DP-6

MULTIPLE LINEAR REGRESSION

S RAMASESHAN
DIRECTOR

STUDY GROUP

S M SETH
N K GOEL

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAVAN
ROORKEE-247667 (UP)
INDIA

1983-84

CONTENTS

ABSTRACT

The association of three or more variables can be investigated by multiple linear regression and correlation analysis. The derivation of relationships among hydrologic variables is of importance for the transfer of information from few gauged stations to many ungauged stations. The general form of the multiple linear regression is:

$$X_1 = B_1 + B_2X_2 + B_3X_3 + \ldots\ldots\ldots\ldots + B_mX_m + \varepsilon$$

where, X_1 is dependent variable and $X_2, X_3 \ldots\ldots X_m$ are independent variables. ε is the error term.

In the documentation, listing of the source programme for multiple linear regression analysis, input data file and output file is given with test data and example calculations. In the programme the selection of different sets of independent variables ana designation of dependent variable can be made as many times as desired.

The programme calculates means, standard derivations of dependent and independent variables, correlation coefficients between dependent and independent variables, regression coefficients, standard error of regression coefficients, computed t-values, intercept, multiple correlation coefficient, standard error of estimate, analysis of variance for multiple regression and table of residuals.

1.0 INTRODUCTION

The derivation of relationships among hydrological variables is of great importance for the transfer of information from few gauged sites to many ungauged sites. Such relationships may be between one dependent variable and one independent variable. The relationship may also be among one dependent variable and more than one independent variables. The statistical technique of regression analysis is used for this purpose. The use of correlation and regression techniques in hydrologic analysis has increased manifold in recent years due to the advent of computers.

The relationship among three or more variables can be developed by multiple regression and correlation analysis. The general form of the multiple linear regression is given below:

$$X_1 = B_1 + B_2 X_2 + \dots\dots\dots + B_m X_m + \varepsilon \qquad \dots(1)$$

where,

X_1 : Dependent variable

$(X_2, X_3 \dots X_m)$: Independent variables

ε : Error term

If equation (1) is linear i.e. all the variables (dependent and independent) are in linear form, the regression is referred to as the multiple linear regression. The association between the variables is referred to as multiple correlation. The variables of non-linear relationships in hydrology are often transformed to linear relationships for the multiple regression analysis as it is easier to treat linear equations.

In designing the multiple linear relationships, the selection of dependent and independent variables is of great importance. The dependent variable is defined by the problem itself. The independent variables are generally selected by the following two criteria:

(a) The variables have been observed in the past concurrently with the dependent variable so that the regression equation may be established, and they will continue to be observed in the future also so that dependent variable may be predicted from them when necessary.

(b) The dependent variable should have dependence upon independent variables, from physical point of view.

The programme for multiple regression analysis described in this documentation has been taken from IBM's Scientific Subroutine Package and has been implemented/tested on VAX-11/780 computer system of National Institute of Hydrology, Roorkee.

2.0 PURPOSE OF THE PROGRAMME

The multiple linear regression analysis is performed for a set of independent variables and a dependent variable. In this programme selection of different sets of independent variables and designation of a dependent variable can be made as many times as desired. The programme carries out the following operations:

(a) Reads the title of the problem for multiple regression.

(b) Reads subset selections i.e. different sets of selections of independent and dependent variables for multiple regression analysis.

(c) Calls various subroutines to calculate means and standard deviations of dependent and independent variables, simple and multiple correlation coefficients, regression coefficients, t-values and analysis of variances for multiple regression.

(d) Prints the results.

3.0 SPECIFIC METHOD

The method is based on estimation of the regression coefficients by the least squares technique and involves computation of different statistical parameters such as mean, standard deviation, regression coefficients, correlation coefficients etc.

3.1 Parameters Estimation

The parameters or the regression coefficients are estimated by the method of least squares.

$$\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (X_1 - b_1 - b_2 X_2 \ldots\ldots\ldots b_m X_m)^2 \quad \ldots(2)$$

If $\sum_{i=1}^{N} \varepsilon_i^2$ is equal to Z, then m partial differential equations will be:

$$\frac{\partial Z}{\partial b_1} = 0; \quad \frac{\partial Z}{\partial b_2} = 0; \ldots\ldots\ldots; \frac{\partial Z}{\partial b_m} = 0$$

The above m partial differential equations will have following m linear equations:

$$b_2 \Sigma (\Delta X_2)^2 + b_3 \Sigma \Delta X_3 \Delta X_2 + \ldots\ldots + b_m \Sigma \Delta X_m \Delta X_2 = \Sigma \Delta X_1 \Delta X_2$$

$$b_2\Sigma\, \Delta X_2\, \Delta X_3 + b_3\Sigma(\Delta X_3)^2 + \ldots\ldots + b_m\Sigma\Delta X_m\, \Delta X_3 = \Sigma\Delta X_1\, \Delta X_3$$

$$\vdots$$

$$b_2\Sigma\Delta X_2\, \Delta X_m + b_3\Sigma\, (\Delta X_3\Delta X_m) + \ldots\ldots + b_m\Sigma(\Delta X_m)^2 = \Sigma\, \Delta X_1\, \Delta X_m \qquad \ldots(3)$$

in which $\Delta X_i = X_i - \bar{X}$ with i = 1 to m. The above equations enable the determination of m parameters $b_1, b_2, \ldots\ldots, b_m$ which are the estimates for $B_1, B_2, \ldots\ldots, B_m$.

