UM23

GENERALIZED COMPUTER PROGRAM

MULTIPLE LINEAR REGRESSION

SEPTEMBER 1970

HEC HYDROLOGIC ENGINEERING CENTER

Water Resources Support Center U.S. Army Corps of Engineers

MULTIPLE LINEAR REGRESSION

THE HYDROLOGIC ENGINEERING CENTER
GENERALIZED COMPUTER PROGRAM
704-G1-L2020

INTRODUCTION

1. ORIGIN OF PROGRAM

This program was developed in The Hydrologic Engineering Center by R. G. Willey and H. E. Kubik. A two-part FORTRAN II version was written in 1968; a September 1970 version was written in FORTRAN IV and included increased capability, and this version basically modifies the input to conform with present standards.

It is recognized that nearly all computer centers have access to multiple linear regression programs, but this program includes some different statistical philosophy and some capability that is particularly useful for regional analyses.

2. CAPABILITIES OF PROGRAM

The program makes a multiple linear regression analysis and has the following special features:

a. Automatic deletion. The analysis is first made with all of the specified independent variables and then the least significant variable is deleted and the analysis repeated.

b. Variable selection. All independent and dependent variables can be stored and then only those variables desired for a particular analysis may be used.

c. Combination of variables. New variables may be computed from the input variables, e.g., a parameter $A\sqrt{S}$ [basin area(A) times the square root of main channel slope (S)] can be computed when area and slope have been input.

d. Transformations. The variables may be transformed (square root, logarithmic or reciprocal) to more nearly linearize the relations.

e. Input regression parameters. The regression constant and regression coefficients may be input and the residuals (difference between observed and calculated), the adjusted multiple determination coefficient and the standard error of estimate will be computed.

3. HARDWARE AND SOFTWARE REQUIREMENTS

This program has been developed and tested primarily on the UNIVAC 1108 and the CDC 7600 computer. No tape units are required.

PROGRAM DESCRIPTION

4. PROGRAM ORGANIZATION

The program consists of a main program and two subroutines. Subroutine COMB is used to compute a new variable based on a combination of one or more variables. Subroutine CROUT solves a set of simultaneous equations to obtain the regression coefficients. The program will exit normally only in the main program. A macro-flow chart of the program organization is shown in figure 1.

5. THEORETICAL ASSUMPTION AND LIMITATIONS

The computations are made in accordance with the procedures given in "Statistical Methods in Hydrology" by Leo R. Beard, January 1962. The independent variables are deleted, in turn, based on the minimum adjusted partial determination coefficient ($\bar{r}^2$).

The adjusted multiple determination coefficient is used to compute the standard error of estimate. Care must be taken in insuring that there are more observations than variables. If the number of observations are very close to the number of variables, the results may be unreasonable.

Variables with zero or very little (.000001) variance are deleted from the analysis. A diagnostic is printed stating the name of the variable deleted because of this limitation.

6. METHODS OF COMPUTATION

The regression analysis is performed on either the transformed or non-transformed variables in a like manner. The basic equation is:

$$Y = a + \sum_{i=1}^{n} b_i X_i \qquad (1)$$

where the X_i values are the independent variables and the Y value is the dependent variable. "n" is the number of independent variables in the analysis. The regression coefficients (b_i) are calculated using the Crout Method (Exhibit 1) for solution of the following equations:

$$\begin{array}{l} \Sigma(x_1)^2\, b_1 + \Sigma(x_1x_2)\, b_2 + \ldots + \Sigma(x_1x_n)\, b_n = \Sigma(yx_1) \\ \Sigma(x_1x_2)\, b_1 + \Sigma(x_2)^2\, b_2 + \ldots + \Sigma(x_2x_n)\, b_n = \Sigma(yx_2) \\ \quad\vdots \qquad\qquad\quad \vdots \qquad\qquad\qquad\quad \vdots \qquad\qquad \vdots \\ \Sigma(x_1x_n)\, b_1 + \Sigma(x_2x_n)\, b_2 + \ldots + \Sigma(x_n)^2\, b_n = \Sigma(yx_n) \end{array} \tag{2}$$

