. Line 1970 – 2400 were meant to test for the types of relationship that exist between the parameters using correlation coefficient after which decisions were taken as to which is correlated to one another.

Finally, line number 2410 and 2420 were files created are closed and the END of the entire program. The end statement tells the computer that there is no more statement to execute. The entire program is saved using the name BABAPROJ.BAS and can be accessed at the QBASIC environment. It is equally listed in Appendix 'B'.

4.3 SYSTEM JUSTIFICATION

By hand, the computations necessary to perform a regression analysis are at best tedious, even with the assistance of a hand-operated calculator. Such tedious can be alleviated by the use of a digit computer. The computer only save time and energy but also prevent the inevitable cascade of errors that occur whenever one error such as mispunching the key of a calculator is committed and provides greater level of accuracy. Although there are full-featured statistical package that can handle regression and correlation analysis problem apart from being expensive they imposes heavy hardware requirements. For example statistical package for social sciences (SPSS) require a 16-bit microcomputer 7million bytes of fixed disc for programs and data. Moreover, in most cases only users with sound statistical background can make use of the available features. This program is design with a view to getting rid of these problems.

4.4 OUTPUT EXPECTED/SAMPLE RESULT

The program is expected to output a set of results as follows: -

(i) The mean of Y, X_1, X_2, that is the mean rice production, the mean Rainfall and the mean Temperature recorded for the period under study.

(ii) The Regression model for future forecast and the significance of the parameter of the model.

(iii) The strength of linear relationship that may exist between Amount of rainfall, temperature recorded and quantity of rice production for the period under study.

The results obtained from the execution of the program are given in the sample results at Appendix "C"

CHAPTER FIVE

SUMMARY AND RECOMMENDATION

5.1 SUMMARY AND CONCLUSION

The computerised of Regression and Correlation analysis have been carried out in line with the objective of the study. The earlier chapters dwelled on defining and explains some basic concepts on the application of computer to regression and correlation analysis and some Statistical tools used.

In order to carry out all the necessary computation, the Basic programming language is choose to do the analysis. After the practical operation of the program the following objective have been achieved. The multiple regression line for the data has been build; the parameter b0, b1 and b2 were estimated for the predication equation formulated. The test was conducted to ascertain whether the parameter b0, b1 and b2 are significant. The result for b0, b1 test indicated that their not significant and that of b2 indicate that is significant. To compute partial correlation, simple correlation were computed between Y, X1 and X2 and results suggested the existence of week relationship in the number of rice yield, rainfall and Temperature recorded and the test of significance carried out confirm it.

It was also observed that in the course of this project work, the successful of regression analysis depends heavenly on the person doing the modelling the analyst must select relevant predictor variables and decide on the conclusion reach when testing the adequacy of the regression model.

5.2 RECOMMENDATION

For this system to function well and indeed to perform regression analysis these are the requirement on the type and amount of data needed. All data element must be in numerical form that is real integer.

(i) To obtain estimate of the regression parameter, the number of data points N must exceed the number of the predictor variable.

(ii) If any of the predictor variables must be forecast prior to using regression equation, then the forecast of the variable must be easy to obtain. Modelling y as functions of variable that are themselves difficult to estimate will only decrease the accuracy of the y forecast.

(iii) To perform any diagnostic procedure, we need to have N greater than or N k exceed k by reasonable amount. The large N – k is the more accurate the estimates will be since all internal estimate and hypothesis test in regression analysis are based on Se that is sample standard error.

In view of this it is strongly recommended that further study in to the area of non-linearity in variable and parameters and an extension of the program to accommodate data transformation be carried out. Interested researchers could extend this study to methods not covered in this work. Such effort may also incorporate the uses of some full features of statistical programming packages.

REFERENCES

1. **BIYI AFONJA (1982):** **INTRODUCTORY STATISTICS: Evan Brother Limited, Ibadan – Nigeria**

2. **HAYS, W. L (1973):** **STATISTICS FOR THE SOCIAL SCIENCES: 2nd Edition, Holden Day Ing /Sam Francisco**