The parameters can be estimated with the help of matrices also as given below:

$$[Y]=[X][B] \quad \text{for} \quad \sum_{i=1}^{N} \varepsilon_i^2 = 0 \qquad \ldots(4)$$

$$\text{or } [X]^T [Y] = [X]^T [X][B] \qquad \ldots(5)$$

$$\text{or } [X^T X]^{-1} [X]^T [Y] = [X^T X]^{-1} [X^T X] [B] \qquad \ldots(6)$$

$$\text{or } [B] = [X^T X]^{-1} [X]^T [Y] \qquad \ldots(7)$$

where,

[B] : Matrix containing M regression coefficients

[X] : (NxM) matrix

[Y] : (Nx1) matrix

3.2 Statistical Parameters

Various statistical parameters given in the output are computed by the following equations in the programme:

a) Mean

$$\bar{X}_j = \frac{\sum_{i=1}^{N} X_{ij}}{N} \qquad \ldots(8)$$

where,

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

m : Number of variables

N : Number of observations

b) Correlation coefficients

$$r_{jk} = \frac{S_{jk}}{(\sqrt{S_{jj}})\ (\sqrt{S_{kk}})} \qquad \ldots(9)$$

where,

$$S_{jk} = (1/N)\ \left(\sum_{i=1}^{N} (X_{ij} - T_j)(X_{ik} - T_k) - \sum_{i=1}^{N} (X_{ij} - T_j) \sum_{i=1}^{N} (X_{ik} - T_k) \right) \qquad \ldots(10)$$

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

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

$$T_j = \sum_{i=1}^{m} X_{ij} / m \qquad \ldots(11)$$

The temporary means T_j and T_k are used in the equation (10) to obtain computational accuracy.

c) Standard deviation

$$S_j = \sqrt{S_{jj} / N-1} \qquad \ldots(11)$$

where,

$j = 1, 2, \ldots\ldots, m.$

d) Regression coefficients

$$b_j = B_j \; (S_y / S_j) \qquad \ldots(12)$$

where,

b_j : Regression coefficient

B_j : Beta coefficient

S_y : Standard deviation of dependent variable

S_j : Standard deviation of j^{th} independent variable

e) Beta coefficients

$$B_j = \sum_{i=1}^{k} (r_{iy}) \, (r_{ij})^{-1} \qquad \ldots(13)$$

where,

r_{iy} : Intercorrelation of i^{th} independent variable with dependent variable

r_{ij}^{-1} : Inverse of intercorrelation r_{ij}

$i,j = 1,2,\ldots\ldots\ldots, k$ imply independent variables

f) Standard error of regression coefficients

$$S_{b_j} = \sqrt{\frac{r_{jj}^{-1}}{D_{jj}} \cdot S^2_{y\,.\,1,2\,\ldots\ldots\ldots k}} \qquad \ldots(14)$$

where,

D_{jj} : Sum of squares of deviations from mean for j^{th} independent variable

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

$S^2_{y.1,2.\ldots..k}$: Mean sum of squares due to regression

g) t-values

$$t_j = (b_j / S_{b_j}) \qquad \ldots(15)$$

where,

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

h) Intercept

$$b_0 = \bar{Y} - \sum_{i=1}^{k} b_j \ \bar{X}_j \qquad \ldots(16)$$

where,

$\bar{Y}$: Mean of dependent variable

$\bar{X}_j$: Mean of j^{th} independent variable

i) Multiple correlation coefficient

$$R = \sqrt{R^2_*} \qquad \ldots(17)$$

where,

R : Multiple correlation coefficient

R^2_*: Coefficient of determination

Coefficient of determination R^2_* is calculated by the following equation:

$$R^2_* = \sum_{i=1}^{k} B_i \ r_{ij} \qquad \ldots(18)$$

where,

B_i : Beta coefficient

r_{iy} : Intercorrelation of i^{th} independent variable with dependent variable

j) Analysis of variance for the multiple regression

The variance for the multiple regression analysis is tested by the F-value given by :

$$\text{F-value} = \frac{\text{Mean squares due to regression}}{\text{Mean squares from the regression}} \quad \ldots(19)$$

k) Standard error of estimate

The positive square root of Var (ε_i) is known as the standard error of estimate.

4.0 COMPUTER PROGRAMME

The programme consists of one main routine and five subroutine. The programme capacity can be changed by suitably changing the dimension statements.

4.1 Programme Subroutines

The multiple linear regression programme consists of the main routine named, MREG, a special input subroutine DATA and four other subroutines. The subroutines have been described below:

a) SUBROUTINE DATA (M,D)

The purpose of this subroutine is to read an observation from input device. This subroutine is called by subroutine CORRE and must be provided by the user. If size and location of data fields are different from problem to problem, this subroutine must be recompiled with a proper format statement. Various calling arguments are:

M : The number of variables in an observation

D : Output vector of length M containing the observation data

b) SUBROUTINE MINV (A, N, D, L, M)

The purpose of the subroutine is to invert a matrix. Various calling arguments are:

A : Input matrix, destroyed in computation and replaced by resultant inverse

N : Order of matrix A

D : Resultant determinant

L : Work vector of length N

M : Work vector of length N

c) SUBROUTINE CORRE (N,M,IO, X, XBAR, STD, RX, R,B,D,T)

This subroutine computes means, standard deviations, sums of cross products of deviations and correlation coefficients. Various calling arguments are:

N : Number of observations. N must be greater than or equal to 2.