where x_i is the deviation of the value X for variable i, the observation number being implied, from the mean ($\overline{X}_i$). $\Sigma(x_i)^2$ and $\Sigma(x_ix_j)$ can be determined from the following equations:

$$\Sigma(x_i)^2 = \Sigma(X_i - \overline{X}_i)^2 = \Sigma(X_i)^2 - (\Sigma X_i)^2/N \tag{3}$$

$$\Sigma(x_ix_j) = \Sigma(X_i - \overline{X}_i)(X_j - \overline{X}_j) = \Sigma(X_iX_j) - \Sigma X_i\ \Sigma X_j/N \tag{4}$$

where N is the number of events or observations.

The regression constant (a) is calculated by use of the equation:

$$a = \overline{Y} - \sum_{i=1}^{n} b_i\overline{X}_i \tag{5}$$

where $\overline{Y}$ is the mean of the dependent variable for N events and the $\overline{X}_i$ values are the means of each of the independent variables for N events.

The unadjusted (R^2) and the adjusted ($\overline{R}^2$) determination coefficients are determined by the following equations:

$$R^2 = \frac{b_1\ \Sigma(yx_1) + b_2\ \Sigma(yx_2) + \ldots + b_n\ \Sigma(yx_n)}{\Sigma(y)^2} \tag{6}$$

$$\bar{R}^2 = 1 - (1 - R^2)(N - 1)/(N - NVAR) \tag{7}$$

where NVAR is total number of variables (both dependent and independent) in the analysis.

The partial determination coefficient for a variable is the percent increase in unexplained variance caused by deleting that variable from the regression equation. The adjusted partial determination coefficient is calculated by the following equation:

$$\bar{r}^2 14.23 = \frac{(1-\bar{R}^2{}_{1.23}) - (1-\bar{R}^2{}_{1.234})}{1 - \bar{R}^2{}_{1.23}} \tag{8}$$

where the first subscript for the adjusted partial determination coefficient indicates the dependent variable, the second indicates the independent variable whose partial determination coefficient is being computed, and the subscripts after the decimal indicate the other independent variables involved in the computation.

The standard error of estimate (S_e) is calculated by the following equation:

$$S_e = \sqrt{(1-\bar{R}^2)\Sigma(y)^2/(N-1)} = S_y \sqrt{1-\bar{R}^2} \tag{9}$$

where S_y is the standard deviation of the dependent variable.

When the residual prediction option is specified the "RESIDUAL" column is calculated as the difference between the observed and predicted values. The "ERROR AS DEVIATE" column is calculated as the residual divided by the standard deviation. The "RATIO" column is calculated as the residual divided by the observed value.

Crout's Method*

One of the best methods for solving systems of linear equations on desk calculating machines was developed by P. D. Crout in 1941. This method is based on the elimination method, with the calculations arranged in systematic order so as to facilitate their accomplishment on a desk calculator. In this method the coefficients and constant terms of the equations are written in the form of a "matrix," which is a rectangular array of quantities arranged in rows and columns.

The method is best explained by an example. Suppose that in a multiple correlation analysis it is required to solve the following system of linear equations to obtain the unknown values of b_2, b_3, b_4 and b_5.

$$\begin{aligned}
\Sigma x_2^2\ b_2 + \Sigma x_2x_3\ b_3 + \Sigma x_2x_4\ b_4 + \Sigma x_2x_5\ b_5 &= \Sigma x_1x_2 \\
\Sigma x_2x_3\ b_2 + \Sigma x_3^2\ b_3 + \Sigma x_3x_4\ b_4 + \Sigma x_3x_5\ b_5 &= \Sigma x_1x_3 \\
\Sigma x_2x_4\ b_2 + \Sigma x_3x_4\ b_3 + \Sigma x_4^2\ b_4 + \Sigma x_4x_5\ b_5 &= \Sigma x_1x_4 \\
\Sigma x_2x_5\ b_2 + \Sigma x_3x_5\ b_3 + \Sigma x_4x_5\ b_4 + \Sigma x_5^2\ b_5 &= \Sigma x_1x_5
\end{aligned}$$