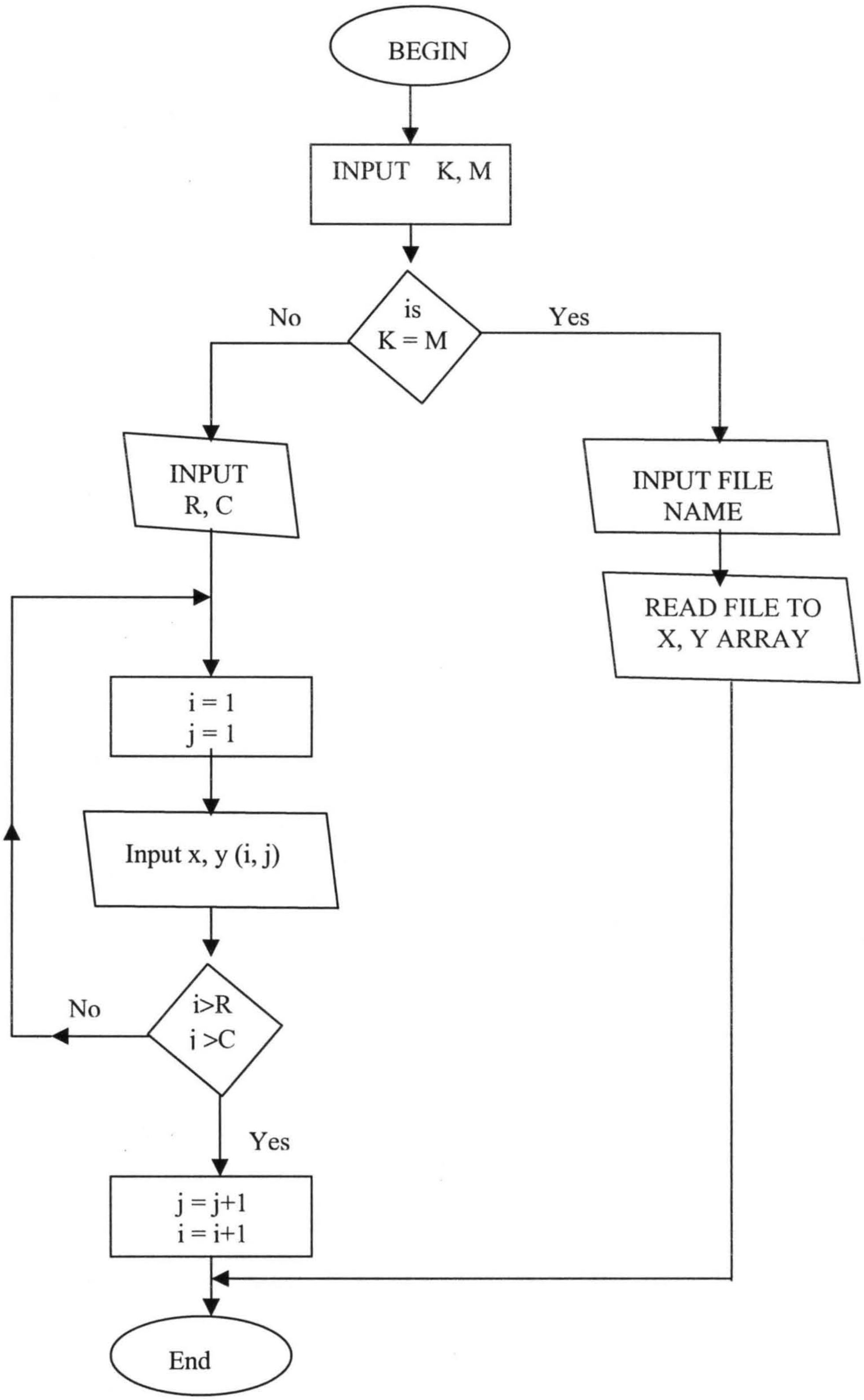
3. **JUSTICE E. O., SUNDAY R.A.& KOLA F. (2003):** FUNDAMENTALS OF COMPUTER **SCIENCE AND TECHNOLOGY: Kay–Em Publishers Owode, Ogbomsho Nigeria**

4. **BADMUS R. O. (2004):** **SYSTEM ANALYSIS AND DESIGN: Department of Maths/Computer Sciences FUT Minna. Unpublished**