M : Number of variables, M must be greater than or equal to 1.

IO : Option code for input data

0 if data are to be read in from input device in the special subroutine DATA

1 if data are already in core

X : If IO = 0 , the value of X is 0.0

if IO = 1, X is the input matrix (NxM) containing data

XBAR: Output vector of length M containing means

STD : Output vector of length M containing standard deviations

RX : Output matrix (MxM) containing sums of cross products of deviations from means

R : Output matrix containing correlation coefficients

B : Output vector of length M containing the diagonal of the matrix of sums of cross products of deviations from mean

D : Working vector of length M

T : Working vector of length M

d) SUBROUTINE ORDER (M,R,NDEP,K,ISAVE,RX,RY)

The purpose of this subroutine is to construct from larger matrix of correlation coefficients a subset matrix of intercorrelations among independent variables and a vector of intercorrelations of independent variables with dependent variable. Various calling arguments are:

M : Number of variables and order of matrix R

R : Input matrix containing correlation coefficients. This subroutine expects only upper triangular portion of the symmetric matrix to be stored by column in R

NDEP : The subscript number of dependent variable

K : Number of independent variables to be included in the forthcoming regression analysis. This must be greater than or equal to one

ISAVE: Input vector of length K+1, containing in ascending order the subscript numbers of K independent variables to be included in the forthcoming regression. Upon returning to the routine this vector contains, in addition, the subscript number of the dependent variable in K+1 position

RX : Output matrix (KxK) containing intercorrelations among independent variables to be used in forthcoming regression

RY : Output vector of length K containing intercorrelations of independent variables with dependent variable

e) SUBROUTINE MULTR (N,K,XBAR, STD, D, RX, RY, ISAVE, B, SB, T , ANS)

This subroutine performs multiple linear regression analysis for a dependent variable and a set of independent variables. Various calling arguments are:

N : Number of observations

K : Number of independent variables in the regression

XBAR : Input vector of length M containing means of all variables. M is the number of variables

STD : Input vector of length M containing standard deviations of all variables

D : Input vector of length M containing the diagonal of the matrix of sums of cross products of deviations from means of all variables

RX : Input matrix (K x K) containing the inverse of intercorrelations among independent variables

RY : Input vector of length K containing intercorrelations of independent variables with dependent variable

ISAVE : Input vector of length K+1 containing subscripts of independent variables in ascending order The subscript of the dependent variable is stored in the K+1 position

B : Output vector of length K containing regression coefficients

SB : Output vector of length K containing standard deviations of regression coefficients

T : Output vector of length K containing t-values

ANS : Output vector of length 10 containing following information:

ANS (1) : Intercept

ANS (2) : Multiple correlation coefficient

ANS (3) : Standard error of estimate

ANS (4) : Sum of squares attributable to regression (SSAR)

ANS (5) : Degrees of freedom associated with SSAR

ANS (6) : Mean squares of SSAR

ANS (7) : Sum of squares of deviations from regression (SSDR)

ANS (8) : Degrees of freedom associated with SSDR

ANS (9) : Mean squares of SSDR

ANS (10) : F value

The listing of the source programme has been given in appendix I.

4.2 Programme Modifications

Programme capacity can be increased or decreased by making changes in the dimension statements. Input data in a different format can also be handled by providing a specific format statement. The following are the general rules for programme modifications:

I. Changes in the dimension statements of the main programme MREG

a. The dimension of arrays XBAR, STD, D, RY, ISAVE, B, SB, T and W must be greater than or equal to the number of variables M.

b. The dimension of array RX must be greater than or equal to the product of MxM.

c. The dimension of array R must be greater than or equal to (M+1)M/2.

II. Changes in the input format statement of the special input subroutine DATA

The special input subroutine data is normally written by the user to handle different formats for different problems. The user may modify this subroutine to perform testing of input data, transforming of data and so on.

5.0 INPUT SPECIFICATIONS, OUTPUT DESCRIPTION AND RESTRICTIONS ON USE

5.1 Input Specifications

Input data file contains control cards, data cards and selection cards.

5.1.1 Control cards

In the first card the title of the problem is given in A format. In the second card the number of observations, number of variables and number of selections are given in free format.

5.1.2 Data cards

Since input data are read into the computer one observation at a time, each row of the data is given in one card in free format. If all the variables are not coming in one card, each row of data is continued on the second and third cards till the last data point comes in. However each row of data must be given in a new card. As indicated earlier the format for the data cards can be modified by the user.

5.1.3 Selection cards

The selection card is used to specify a dependent variable and a set of independent variables in multiple linear regression analysis.

Any variable in the set of original variables can be designated as a dependent variable and any number of remaining variables can be specified as independent variables. Selection of a dependent variable and a set of independent variables can be made as many times as desired. The selection card is prepared as follows:

Columns	Contents
1-2	Option code for table of residuals 00 if it is not required 01 if it is required
3-4	Dependent variable designated for the forthcoming regression analysis
5-6	Number of independent variables included in the forthcoming regression analysis
7-8	First independent variable included
9-10	Second independent variable included

Rest of the columns are for other independent variables. The input format (36I2) is used for the selection line.