For simplicity let us replace the coefficients of the b's by the letters p, q, r and s, and the constant terms by the letter t, using subscripts 1, 2, 3 and 4 to denote the respective equations:

$$\begin{aligned}
p_1\ b_2 + q_1\ b_3 + r_1\ b_4 + s_1\ b_5 &= t_1 \\
p_2\ b_2 + q_2\ b_3 + r_2\ b_4 + s_2\ b_5 &= t_2 \\
p_3\ b_2 + q_3\ b_3 + r_3\ b_4 + s_3\ b_5 &= t_3 \\
p_4\ b_2 + q_4\ b_3 + r_4\ b_4 + s_4\ b_5 &= t_4
\end{aligned}$$

A continuous check on the computations as they progress may be obtained by adding to the matrix of the above system a column of u's, such that $u = p + q + r + s + t$. The matrix and check column are written as follows:

---

*The Crout Method was presented at the AIEE Summer Convention in June 1941 by Prescott D. Crout. The method was developed by Gauss and refined by Doolittle.

$$\begin{vmatrix} p_1 & q_1 & r_1 & s_1 & t_1 & u_1 \\ p_2 & q_2 & r_2 & s_2 & t_2 & u_2 \\ p_3 & q_3 & r_3 & s_3 & t_3 & u_3 \\ p_4 & q_4 & r_4 & s_4 & t_4 & u_4 \end{vmatrix}$$

The elements p_1, q_2, r_3 and s_4 form the "principal diagonal" of the matrix. Examination of the original equations shows that the coefficients are symmetrical about the principal diagonal, i.e., $q_1 = p_2$, $r_1 = p_3$, $r_2 = q_3$, $s_1 = p_4$, $s_2 = q_4$, and $s_3 = r_4$.
This is characteristic of the system of equations to be solved in any multiple correlation analysis. Because of this symmetry, the computations are considerably simplified. While the Crout method may be used to solve any system of linear equations, the computational steps given here are applicable only to those with symmetrical coefficients.

The solution consists of two parts, viz., the computation of a "derived matrix" and the "back solution." Let the derived matrix be denoted as follows:

$$\begin{vmatrix} P_1 & Q_1 & R_1 & S_1 & T_1 & U_1 \\ P_2 & Q_2 & R_2 & S_2 & T_2 & U_2 \\ P_3 & Q_3 & R_3 & S_3 & T_3 & U_3 \\ P_4 & Q_4 & R_4 & S_4 & T_4 & U_4 \end{vmatrix}$$

The elements of the derived matrix are computed as follows:

$$P_1 = p_1 \qquad P_2 = p_2 \qquad P_3 = p_3 \qquad P_4 = p_4$$

$$Q_1 = \frac{q_1}{p_1} \qquad R_1 = \frac{r_1}{p_1} \qquad S_1 = \frac{s_1}{p_1} \qquad T_1 = \frac{t_1}{p_1} \qquad U_1 = \frac{u_1}{p_1}$$

$$Q_2 = q_2 - P_2Q_1 \qquad Q_3 = q_3 - P_3Q_1 \qquad R_2 = \frac{Q_3}{Q_2}$$

$$Q_4 = q_4 - P_4Q_1 \qquad S_2 = \frac{Q_4}{Q_2} \qquad T_2 = \frac{t_2 - T_1P_2}{Q_2} \qquad U_2 = \frac{u_2 - U_1P_2}{Q_2}$$

$$R_3 = r_3 - Q_3R_2 - P_3R_1 \qquad R_4 = r_4 - Q_4R_2 - P_4R_1 \qquad S_3 = \frac{R_4}{R_3}$$

$$T_3 = \frac{t_3 - T_2Q_3 - T_1P_3}{R_3} \qquad U_3 = \frac{u_3 - U_2Q_3 - U_1P_3}{R_3}$$

$$S_4 = s_4 - R_4S_3 - Q_4S_2 - P_4S_1$$

$$T_4 = \frac{t_4 - T_3R_4 - T_2Q_4 - T_1P_4}{S_4} \qquad U_4 = \frac{u_4 - U_3R_4 - U_2Q_4 - U_1P_4}{S_4}$$

The general pattern of the above computations, which may be applied to a system containing any number of equations, is as follows:

(1) The first column of the derived matrix is copied from the first column of the given matrix.