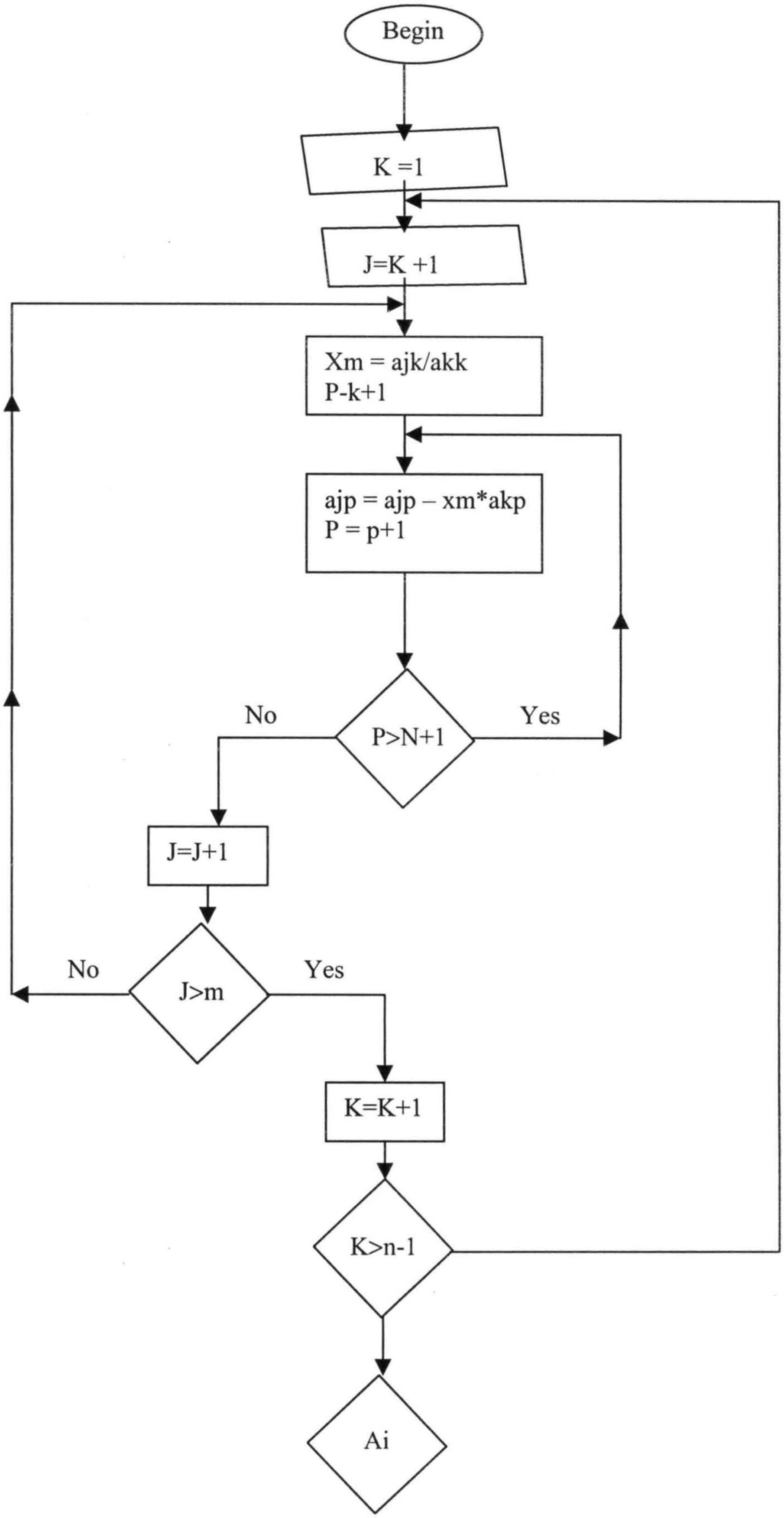
5. **JIYA M. (2004):** **COMPUTER PROGRAMMING (BASIC LANGUAGE): Department of Maths/Computer Science FUT Minna. Unpublished**

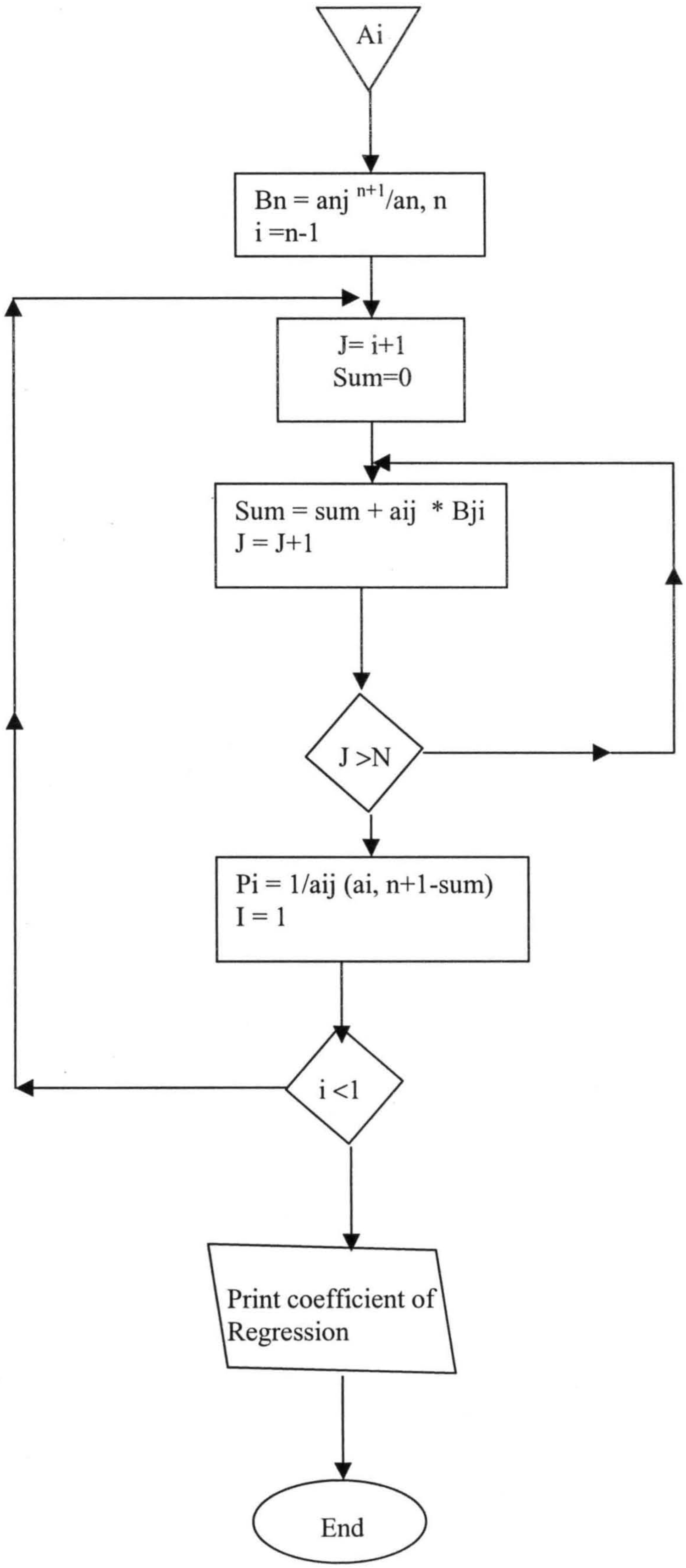
6. **KOLA R. (1971):** **COMPUTER PROGRAMMING: Federal University Of Technology Minna. Unpublished.**