5.2 Output Description

The output of the multiple linear regression analysis programme includes the followings:

a. Means

b. Standard deviations

c. Correlation coefficient between the independent variables and dependent variable

d. Regression coefficients

e. Standard error of regression coefficients

f. Computed t-values

g. Multiple correlation coefficient

h. Analysis of variance for the multiple regression

i. Standard error of estimate

6.0 TEST DATA

The programme for multiple regression analysis has been run on the data of sub zone 3 (f) (Lower Godavari Sub Zone) to relate unit hydrograph parameters to basin characteristics. The data has been taken from 'Flood Estimation Report for Lower Godavari Sub Zone (SUB ZONE -3f)' design office report No.3/1980. The data has been logarithmically transformed to have linear relationships.

7.0 EXAMPLE CALCULATIONS

The following relationships have been derived by the regression analysis:

(i) $t_p = 0.3468 \; (\frac{LL_c}{S})^{0.453}$

Coefficient of correlation is 0.85560

(ii) $q_p = 1.9527 \; (t_p)^{-0.837}$

Coefficient of correlation is 0.93826

(iii) $W_{50} = 2.333 \; (q_p)^{-1.012}$

Coefficient of correlation is 0.97041

(iv) $W_{75} = 1.3414 \; (q_p)^{-1.020}$

Coefficient of correlation is 0.95832

(v) $W_{R50} = 0.953 \; (q_p)^{-1.078}$

Coefficient of correlation is 0.95832

(vi) $W_{R75} = 0.5810 \; (q_p)^{-1.035}$

Coefficient of correlation is 0.93073

where,

t_p : Basin lag

q_p : Unit peak discharge

W_{50} : Width of the representative unit hydrograph at 50% of the peak of the hydrograph

W_{75} : Width at 75% peak flow

W_{R50}: Width of rising limb at 50% of peak flow

W_{R75}: Width of rising limb at 75% of peak flow

L : Length of mainstream in kilometers

L_c : Length of mainstream in kilometers from gauging site to the centre of gravity of catchment in kilometers

S : Statistical stream slope in meters per kilometer

8.0 APPLICATION, SAMPLE INPUT AND SAMPLE OUTPUT

The programme for multiple regression analysis has been run on the data of sub zone 3(f), (Lower Godavari Sub Zone) to relate unit hydrograph parameters to basin characteristics. The data has been log transformed to have linear relationships. Subroutine DATA can be modified accordingly.

8.1 Sample Input

The listing of sample input (data file) has been given in appendix II. Total number of observations, total number of variables and number of selections are 22, 13 and 6 respectively. So in the first line title of the problem has been given. In the second line total number of observations, total number of variables and total number of selections are given in free format.

In the data lines A (catchment area in square kilometers) L, L_c, W_c (minimum width of the catchment through the centre of gravity of the catchment in kilometers), S, t_p, q_p, W_{50}, W_{75}, W_{R50}, W_{R75}, (LL_c/S) and q_p are given in the free format. Different options have been given in selection lines.

8.2 Sample Output

The listing of the sample output has been given in appendix III.

9.0 RECOMMENDATIONS

The programme for multiple regression analysis can deal with upto 40 variables (including both dependent and independent variables). Therefore if the number of variables involved in a problem are less than 40, then the programme can be used as such. If there are more than 40 variables, the dimension statements in the main programme have to be modified. Input format in the subroutine DATA is also to be modified according to the problem.