(2) The remaining elements in the first row of the derived matrix are computed by dividing the corresponding elements in the first row of the given matrix by the first element in that row.

(3) After completing the $n^{\underline{th}}$ row, the remaining elements in the $(n+1)^{\underline{th}}$ column are computed. Such an element (X) equals the corresponding element of the given matrix minus the product of the element immediately to the left of (X) by the element immediately above the principal diagonal in the same column as (X), minus the product of the second element to the left of (X) by the second element above the principal diagonal in the same column as (X), etc. After each element below the principal diagonal is recorded, and while that element is still in the calculator, it is divided by the element of the principal diagonal which is in the same column. The quotient is the element whose location is symmetrical to (X) with respect to the principal diagonal.

(4) When the elements in the $(n+1)^{\underline{th}}$ column and their symmetrical counterparts have been recorded, the $(n+1)^{\underline{th}}$ row will be complete except for the last two elements, which are next computed. Such an element (X) equals the corresponding element of the given matrix minus the product of the element immediately above (X) by the element immediately to the left of the principal diagonal in the same row as (X), minus the product of the second element above (X) by the second element to the left of the principal diagonal in the same row as (X), etc., all divided by the element of the principal diagonal in the same row as (X).

The check column (U) of the derived matrix serves as a continuous check on the computations in that each element in the column equals one plus the sum of the elements in the same row to the right of the principal diagonal. That is,

$$U_1 = 1 + Q_1 + R_1 + S_1 + T_1$$

$$U_2 = 1 + R_2 + S_2 + T_2$$

$$U_3 = 1 + S_3 + T_3$$

$$U_4 = 1 + T_4$$

This check should be made after completing each row.

The elements of the derived matrix to the right of the principal diagonal form a system of equations which may now be used to compute the unknown values of b_2, b_3, b_4 and b_5 by successive substitution. This is known as the "back solution." The computations are as follows:

$$b_5 = T_4$$

$$b_4 = T_3 - S_3 b_5$$

$$b_3 = T_2 - S_2 b_5 - R_2 b_4$$

$$b_2 = T_1 - S_1 b_5 - R_1 b_4 - Q_1 b_3$$

It is very important that the computations be carried to a sufficient number of digits, both in computing the coefficients and constant terms of the original equations, and in computing the elements of the derived matrix. It is possible for relatively small errors in the coefficients and constant terms of the original equations to result in relatively large errors in the computed solutions of the unknowns. The greatest source of error in computing the elements of the derived matrix arises from the loss of leading significant digits by subtraction. This must be guarded against and can be done by carrying the computations to more figures than the data. As a general rule, it is recommended that the coefficients and constant terms of the original equations be carried to a sufficient number of decimals to produce at least five significant digits in the smallest quantity, and that the elements of the derived matrix be carried to one more decimal than this, but to not less than six significant digits.

```
T1       TEST NO. 1     JANUARY 1975
T2  STATISTICAL METHODS (BEARD 1962) EX 33
T3      FIRST ANALYSIS-DATA SUPPLIED
J1      4
NM          LOG Q LOG SNO  LOG GWLOG PRCP
TR       1  1  1  1  1                                              2
DT   1936     .939     .399     .325     .710
DT   1937     .945     .343     .385     .634
DT   1938    1.052     .369     .408     .886
DT   1939     .744     .246     .428     .581
DT   1940     .666     .181     .316    1.027
DT   1941    1.081     .297     .460    1.315
DT   1942    1.060     .299     .511    1.097
DT   1943     .892     .354     .379     .707
DT   1944    1.021     .295     .395    1.240
DT   1945     .920     .321     .376    1.091
DT   1946     .755     .168     .413    1.038
DT   1947     .960     .280     .410     .979
ED
NJ      1
T3      DATA FROM PREVIOUS ANALYSIS
TR      3 0 0 1 1
NJ
T1       TEST NO.2      JANUARY 1975
T2       DATA WITH  SPECIAL FORMAT
T3       COMPUTATION OF NEW VARIABLE
J1      4       1               1
NM           AREA  LENGTH    SLOPE    LOG Q
FT  (A2,A6,3F8.0/8X,F8.0)
CJ      9        1 L/S**.5
CO     -3        1       8       5       6      -2       1       7       4
CC     .5
TR      4 3 0 0 1 3
DT IDP R       230     38.8     1.27
DT IDP R     3.082
DT BUF C      19.4     10.9    15.42
DT BUF C     2.377
DT MCD C      7.52      7.0     9.66
DT MCD C     2.062
DTSALT 1      32.5     10.9    13.04
DTSALT 1     2.568
DTSALT 2       114     36.8     3.01
DTSALT 2     3.004
DT 2DP R       635     88.1     1.06
DT 2DP R     3.557
ED
ED
NJ
T1       TEST NO. 3     JANUARY 1975
T2       USE DATA FROM TEST NO. 1
T3       READ REGRESSION COEFFICIENTS
J1      4               1
NM          LOG Q LOG SNO  LOG GWLOG PRCP
TR      1  1  1  1  1                                              -1
RP  -.22      1.6      1.0       .3
DT   1936     .939     .399     .325     .710
DT   1937     .945     .343     .385     .634
DT   1938    1.052     .369     .408     .886
DT   1939     .744     .246     .428     .581
DT   1940     .666     .181     .316    1.027
DT   1941    1.081     .297     .460    1.315
DT   1942    1.060     .299     .511    1.097
DT   1943     .892     .354     .379     .707
DT   1944    1.021     .295     .395    1.240
DT   1945     .920     .321     .376    1.091
DT   1946     .755     .168     .413    1.038
DT   1947     .960     .280     .410     .979
ED
NJ
T1
```

TEST DATA OUTPUT

```
*********************************
   MULTIPLE LINEAR REGRESSION

   704-G1-L2020         JAN 1975
   REVISED              JAN 1983
*********************************
```

TEST NO. 1   JANUARY 1975
STATISTICAL METHODS (BEARD 1962) EX 33
FIRST ANALYSIS-DATA SUPPLIED

NVAL	NCOMB	IPRNT	IFRMT	IDELE	DELTA
4	0	0	0	0	0.0000

+ + + + ANALYSIS NO 1 + + + +

DEPENDENT VARIABLE -- LOG Q

IDEP	.......TRANSFORMATION CODES........	NOBR	IRES	IFORC
1	1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	2	0

VARIABLE	TRANSFORMATION
LOG Q	NONE
LOG SNO	NONE
LOG GW	NONE
LOG PRCP	NONE

INPUT DATA

OBS NO	OBS ID	LOG Q	LOG SNO	LOG GW	LOG PRCP
1	1936	0.939	0.399	0.325	0.710
2	1937	0.945	0.343	0.385	0.634
3	1938	1.052	0.369	0.408	0.886
4	1939	0.744	0.246	0.428	0.581
5	1940	0.666	0.181	0.316	1.027
6	1941	1.081	0.297	0.460	1.315
7	1942	1.060	0.299	0.511	1.097
8	1943	0.892	0.354	0.379	0.707
9	1944	1.021	0.295	0.395	1.240
10	1945	0.920	0.321	0.376	1.091
11	1946	0.755	0.168	0.413	1.038
12	1947	0.960	0.280	0.410	0.979

STATISTICS OF DATA

BASED ON 12 OBSERVATIONS

VARIABLE	AVERAGE	VARIANCE	STANDARD DEVIATION	
LOG SNO	0.2960	0.0050	0.0704	
LOG GW	0.4005	0.0028	0.0531	
LOG PRCP	0.9421	0.0572	0.2392	
LOG Q	0.9196	0.0181	0.1346	DEPENDENT VARIABLE