DATA ENTRY PROCEDURE FLOWCHART



COEFFICIENT OF REGRESSION ESTIMATE PROCEDURE




Ai
Bn = anj n+1/an, n
i =n-1
J= i+1
Sum=0
Sum = sum + aij * Bji
J = J+1
J >N
Pi = 1/aij (ai, n+1-sum)
I = 1
i <1
Print coefficient of Regression
End

APPENDIX "B"

```
10 REM THIS PROGRAM USES REGRESSION AND CORRECTION TO ANALISED A SET OF DATA
20 REM IT IS WRITTEN BY ISAH BABA GANA
30 DIM X1(100), Y(100), X2(100), S(100), YR(100)
40 DIM E(100), F(100), G(100), L(100), M(100), N(100)
50 KEY OFF: CLS : LOCATE 12, 35: INPUT "ENTER NUMBER OF THE YEARS"; YN
60 CLS : LOCATE 23, 1: PRINT "ENTER YEARS"
70 OPEN "YEAR" FOR OUTPUT AS #1
80 FOR I = 1 TO YN
90 INPUT YR(I)
100 NEXT I
110 CLOSE #1: OPEN "RICE" FOR OUTPUT AS #2
120 CLS : LOCATE 23, 1: PRINT "READING THE YEARLY RICE PRODUCTION DATA WAIT..."
130 FOR J = 1 TO YN
140 INPUT Y(J)
150 NEXT J
160 CLOSE #2: OPEN "RAIN" FOR OUTPUT AS #3
170 CLS : LOCATE 23, 1: PRINT "READING THE YEARLY RAINFALL DATA WAIT..."
180 FOR k = 1 TO YN
190 INPUT X1(k)
200 NEXT k
210 CLOSE #3: OPEN "TEMP" FOR OUTPUT AS #4
220 CLS : LOCATE 23, 1: PRINT "READING THE YEARLY TEMPERATURE DATA WAIT..."
230 FOR I = 1 TO YN
240 INPUT X2(I)
250 NEXT I
260 REM USING REGRESSION ANALYSIS TO ANALYSE THE DATA IN FILE CREATED
270 REM FIND THE MULTIPLE REGRESSION LINE FOR THE DATA WHICH IS GIVEN BY Y =
bo+b1x1+b2x2
280 SUMX1 = 0: SUMX2 = 0: SUMY = 0
290 FOR I = 1 TO YN
300 SUMX1 = SUMX1 + X1(I): SUMX2 = SUMX2 + X2(I): SUMY = SUMY + Y(I)
310 NEXT I
320 REM LET THE MEANS OF Y, X1 AND X2 BE DENOTED BY A, B, AND C RESPECTIVELY
330 A = SUMY / YN: B = SUMX1 / YN: C = SUMX2 / YN
340 FOR I = 1 TO YN
350 P(I) = Y(I) - A: Q(I) = X1(I) - B: R(I) = X2(I) - C: E(I) = (P(I)) ^ 2: F(I)
= (Q(I)) ^ 2
360 G(I) = (R(I)) ^ 2: L(I) = P(I) * Q(I): M(I) = P(I) * R(I): N(I) = Q(I) *
R(I)
370 NEXT I
380 SUMP = 0: SUMQ = 0: SUMR = 0: SUME = 0: SUMF = 0: SUMG = 0: SUML = 0: SUMM =
0: SUMN = 0
390 FOR I = 1 TO YN
400 SUMP = SUMP + P(I): SUMQ = SUMQ + Q(I): SUMR = SUMR + R(I): SUME = SUME +
E(I): SUMF = SUMF + F(I):
410 SUMG = SUMG + G(I): SUML = SUML + L(I): SUMM = SUMM + M(I): SUMN = SUMN +
N(I)
420 NEXT I
430 W = (SUML * SUMG) - (SUMM * SUMN): V = (SUMF * SUMG) - ((SUMN) ^ 2): S =
(SUMM * SUMF) - (SUML * SUMN)
440 b1 = W / V: b2 = S / V: bO = A - (b1 * B) - (b2 * C)
450 PRINT : PRINT "----------------------------------------"
460 CLS : LOCATE 1, 3: PRINT "WORKSHEET FOR REGRESSION EQUATION Y=bo+b1x1+b2x2"
470 PRINT "-----------------------------------------------------"
480 PRINT TAB(3); "YEARS"; TAB(12); "RICE Y"; TAB(21); "RAIN X1"; TAB(30); "TEMP
X2"; TAB(39); "yi=Yi-A"; TAB(47); "X1i-B"; TAB(57); "X2i=X2i-C"
```

```
490 YR = 1994
500 O = 1
510 FOR I = 1 TO YN
520 PRINT TAB(2); YR(I); TAB(13); Y(I); TAB(23); X1(I); TAB(30); USING "##.##";
X2(I); TAB(40); P(I); TAB(49); Q(I); TAB(59); R(I)
530 O = O + 1
540 NEXT I
542 PRINT
544 PRINT
546 PRINT
550 PRINT TA(3); "(yi)^2"; TAB(16); "(X1i)^2"; TAB(27); "(X2i)^2"; TAB(39);
"yiX1i"; TAB(47); "yiX2i"; TAB(57); "X1iX2i"
560 FOR I = 1 TO YN
570 PRINT TAB(3); E(I); TAB(16); F(I); TAB(27); USING "##.##"; G(I); TAB(39);