REFERENCES

1. Haan, C.T.(1972),'Statistical Methods in Hydrology', Iowa State University Press, Ames, Iowa.

2. Scientific Subroutine Package, International Business Machines, White Plains, N.Y.

APPENDIX I

MULTIPLE LINEAR REGRESSION

```
C       MASTER MULTIPLE REGRESSION
        DIMENSION XBAR(40),STD(40),D(40),RY(40),ISAVE(40),B(40),
        1SB(40),T(40),W(40),AW(40)
        DIMENSION RX(1600),R(820),ANS(10),TITLE(80)
        OPEN(UNIT=5,FILE='MREG.DAT',STATUS='OLD')
        OPEN(UNIT=6,FILE='MREG.OUT',STATUS='NEW')
        OPEN(UNIT=13,FILE='R.DAT',STATUS='NEW')
  1     FORMAT(80A1)
  2     FORMAT(5X,'MULTIPLE REGRESSION.......'//,6X,
        1'SELECTION.......',I2//)
  3     FORMAT(1X,'VARIABLE',3X,'MEAN',3X,'STANDARD',3X,'CORRELATION',
        11X,'REGRESSION',1X,'STD. ERROR',2X,'COMPUTED'/6H   NO.,13X,
        2'DEVIATION',4X,6HX VS Y,4X,'COEFFIENT',1X,'OF REG.COFF.',
        31X,7HT VALUE)
  4     FORMAT(1H ,I4,6F11.3)
  5     FORMAT(10H DEPENDENT)
  6     FORMAT(1H0/10H INTERCEPT,10X,F16.5//' MULTIPLE CORRELATION  ',
        1F13.5//' STD. ERROR OF ESTIMATE',F13.5//)
  7     FORMAT(1H0,21X,' ANALYSIS OF VARIANCE FOR REGRESSION'//5X,
        1'SOURCE OF VARIATION',7X,'DEGREES',7X,'SUM OF',10X,'MEAN',
        26X,'F VALUE'/30X,'OF FREEDOM',4X,'SQUARES',9X,'SQUARES')
  8     FORMAT(' ATTRIBUTABLE TO REGRESSION  ',I6,2F16.3,F11.3/' DEVIA
        1TION FROM REGRESSION   ',I6,2F16.3)
  9     FORMAT(1H ,5X,5HTOTAL,19X,I6,F15.3)
 10     FORMAT(36I2)
 11     FORMAT(1H ,15X,'TABLE OF RESIDUALS'//' CASE NO.',5X,7HY VALUE,
        15X,10HY ESTIMATE,6X,8HRESIDUAL)
 12     FORMAT(1H ,I6,F15.5,2F14.5)
 13     FORMAT(' NUMBER OF SELECTION NOT SPECIFIED. JOB TERMINATED')
 14     FORMAT(' THIS MATRIX IS SINGULAR. THIS SELECTION IS SKIPPED')
 17     FORMAT(13F10.5)
C       READ  PROBLEM PARAMETER CARD
 100    READ(5,1)TITLE
        WRITE(6,1)TITLE
        READ(5,*)N,M,NS
        IO=0
        X=0.0
        CALL CORRE (N,M,IO,X,XBAR,STD,RX,R,D,B,T)
        REWIND 13
C       TEST NUMBER OF SELECTIONS
        IF(NS)108,108,109
 108    WRITE(6,13)
        GO TO 300
 109    DO 200 I=1,NS
        WRITE(6,2)I
        READ(5,10) NRESI,NDEP,K,(ISAVE(J),J=1,K)
C       NRESI.....OPTION CODE FOR TABLE RESIDUALS
C                 0 IF TABLE IS NOT REQUIRED
C                 1 IF TABLE IS NOT RQUIRED
C       NDEP..... DEPENDENT VARIABLE
```

```
C       K......... NO OF INDEPENDENT VARIABLES INCLUDED
C       ISAVE..... A VECTOR CONTAINING THE INDEPENDENT VARIABLES INCLUDED
        CALL ORDER (M,R,NDEP,K,ISAVE,RX,RY)
        CALL MINV(RX,K,DET,B,T)
C       TEST SINGULARITY OF MATRIX INVERTED
        IF(DET) 112,110,112
110     WRITE(6,14)
        GO TO 200
112     CALL MULTR(N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)
C       PRINT MEANS,STANDARD DEVIATIONS,INTERCORRELATIONS BETWEEN
C       X AND Y,REGRESSION COEFFICIENTS,STANDARD DEVIATIONS OF
C       REGRESSION COEFFICIENTS,AND COMPUTED T-VALUES
        MM=K+1
        WRITE(6,3)
        DO 115 J=1,K
        L=ISAVE(J)
115     WRITE(6,4) L,XBAR(L),STD(L),RY(J),B(J),SB(J),T(J)
        WRITE(6,5)
        L=ISAVE(MM)
        WRITE(6,4) L,XBAR(L),STD(L)
C       PRINT INTERCEPT,MULTIPLE CORRELATION COEFFICIENT,AND
C       STANDARD ERROR OF ESTIMATE
        WRITE(6,6)ANS(1),ANS(2),ANS(3)
C       PRINT ANALYSIS OF VARIANCE FOR REGRESSION
        WRITE(6,7)
        L=ANS(8)
        WRITE(6,8) K,ANS(4),ANS(6),ANS(10),L,ANS(7),ANS(9)
        L=N-1
        SUM=ANS(4)+ANS(7)
        WRITE(6,9) L,SUM
        IF(NRESI) 200,200,120
C       PRINT TABLE OF RESIDUALS
120     WRITE(6,2)I
        WRITE(6,11)
        MM=ISAVE(K+1)
        DO 140 II=1,N
        READ (13,17) (W(J),J=1,M)
        SUM=ANS(1)
        DO 130 J=1,K
        L=ISAVE(J)
130     SUM=SUM+W(L)*B(J)
        RESI=W(MM)-SUM
140     WRITE(6,12) II,W(MM),SUM,RESI
        REWIND 13
200     CONTINUE
300     CONTINUE
        STOP
        END
        SUBROUTINE DATA(M,D)
        DIMENSION D(1)
```

```
 1      FORMAT(13F10.3)
 C      THIS SUBROUTINE IS CALLED BY SUBROUTINE CORRE
        READ(5,*)(D(I),I=1,M)
        DO 10 I=1,M
        IF (D(I).EQ.0.)D(I)=1.
 10     D(I)=ALOG(D(I))
        WRITE(13,1)(D(I),I=1,M)
        RETURN
        END
        SUBROUTINE MINV(A,N,D,L,M)
        DIMENSION A(1),L(1),M(1)
 C      SEARCH FOR LARGEST ELEMENT
        D=1.0
        NK=-N
        DO 80 K=1,N
        NK=NK+N
        L(K)=K
        M(K)=K
        KK=NK+K
        BIGA=A(KK)
        DO 20 J=K,N
        IZ=N*(J-1)
        DO 20 I=K,N
        IJ=IZ+I
 10     IF(ABS(BIGA)-ABS(A(IJ))) 15,20,20
 15     BIGA=A(IJ)
        L(K)=I
        M(K)=J
 20     CONTINUE
 C      INTERCHANGE ROWS
        J=L(K)
        IF(J-K) 35,35,25
 25     KI=K-N
        DO 30 I=1,N
        KI=KI+N
        HOLD=-A(KI)
        JI=KI-K+J
        A(KI)=A(JI)
 30     A(JI)=HOLD
 C      INTERCHANGE COLUMNS
 35     I=M(K)
        IF(I-K)45,45,38
 38     JP=N*(I-1)
        DO 40 J=1,N
        JK=NK+J
        JI=JP+J
        HOLD=-A(JK)
        A(JK)=A(JI)
 40     A(JI)=HOLD
 C      DIVIDE COLUMNS BY MINUS PIVOT
```

```
45      IF(BIGA)48,46,48