SIMPLE CORRELATION COEFFICIENTS

VARIABLE	LOG SNO	LOG GW	LOG PRCP	LOG Q
LOG SNO	1.0000	0.0000	-0.0459	0.6308
LOG GW	0.0000	1.0000	0.1275	0.4170
LOG PRCP	-0.0459	0.1275	1.0000	0.2011
LOG Q	0.6308	0.4170	0.2011	1.0000

INDEPENDENT VARIABLE	REGRESSION COEFFICIENT	PARTIAL DETERMINATION COEFFICIENT
LOG SNO	1.621805	0.9106
LOG GW	1.012931	0.6814
LOG PRCP	0.273387	0.7451

REGRESSION CONSTANT	R SQUARE	R BAR SQUARE	STANDARD ERROR OF ESTIMATE
-0.223704	0.9437	0.9226	0.0374

+++

VARIABLE DELETED IS LOG GW

INDEPENDENT VARIABLE	REGRESSION COEFFICIENT	PARTIAL DETERMINATION COEFFICIENT
LOG SNO	1.647213	0.7467
LOG PRCP	0.348730	0.5963

REGRESSION CONSTANT	R SQUARE	R BAR SQUARE	STANDARD ERROR OF ESTIMATE
0.103475	0.8011	0.7570	0.0663

DEPENDENT VARIABLE LOG Q

OBS NO	OBS ID	OBSERVED	CALCULATED	RESIDUAL	ERROR AS DEVIATE	RATIO
1	1936	0.939	1.008	-0.069	-0.515	-0.074
2	1937	0.945	0.890	0.055	0.412	0.059
3	1938	1.052	1.020	0.032	0.236	0.030
4	1939	0.744	0.711	0.033	0.243	0.044
5	1940	0.666	0.760	-0.094	-0.697	-0.141
6	1941	1.081	1.051	0.030	0.221	0.027
7	1942	1.060	0.979	0.081	0.605	0.077
8	1943	0.892	0.933	-0.041	-0.306	-0.046
9	1944	1.021	1.022	-0.001	-0.006	-0.001
10	1945	0.920	1.013	-0.093	-0.689	-0.101
11	1946	0.755	0.742	0.013	0.095	0.017
12	1947	0.960	0.906	0.054	0.400	0.056

R BAR SQUARED	0.7570
STANDARD ERROR OF ESTIMATE	0.0663
MEAN ERROR	0.0000
MEAN SQUARED ERROR	0.0033

++

VARIABLE DELETED IS LOG PRCP

INDEPENDENT VARIABLE	REGRESSION COEFFICIENT	PARTIAL DETERMINATION COEFFICIENT
LOG SNO	1.286212	0.3979

REGRESSION CONSTANT	R SQUARE	R BAR SQUARE	STANDARD ERROR OF ESTIMATE
0.538865	0.4526	0.3979	0.1044

DEPENDENT VARIABLE LOG Q

OBS NO	OBS ID	OBSERVED	CALCULATED	RESIDUAL	ERROR AS DEVIATE	RATIO
1	1936	0.939	1.052	-0.113	-0.840	-0.120
2	1937	0.945	0.980	-0.035	-0.260	-0.037
3	1938	1.052	1.013	0.039	0.286	0.037
4	1939	0.744	0.855	-0.111	-0.827	-0.150
5	1940	0.666	0.772	-0.106	-0.785	-0.159
6	1941	1.081	0.921	0.160	1.190	0.148
7	1942	1.060	0.923	0.137	1.015	0.129
8	1943	0.892	0.994	-0.102	-0.759	-0.115
9	1944	1.021	0.918	0.103	0.763	0.101
10	1945	0.920	0.952	-0.032	-0.236	-0.034
11	1946	0.755	0.755	0.000	0.000	0.000
12	1947	0.960	0.899	0.061	0.453	0.064