L(I); TAB(47); M(I); TAB(57); N(I)
580 NEXT I
582 PRINT "*****************************************************"
584 PRINT ".........SAMPLE RESULT............."
586 PRINT
588 PRINT
590 PRINT "THE MEAN OF Y ="; A
600 PRINT "THE MEAN OF X1 ="; B
610 PRINT "THE MEAN OF X2 ="; C
620 PRINT "b1 ="; b1
630 PRINT "b2 ="; b2
640 PRINT "bo ="; bO
642 PRINT
644 PRINT
650 PRINT "THE REGRESSION EQUATION Y ="; bO; "+"; b1; "X1"; "+"; b2; "X2"
660 REM TESTING FOR THE SIGNIFICANCE OF THE PARAMETER ESTIMATES IN THE
670 REM MULTIPLE REGRESSION USING THE ANALYSIS OF VARIANCE
680 FOR I = 1 TO YN
690 U1(I) = (Y(I)) ^ 2
700 NEXT I
710 SUMU1 = 0
720 FOR I = 1 TO YN
730 U1 = SUMU1 + U1(I)
740 NEXT I
750 SST = SUMU1 - ((SUMY) ^ 2) / YN
760 SUMY2 = 0
770 FOR I = 1 TO YN
780 U2(I) = (Y2(I))
790 SUMY2 = SUMY2 + U2(I)
800 NEXT I
810 SSR = SUMY2 - ((SUMY) ^ 2) / YN
820 SSE = SST - SSR
830 REM FINDING THE ASSOCIATED DEGREESN OF FREEDOM
840 REM LET DF1, DF2,ANDDF3 REPRESENT THE ASSOCIATED
850 REM DEGREES OF FREEDOM FOR SST, SSR, AND SSE RESPECTIVELY
860 k = 2
870 DF1 = YN - 1: DF2 = k - 1: DF3 = YN - 1 * k - 1
880 MSR = SSR / DF2: MSE = SSE / DF3
890 F1 = MSR / MSE
900 PRINT : PRINT "--------------------------------"
910 CLS : LOCATE 1, 3: PRINT "THE TEST FOR SIGNIFICANCE OF THE PARAMENTER
ESTIMATES USING ANOVA"
920 PRINT "-----------------------------------------------"
```

```
930 PRINT "HO:Bi = 0 (THERE IS NO SIGNIFICANCE RELATIONSHIP BETWEEN Y AND X1,
X2)"
940 PRINT "H1: NOT ALL Bi's ARE EQUAL TO ZERO"
950 PRINT "THE LEVEL OF SIGNIFICANCE IS 0.05"
960 PRINT "SST =", SST
970 PRINT "SSR =", SSR
980 PRINT "SSE =", SSE
990 PRINT "DF1 =", DF1
1000 PRINT "DF2 =", DF2
1010 PRINT "DF3 =", DF3
1020 PRINT "MSR =", MSR
1030 PRINT "MSE =", MSE
1040 PRINT "F1 =", F1
1050 F(CRITICAL) = 4.74
1060 PRINT "F(CRITICAL) AT 2 AND 7 DEGREES OF FREEDOM =", F(CRITICAL)
1070 PRINT : PRINT TAB(10); "--------------------------------------"
1080 CLS : LOCATE 1, 3: PRINT TAB(9); "THE ANOVA TABLE FOR THE DATA"
1090 PRINT TAB(9); "-------------------------------------------------"
1100 PRINT
1110 PRINT
1120 PRINT : PRINT TAB(3); "-------------------------------------"
1130 PRINT TAB(3); "SOURCE OF"; TAB(18); "SUM OF"; TAB(33); "DEGREEE OF";
TAB(48); "MEAN"; TAB(63); "F*"
1140 PRINT TAB(3); "VARIATION"; TAB(18); "SQUARES"; TAB(33); "FREEDOM"; TAB(48);
"SQUARE"
1150 PRINT TAB(3); "-------------------------------------------------"
1160 PRINT TAB(3); "REGRESSION"; TAB(17); SSR; TAB(37); k; TAB(47); MSR;
TAB(62); F1
1170 PRINT TAB(3); "ERROR"; TAB(17); SSE; TAB(37); YN - k - 1; TAB(47); MSE
1180 PRINT TAB(3); "--------------------------------------"
1190 PRINT TAB(3); "TOTAL"; TAB(17); SST; TAB(37); YN - 1
1200 PRINT TAB(3); "-------------------------------------------"
1210 PRINT
1220 PRINT
1230 IF ABS(F1) >= F(CRITICAL) THEN LPRINT " WE REJECT HO AND CONCLUDE THAT RICE
YIELD IS SIGNIFICANTLY RELATED TO THE AVERAGE RAINFALL AND AVERAGE TEMPERATURE"
1240 IF ABS(F1) < F(CRITICAL) THEN LPRINT " WE ACCEPT HO AND CONCLUDE THAT RICE