46      D=0.0
        RETURN
48      DO 55 I=1,N
        IF(I-K)50,55,50
50      IK=NK+I
        A(IK)=A(IK)/(-BIGA)
55      CONTINUE
C       REDUCE MATRIX
        DO 65 I=1,N
        IK=NK+I
        HOLD=A(IK)
        IJ=I-N
        DO 65 J=1,N
        IJ=IJ+N
        IF(I-K)60,65,60
60      IF(J-K)62,65,62
62      KJ=IJ-I+K
        A(IJ)=HOLD*A(KJ)+A(IJ)
65      CONTINUE
C       DIVIDE ROW BY PIVOT
        KJ=K-N
        DO 75 J=1,N
        KJ=KJ+N
        IF(J-K)70,75,70
70      A(KJ)=A(KJ)/BIGA
75      CONTINUE
C       PRODUCT OF PIVOTS
        D=D*BIGA
C       REPLACE PIVOT BY RECIPROCAL
        A(KK)=1.0/BIGA
80      CONTINUE
C       FINAL ROW AND COLUMN INTERCHANGE
        K=N
100     K=(K-1)
        IF(K)150,150,105
105     I=L(K)
        IF(I-K)120,120,108
108     JQ=N*(K-1)
        JR=N*(I-1)
        DO 110 J=1,N
        JK=JQ+J
        HOLD=A(JK)
        JI=JR+J
        A(JK)=-A(JI)
110     A(JI)=HOLD
120     J=M(K)
        IF(J-K)100,100,125
125     KI=K-N
        DO 130 I=1,N
```

```
      KI=KI+N
      HOLD=A(KI)
      JI=KI-K+J
      A(KI)=-A(JI)
130   A(JI)=HOLD
      GO TO 100
150   RETURN
      END
      SUBROUTINE CORRE(N,M,IO,X,XBAR,STD,RX,R,B,D,T)
      DIMENSION X(1),XBAR(1),STD(1),RX(1),R(1),B(1),
      1D(1),T(1)
C     INITIALISATION
      DO 100 J=1,M
      B(J)=0.0
100   T(J)=0.0
      K=(M*M+M)/2
      DO 102 I=1,K
102   R(I)=0.0
      FN=N
      L=0
      IF(IO) 105,127,105
105   DO 108 J=1,M
      DO 107 I=1,N
      L=L+1
107   T(J)=T(J)+X(L)
      XBAR(J)=T(J)
108   T(J)=T(J)/FN
      DO 115 I=1,N
      JK=0
      L=I-N
      DO 110 J=1,M
      L=L+N
      D(J)=X(L)-T(J)
110   B(J)=B(J)+D(J)
      DO 115 J=1,M
      DO 115 K=1,J
      JK=JK+1
115   R(JK)=R(JK)+D(J)*D(K)
      GO TO 205
127   IF(N-M) 130,130,135
130   KK=N
      GO TO 137
135   KK=M
137   DO 140 I=1,KK
      CALL DATA(M,D)
      DO 140 J=1,M
      T(J)=T(J)+D(J)
      L=L+1
140   RX(L)=D(J)
      FKK=KK
```

```
      DO 150 J=1,M
      XBAR(J)=T(J)
  150 T(J)=T(J)/FKK
      L=0
      DO 180 I=1,KK
      JK=0
      DO 170 J=1,M
      L=L+1
  170 D(J)=RX(L)-T(J)
      DO 180 J=1,M
      B(J)=B(J)+D(J)
      DO 180 K=1,J
      JK=JK+1
  180 R(JK)=R(JK)+D(J)*D(K)
      IF(N-K) 205,205,185
  185 KK=N-KK
      DO 200 I=1,KK
      JK=0
      CALL DATA(M,D)
      DO 190 J=1,M
      XBAR(J)=XBAR(J)+D(J)
      D(J)=D(J)-T(J)
  190 B(J)=B(J)+D(J)
      DO 200 J=1,M
      DO 200 K=1,J
      JK=JK+1
  200 R(JK)=R(JK)+D(J)*D(K)
  205 JK=0
      DO 210 J=1,M
      XBAR(J)=XBAR(J)/FN
      DO 210 K=1,J
      JK=JK+1
  210 R(JK)=R(JK)-B(J)*B(K)/FN
      JK=0
      DO 220 J=1,M
      JK=JK+J
  220 STD(J)=SQRT(ABS(R(JK)))
      DO 230 J=1,M
      DO 230 K=J,M
      JK=J+(K*K-K)/2
      L=M*(J-1)+K
      RX(L)=R(JK)
      L=M*(K-1)+J
      RX(L)=R(JK)
      IF(STD(J)*STD(K))225,222,225
  222 R(JK)=0.0
      GO TO 230
  225 R(JK)=R(JK)/(STD(J)*STD(K))
  230 CONTINUE
      FN=SQRT(FN-1.0)
```

```
      DO 240 J=1,M
240   STD(J)=STD(J)/FN
      L=-M
      DO 250 I=1,M
      L=L+M+1
250   B(I)=RX(L)
      RETURN
      END
C
      SUBROUTINE ORDER(M,R,NDEP,K,ISAVE,RX,RY)
      DIMENSION R(1),ISAVE(1),RX(1),RY(1)
      MM=0
      NN=1
50    FORMAT(/30X,'MATRIX AND VECTOR SELECTED BY ORDER'/)
      DO 140 J=1,K
      L2=ISAVE(J)
      IF(NDEP-L2)122,123,123
122   L=NDEP+(L2*L2-L2)/2
      GO TO 125
123   L=L2+(NDEP*NDEP-NDEP)/2
125   RY(J)=R(L)
      DO 130 I=1,K
      L1=ISAVE(I)
      IF(L1-L2)127,128,128
127   L=L1+(L2*L2-L2)/2
      GO TO 129
128   L=L2+(L1*L1-L1)/2
129   MM=MM+1
130   RX(MM)=R(L)
120   FORMAT(10F12.6)
      NN=NN+K
140   CONTINUE
150   FORMAT(//10F12.6)
      ISAVE(K+1)=NDEP
      RETURN
      END
C
      SUBROUTINE MULTR(N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)
      DIMENSION XBAR(1),STD(1),RX(1),RY(1),D(1),ISAVE(1),
     1B(1),SB(1),T(1),ANS(1)
      MM=K+1
C     BETA WEIGHTS
      DO 100 J=1,K
100   B(J)=0.0
      DO 110 J=1,K
      L1=K*(J-1)
      DO 110 I=1,K
      L=L1+I
110   B(J)=B(J)+RY(I)*RX(L)
      RM=0.0
```

```
      BO=0.0
      L1=ISAVE(MM)
C     COEFFICIENTS OF DETERMINATION
      DO 120 I=1,K
      RM=RM+B(I)*RY(I)
C     REGRESSION COEFFICIENTS
      L=ISAVE(I)
      B(I)=B(I)*(STD(L1)/STD(L))
C     INTERCEPT
120   BO=BO+B(I)*XBAR(L)
      BO=XBAR(L1)-BO
C     SUM OF SQUARES ATTRIBUTABLE TO REGRESSION
      SSAR=RM*D(L1)
C     MULTIPLE CORRELATION COEFFICIENT
122   RM=SQRT(ABS(RM))
C     SUM OF SQUARES OF DEVIATIONS FROM REGRESSION
      SSDR=D(L1)-SSAR
C     VARIANCE OF ESSTIMATE
      FN=N-K-1
      SY=SSDR/FN
C     STANDARD DEVIATIONS OF REGRESSION COEFFICIENTS
      DO 130 J=1,K
      L1=K*(J-1)+J
      L=ISAVE(J)
125   SB(J)=SQRT(ABS((RX(L1)/D(L))*SY))
C     COMPUTED T-VALUES
130   T(J)=B(J)/SB(J)
C     STANDARD ERROR OF ESTIMATE
135   SY=SQRT(ABS(SY))
C     F VALUE
      FK=K
      SSARM=SSAR/FK
      SSDRM=SSDR/FN
      F=SSARM/SSDRM
      ANS(1)=BO
      ANS(2)=RM
      ANS(3)=SY
      ANS(4)=SSAR
      ANS(5)=FK
      ANS(6)=SSARM
      ANS(7)=SSDR
      ANS(8)=FN
      ANS(9)=SSDRM
      ANS(10)=F
      RETURN
      END
```