R BAR SQUARED	0.3979
STANDARD ERROR OF ESTIMATE	0.1044
MEAN ERROR	0.0000
MEAN SQUARED ERROR	0.0091

++

```
********************************
   MULTIPLE LINEAR REGRESSION

   704-G1-L2020       JAN 1975
   REVISED            JAN 1983
********************************
```

```
       TEST NO. 1    JANUARY 1975
  STATISTICAL METHODS (BEARD 1962) EX 33
      DATA FROM PREVIOUS ANALYSIS
```

```
                                          + + + + ANALYSIS NO  2 + + + +

                                          DEPENDENT VARIABLE --   LOG GW
```

IDEP	.......TRANSFORMATION CODES........	NOBR	IRES	IFORC
3	0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0	0

VARIABLE	TRANSFORMATION
LOG Q	NOT USED
LOG SNO	NOT USED
LOG GW	NONE
LOG PRCP	NONE

INPUT DATA

OBS NO	OBS ID	LOG GW	LOG PRCP
1	1936	0.325	0.710
2	1937	0.385	0.634
3	1938	0.408	0.886
4	1939	0.428	0.581
5	1940	0.316	1.027
6	1941	0.460	1.315
7	1942	0.511	1.097
8	1943	0.379	0.707
9	1944	0.395	1.240
10	1945	0.376	1.091
11	1946	0.413	1.038
12	1947	0.410	0.979

STATISTICS OF DATA

BASED ON 12 OBSERVATIONS

VARIABLE	AVERAGE	VARIANCE	STANDARD DEVIATION	
LOG PRCP	0.9421	0.0572	0.2392	
LOG GW	0.4005	0.0028	0.0531	DEPENDENT VARIABLE

SIMPLE CORRELATION COEFFICIENTS

VARIABLE	LOG PRCP	LOG GW
LOG PRCP	1.0000	0.1275
LOG GW	0.1275	1.0000

INDEPENDENT VARIABLE	REGRESSION COEFFICIENT	PARTIAL DETERMINATION COEFFICIENT
LOG PRCP	0.072132	0.0163

REGRESSION CONSTANT	R SQUARE	R BAR SQUARE	STANDARD ERROR OF ESTIMATE
0.332546	0.1057	0.0163	0.0526

++

```
********************************
  MULTIPLE LINEAR REGRESSION

  704-G1-L2020       JAN 1975
  REVISED            JAN 1983
********************************
```

TEST NO.2   JANUARY 1975
DATA WITH SPECIAL FORMAT
COMPUTATION OF NEW VARIABLE

NVAL	NCOMB	IPRNT	IFRMT	IDELE	DELTA
4	1	0	1	0	0.0000

DATA FORMAT (A2,A6,3F8.0/8X,F8.0)

COMBINATION 1	STEP	1	2	3	4	5	6	7	8	9
	OPERATION CODE	-3	1	8	5	6	-2	1	7	4
	CONSTANTS	1/	0.5000							

+ + + + ANALYSIS NO 1 + + + +

DEPENDENT VARIABLE --   LOG Q

IDEP	.......TRANSFORMATION CODES........	NOBR	IRES	IFORC
4	3 0 0 1 3 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0	0

VARIABLE	TRANSFORMATION
AREA	COMMON LOG
LENGTH	NOT USED
SLOPE	NOT USED
LOG Q	NONE
L/S**.5	COMMON LOG

INPUT DATA

OBS NO	OBS ID	AREA	LOG Q	L/S**.5
1	IDP R	230.000	3.082	34.429
2	BUF C	19.400	2.377	2.776
3	MCD C	7.520	2.062	2.252
4	SALT 1	32.500	2.568	3.018
5	SALT 2	114.000	3.004	21.211
6	2DP R	635.000	3.557	85.570

STATISTICS OF DATA

BASED ON 6 OBSERVATIONS

VARIABLE	AVERAGE	VARIANCE	STANDARD DEVIATION	
AREA	1.8162	0.5168	0.7189	
L/S**.5	1.0119	0.4526	0.6727	
LOG Q	2.7750	0.2936	0.5418	DEPENDENT VARIABLE