YIELD IS NOT SIGNIFICANTLY RELATED TO THE AVERAGE RAINFALL AND AVERAGE
TEMPERATURE"
1250 REM PERFORMING THE t TEST FOR THE MULTIPLE REGRESSION COEFFICIENTS
1260 REM ESTIMATING THE COEFFICIENT OF MULTIPLE DETERMINATION R^2 DENOTED AS DR
1270 DR = ((b1 * SUML) + (b2 * SUMM)) / SUME
1280 Vb1 = MSE * (SUMG / V): sb1 = (Vb1) ^ .5
1290 Vb2 = MSE * (SUMF / V): sb1 = (Vb2) ^ .5
1300 k2 = (B ^ 2) * SUMG: H2 = (C ^ 2) * SUMF: z = 2 * B * C * SUMN
1310 Vbo = MSE * ((1 / YN) + (k2 + H2 - z) / V): sbo = (Vbo) ^ .5
1320 REM PERFORMING THE t TEST FOR b1
1330 t(CRITICAL) = 2.365
1340 t0 = bO / sbo
1350 PRINT : PRINT "------------------------------------------------"
1360 PRINT "TEST FOR THE SIGNIFICANCE OF bO REPORT"
1370 PRINT "------------------------------------------------"
1380 PRINT
1390 PRINT
1400 PRINT "HO: BO =0"
1410 PRINT "H1: BO IS NOT EQUAL TO ZERO"
1420 PRINT
```

```
) PRINT "Vbo =", Vbo
) PRINT "sbo =", sbo
) PRINT "t0 = ", t0
) IF ABS(t0) >= t(CRITICAL) THEN PRINT "WE REJECT HO AND CONCLUDE THAT bO IS
TISTICALLY SIGNIFICANT AT THE 5% LEVEL"
) IF ABS(t0) < t(CRITICAL) THEN PRINT "WE ACCEPT HO AND CONCLUDE THAT bO IS
 STATISTICALLY SIGNIFICANT AT THE 5% LEVEL"
) t1 = b1 / sb1
) PRINT : PRINT "----------------------------------------------"
) PRINT "TEST FOR THE SIGNIFICANCE OF b1 REPORT"
) PRINT "-------------------------------------"
) PRINT
) PRINT
) PRINT "HO: B1 = 0"
) PRINT "H1: B1 IS NOT EQUAL TO ZERO"
) PRINT
) PRINT "Vb1 =", Vb1
) PRINT "sb1 =", sb1
) PRINT "t1 =", t1
) IF ABS(t1) >= t(CRITICAL) THEN LPRINT "WE REJECT HO AND CONCLUDE THAT b1 IS
TISTICALLY SIGNIFICANT AT THE 5% LEVEL"
) IF ABS(t1) < t(CRITICAL) THEN LPRINT "WE ACCEPT HO AND CONCLUDE THAT b1 IS
 STATISTICALLY SIGNIFICANT AT THE 5% LEVEL"
) t2 = b2 / sb2
) PRINT : LPRINT "------------------------------------"
) PRINT "TEST FOR THE SIGNIFICANCE OF b2 REPORT"
) PRINT "-------------------------------------"
) PRINT
) PRINT
) PRINT "HO: B2 = 0"
) PRINT "H1: B2 IS NOT EQUAL TO ZERO"
) PRINT "Vb2 =", Vb2
 PRINT "sb2 =", sb2
 PRINT "t2 =", t2
 IF ABS(t2) >= t(CRITICAL) THEN LPRINT "WE REJECT HO AND CONCLUDE THAT b2 IS
ISTICALLY SIGNIFICANT AT THE 5% LEVEL"
 IF ABS(t2) < t(CRITICAL) THEN LPRINT "WE ACCEPT HO AND CONCLUDE THAT b2 IS
STATISTICALLY SIGNIFICANT AT THE %% LEVEL"
 PRINT
 PRINT "THE COEFFICIENT OF MULTIPLE DETERMINATION DR ="; DR
 REM USING CORRELATION ANALYSIS TO ANALYSE THE SET OF DATA"
 REM FIND THE CORRELATION COEFFICIENTS BETWEEN Y, X1 AND X2
 r12 = SUML / (((SUMF) ^ .5) * ((SUME) ^ .5))
 r13 = SUMM / (((SUMG) ^ .5) * ((SUME) ^ .5))
 r23 = SUMN / (((SUMF) ^ .5) * ((SUMG) ^ .5))
 REM FIND THE PARTIAL CORRELATION COEFFICIENTS BETWEEN Y, X1 AND X2
 r123 = (r12 - (r13 * r23)) / (((1 - (r13) ^ 2) * (1 - (r23) ^ 2)) ^ .5)
 r132 = (r13 - (r12 * r23)) / (((1 - (r12) ^ 2) * (1 - (r23) ^ 2)) ^ .5)
 r231 = (r23 - (r12 * r13)) / (((1 - (r12) ^ 2) * (1 - (r13) ^ 2)) ^ .5)
 RINT : PRINT "-----------------------------------------------"
 RINT " THE CORRELATION COEFFICIENTS REPORT"
 RINT "----------------------------------------------------"
 RINT
 RINT