APPENDIX II

TEST INPUT

```
UNIT HYDROGRAPH PARAMETERS WITH CATCHMENT CHARACTERSTICS
22,13,6
   824.0 67.2 25.8 12.9  2.3  4.5  650.0  2.9  1.7  1.2  0.6 1143.0 0.79
   750.0 61.1 29.0 20.1  1.8  9.5  290.0  6.0  4.0  3.0  1.6 1321.0 0.39
   750.0 61.1 23.8 17.7  1.4  9.5  214.0  7.5  4.0  3.8  1.8 1229.0 0.29
   731.0 92.3 43.1 21.6  2.0 10.5  184.0  9.2  5.5  5.9  3.5  890.0 0.25
   483.0 41.8 17.7 18.5  3.8  4.5  280.0  3.3  2.0  1.3  1.0  380.0 0.58
   459.0 33.1  8.4 16.6  1.3  5.0  234.0  4.2  3.0  2.2  1.5  744.0 0.51
   364.0 35.2 12.9 16.4  1.8 11.5   60.0 15.3  7.0  7.2  3.4  339.0 0.17
   341.0 45.0 20.5 14.5  1.9  3.5  228.7  3.6  2.3  1.6  1.0  669.0 0.67
   242.0 27.7 11.2 14.1  3.8  3.5  140.8  4.4  3.1  1.2  1.0  159.0 0.58
   233.0 24.1 10.1 11.3  9.1  3.5  190.0  2.4  1.1  1.1  0.5   81.0 0.82
   208.0 25.0  6.8 10.5  2.1  2.5  179.0  2.8  1.7  1.1  0.7  117.0 0.86
   163.0 29.0 15.3  8.1  1.6  6.5   65.0  5.5  3.0  1.8  1.0  351.0 0.40
   139.0 23.0  8.5  7.3  2.3  4.5   80.0  3.9  2.3  1.5  1.0  129.0 0.58
   137.0 19.6  8.4  9.3  4.9  2.5  190.5  1.5  0.8  0.7  0.4   74.0 1.39
   120.0 18.2 10.0 10.3  1.9  3.5   65.1  4.8  3.4  1.8  1.3  132.0 0.54
    87.0 33.7 20.0  5.1  1.3  3.5   65.8  4.4  2.1  1.1  0.7  296.0 0.76
    65.0 18.0 10.0  3.1  3.4  2.5   66.5  2.1  1.3  1.1  0.8   98.0 1.07
    60.0 17.7  8.1  5.6  5.0  2.0   60.8  3.4  1.8  1.0  0.9   64.0 1.01
    54.0 12.2  5.3  6.2  7.5  2.5   41.0  2.9  1.6  1.4  0.7   81.0 0.77
    50.0 12.2  5.3  6.2  7.5  1.5   71.4  1.6  0.9  0.9  0.5   24.0 1.43
    42.0 14.7  7.7  3.5  5.4  1.5   43.5  2.3  1.3  0.8  0.4   49.0 1.04
    35.0 10.1  7.4  2.2  8.8  1.2   55.8  1.3  0.7  0.6  0.3   25.0 1.59
00060112
00130106
00080113
00090113
00100113
00110113
```