 RINT " THE CORRELATION COEFFICIENT BETWEEN Y AND X1 DENOTED r12 ="; r12
 RINT "THE CORRELATION COEFFICIENT BETWEEN Y AND X2 DENOTED r13 ="; r13
 RINT "THE CORRELATION COEFFICIENT BETWEEN X1 AND X2 DENOTED r23 ="; r23
```

```
1940 PRINT "THE PARTIAL CORRELATION COEFFICIENT BETWEEN Y AND X1 DENOTED r123
="; r123
1950 PRINT "THE PARTIAL CORRELATION COEFFICIENT BETWEEN Y AND X2 DENOTED r132
="; r132
1960 PRINT "THE PARTIAL CORRELATION COEFFICIENT BETWEEN X1 AND X2 DENOTED r231
="; r231
1970 REM USING THE FISHER's Z-SCORE TO TEST HYPOTHESES FOR THE TRUE VALUES OF
1980 REM PARTIAL CORRELATION COEFFICIENTS02050 LPRINT
1990 PRINT "THE LEVEL OF SIGNIFICANCE OF THE TEST IS .05"
2000 z(CRITICAL) = 1.96
2010 Zr123 = .5 * LOG((1 + r123) / (1 - r123)): Zr123 = .5 * LOG((1 + r132) / (1
- r132))
2020 Zr231 = .5 * LOG((1 + r231) / (1 - r231)): Kn = 1: Sz = 1 / ((YN - Kn - 3)
^ .5)
2030 Z123 = Zr123 / Sz: Z132 = Zr132 / Sz: Z231 = Zr231 / Sz
2040 PRINT : PRINT "----------------------------------------"
2050 PRINT "TEST FOR THE SIGNIFICANCE OF r123 REPORT"
2060 PRINT "------------------------------------------"
2070 PRINT
2080 PRINT
2090 PRINT "HO: P123 = 0"
2100 PRINT "H1: P123 IS NOT EQUAL TO ZERO"
2110 PRINT
2120 PRINT "Sz =", Sz
2130 PRINT : LPRINT "Zr123 =", Zr123
2140 PRINT "Z123 =", Z123
2150 IF ABS(Z123) >= z(CRITICAL) THEN LPRINT "WE REJECT HO AND CONCLUDE THAT
RICE YEILD AND AVERAGE RAINFALL ARE SIGNIFICANTLY CORRELATED WHEN THE
TEMPERATURE IS HELD CONSTANT"
2160 IF ABS(Z123) < z(CRITICAL) THEN LPRINT "WE ACCEPT HO AND CONCLUDE THAT RICE
YIELD AND AVERAGE RAINFALL ARE NOT SIGNIFICANTLY CORRELATED WHEN THE TEMPERATURE
IS HELD CONSTANT"
2170 PRINT : PRINT "-----------------------------------------------------"
2180 PRINT "TEST FOR THE SIGNIFICANCE OF R132 REPORT"
2190 PRINT "----------------------------------------"
2200 PRINT
2210 PRINT
2220 PRINT "HO:P132 = 0"
2230 PRINT "H1: P132 IS NOT EQUAL TO ZERO"
2240 PRINT
2250 PRINT "Z132 =", Zr132
2260 PRINT "Z132 =", Z132
2270 IF ABS(Z132) >= z(CRITICAL) THEN PRINT "WE REJECT HO AND CONCLUDE THAT RICE
YIELD AND AVERAGE TEMPERATURE ARE SIGNIFICANTLY CORRELATE WHEN THE RAINFALL IS
HELD CONSTANT"
2280 IF ABS(Z132) < z(CRITICAL) THEN PRINT "WE ACCEPT HO AND CONCLUDE THAT RICE
YEILD AND AVERAGE TEMPERATURE ARE NOT SIGNIFICANTLY CORRELATED WHEN THE RAINFALL
IS HELD CONSTANT"
290 PRINT : PRINT "-----------------------------------------------------------"
300 PRINT "TEST FOR THE SIGNIFICANCE OF r231 REPORT"
310 PRINT "------------------------------------------------"
320 PRINT
330 PRINT
340 PRINT "HO: P231 = 0"
50 PRINT "H1: P231 IS NOT EQUAL TO ZERO"
60 PRINT
70 PRINT "Zr231 =", Zr231
```

```
2380 PRINT "Z231 =", Z231
2390 IF ABS(Z231) >= z(CRITICAL) THEN PRINT "WE REJECT HO AND CONCLUDE THAT
RAINFALL AND TEMPERATURE ARE SIGNIFICANTLY CORRELATION WHEN THE RICE YEILD IS
HELD CONSTANT"
2400 IF ABS(Z231) < z(CRITICAL) THEN PRINT "WE ACCEPT HO AND CONCLUDE THAT
RAINFALL AND TEMPERATURE ARE NOT SIGNIFICANTLY CORRELATED WHEN THE RICE YEILD IS
HELD CONSTANT"
2410 CLOSE #1, #2, #3, #4
2420 END
```

APPENDIX C

****

...........SAMPLE RESULT............

The mean of Y = 20.877
The mean of X1 = 1039.11
The mean of X2 = 376.5
B1 = 5.051658
b2 = 4. 148388
b0 = 9.089646

The Regression equation Y = 9.089646 + 5.051658X1 + 4.148388X2

TEST FOR SIGNIFICANCE OF THE PARAMETER ESTIMATES USING ANOVA

H0: B1 = 0 (There is no significant reletionship between Y, X1 and X2)
H1: Not all Bi's are equal to zero
The level of significance is .05
SST = 175.8139
SSR = 100.181
SSE = 75.63281
DF1 = 9
DF2 = 2
DF3 = 7
MSSR = 50.09052
MSSE = 10.80469
F1 = 4.635999
F(Critical) at 2 and 7 degrees of freedom = 4.74

THE ANOVA TABLE FOR THE DATA

Source of Variation:	Sum of square:	Degree of Freedom:	Mean square:	F*
Regression	100.181	2	50.09052	4.63599
Error	75.63281	7	10.80469	
Total	175.8139	9		

le accept H0 and conclude that Rice yield is not significantly related to the
verage Rainfall and average Temperature

EST FOR THE SIGNIFICANCE OF b0 REPORT

0: b0 = 0
1: B0 is not equal to zero
b0 = 50.36749
b0 = 7.097005
0 = 1.280772
e accept H0 and conclude that b0 is not statistically significant at the
level

EST FOR THE SIGNIFICANCE OF b1 REPORT

0: B1 = 0
1: B1 is not equal to zero
b1 = 7.785915

sb1 = 2.790325
t1 = 1.810419
We accept H0 and conclude that b1 is not statistically significant at the 5% level

TEST FOR THE SIGNIFICANCE OF b2 REPORT

H0: B2 = 0
H1: B2 is not equal to zero
vb2 = 2.90431
sb2 = 1.704204
t2 = 2.434209
We reject H0 and conclude that b2 is statistically significant at the 5% level

The coefficient of multiple determination DR = .5698158

THE CORRELATION COEFFICIENT REPORT

The correlation coefficient between Y and X1 denoted r12 = .4535083
The correlation coefficient between Y and X2 denoted r13 = .6069502
The correlation coefficient between X1 and X2 denoted r23 = .769079
The partial correlation coefficient between Y and X1 denoted r123 = .5647211
The partial correlation coefficient between Y and X2 denoted r132 = .677075
The partial correlation coefficient between X1 and X2 denoted r231 = .37764
The level of significance of the test is .05

TEST FOR THE SIGNIFICANCE OF r123 REPORT

H0: P123 = 0
H1: P123 is not equal to zero
Sz = .4082483
Zr123 = .639738
Z123 = 1. 567032
We accept H0 and conclude that Rice yield and average Rainfall are not significantly correlated when the temperature is held constant

----

TEST FOR THE SIGNIFICANCE OF r132 REPORT

H0: P132 = 0
H1: P132 is not equal to zero
Zr132 = .8236945
Z132 = 2.017631
We reject H0 and conclude that Rice yield and average Temperature are significantly correlated when Rainfall is held constant

----

TEST FOR THE SIGNIFICANCE OF r231 REPORT

H0: P231 = 0
H1: P231 is not equal to zero
Zr231 = -.3973048
Z231 = -.9731941
We accept H0 and conclude that Rainfall and Temperature are not significantly correlated when the Rice yeild is held constant