APPENDIX III

TEST OUTPUT

UNIT HYDROGRAPH PARAMETERS WITH CATCHMENT CHARACTERSTICS
MULTIPLE REGRESSION.......

SELECTION....... 1

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFIENT	STD. ERROR OF REG.COEF.	COMPUTED T VALUE
12	5.234	1.208	0.856	0.453	0.061	7.392
DEPENDENT						
6	1.314	0.639				

0

INTERCEPT -1.05669

MULTIPLE CORRELATION 0.85560

STD. ERROR OF ESTIMATE 0.33907

0 ANALYSIS OF VARIANCE FOR REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	1	6.282	6.282	54.643
DEVIATION FROM REGRESSION	20	2.299	0.115	
TOTAL	21	8.582		

MULTIPLE REGRESSION.......

SELECTION....... 2

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFIENT	STD. ERROR OF REG.COEF.	COMPUTED T VALUE
6	1.314	0.639	-0.938	-0.837	0.069	-12.129
DEPENDENT						
13	-0.431	0.570				

0

INTERCEPT 0.66922

MULTIPLE CORRELATION 0.93825

STD. ERROR OF ESTIMATE 0.20222

0 ANALYSIS OF VARIANCE FOR REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	1	6.016	6.016	147.115
DEVIATION FROM REGRESSION	20	0.818	0.041	

TOTAL	21	6.834	

MULTIPLE REGRESSION......

SELECTION...... 3

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFIENT	STD. ERROR OF REG.COEF.	COMPUTED T VALUE
13	-0.431	0.570	-0.970	-1.012	0.056	-17.977
DEPENDENT						
8	1.283	0.595				

0

INTERCEPT 0.84727

MULTIPLE CORRELATION 0.97041

STD. ERROR OF ESTIMATE 0.14720

0 ANALYSIS OF VARIANCE FOR REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	1	6.998	6.998	322.981
DEVIATION FROM REGRESSION	20	0.433	0.022	
TOTAL	21	7.431		

MULTIPLE REGRESSION......

SELECTION...... 4

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFIENT	STD. ERROR OF REG.COEF.	COMPUTED T VALUE
13	-0.431	0.570	-0.956	-1.020	0.070	-14.505
DEPENDENT						
9	0.733	0.609				

0

INTERCEPT 0.29377

MULTIPLE CORRELATION 0.95561

STD. ERROR OF ESTIMATE 0.18380

0 ANALYSIS OF VARIANCE FOR REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	1	7.108	7.108	210.401
DEVIATION FROM REGRESSION	20	0.676	0.034	

TOTAL	21	7.784

MULTIPLE REGRESSION......

SELECTION...... 5

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFIENT	STD. ERROR OF REG.COEF.	COMPUTED T VALUE
13	-0.431	0.570	-0.958	-1.078	0.072	-15.002
DEPENDENT						
10	0.417	0.642				

0

INTERCEPT -0.04757

MULTIPLE CORRELATION 0.95832

STD. ERROR OF ESTIMATE 0.18783

0 ANALYSIS OF VARIANCE FOR REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	1	7.940	7.940	225.056
DEVIATION FROM REGRESSION	20	0.706	0.035	
TOTAL	21	8.646		

MULTIPLE REGRESSION......

SELECTION...... 6

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFIENT	STD. ERROR OF REG.COEF.	COMPUTED T VALUE
13	-0.431	0.570	-0.931	-1.035	0.091	-11.381
DEPENDENT						
11	-0.097	0.634				

0

INTERCEPT -0.54284

MULTIPLE CORRELATION 0.93073

STD. ERROR OF ESTIMATE 0.23776

0 ANALYSIS OF VARIANCE FOR REGRESSION

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	1	7.323	7.323	129.536
DEVIATION FROM REGRESSION	20	1.131	0.057	
TOTAL	21	8. 454		