SIMPLE CORRELATION COEFFICIENTS

VARIABLE	AREA	L/S**.5	LOG Q
AREA	1.0000	0.9658	0.9936
L/S**.5	0.9658	1.0000	0.9566
LOG Q	0.9936	0.9566	1.0000

INDEPENDENT VARIABLE	REGRESSION COEFFICIENT	PARTIAL DETERMINATION COEFFICIENT
AREA	0.780995	0.8004
L/S**.5	-0.034300	0.0000

REGRESSION CONSTANT	R SQUARE	R BAR SQUARE	STANDARD ERROR OF ESTIMATE
1.391251	0.9898	0.9830	0.0705

++

VARIABLE DELETED IS L/S**.5

INDEPENDENT VARIABLE	REGRESSION COEFFICIENT	PARTIAL DETERMINATION COEFFICIENT
AREA	0.749774	0.9872

REGRESSION CONSTANT	R SQUARE	R BAR SQUARE	STANDARD ERROR OF ESTIMATE
1.413246	0.9897	0.9872	0.0614

++

```
********************************
  MULTIPLE LINEAR REGRESSION

  704-G1-L2020        JAN 1975
  REVISED             JAN 1983
********************************
```

TEST NO. 3 JANUARY 1975
USE DATA FROM TEST NO. 1
READ REGRESSION COEFFICIENTS

NVAL	NCOMB	IPRNT	IFRMT	IDELE	DELTA
4	0	1	0	0	0.0000

\+ + + + ANALYSIS NO 1 + + + +

DEPENDENT VARIABLE -- LOG Q

IDEP	.......TRANSFORMATION CODES........	NOBR	IRES	IFORC
1	1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	-1	0

VARIABLE	TRANSFORMATION
LOG Q	NONE
LOG SNO	NONE
LOG GW	NONE
LOG PRCP	NONE

STATISTICS OF DATA

BASED ON 12 OBSERVATIONS

VARIABLE	AVERAGE	VARIANCE	STANDARD DEVIATION	
LOG SNO	0.2960	0.0050	0.0704	
LOG GW	0.4005	0.0028	0.0531	
LOG PRCP	0.9421	0.0572	0.2392	
LOG Q	0.9196	0.0181	0.1346	DEPENDENT VARIABLE

SIMPLE CORRELATION COEFFICIENTS

VARIABLE	LOG SNO	LOG GW	LOG PRCP	LOG Q
LOG SNO	1.0000	0.0000	-0.0459	0.6308
LOG GW	0.0000	1.0000	0.1275	0.4170
LOG PRCP	-0.0459	0.1275	1.0000	0.2011
LOG Q	0.6308	0.4170	0.2011	1.0000

INPUT REGRESSION COEFFICIENTS

CONSTANT	-0.2200
LOG SNO	1.6000
LOG GW	1.0000
LOG PRCP	0.3000

DEPENDENT VARIABLE LOG Q

OBS NO	OBS ID	OBSERVED	CALCULATED	RESIDUAL	ERROR AS DEVIATE	RATIO
1	1936	0.939	0.956	-0.017	-0.129	-0.019
2	1937	0.945	0.904	0.041	0.305	0.043
3	1938	1.052	1.044	0.008	0.058	0.007
4	1939	0.744	0.776	-0.032	-0.237	-0.043
5	1940	0.666	0.694	-0.028	-0.206	-0.042
6	1941	1.081	1.110	-0.029	-0.213	-0.027
7	1942	1.060	1.098	-0.038	-0.286	-0.036
8	1943	0.892	0.937	-0.045	-0.338	-0.051
9	1944	1.021	1.019	0.002	0.015	0.002
10	1945	0.920	0.997	-0.077	-0.571	-0.084
11	1946	0.755	0.773	-0.018	-0.135	-0.024
12	1947	0.960	0.932	0.028	0.210	0.029

R BAR SQUARED	0.9190
STANDARD ERROR OF ESTIMATE	0.0383
MEAN ERROR	-0.0171
MEAN SQUARED ERROR	0.0013

++

DESCRIPTION OF INPUT DATA
GENERALIZED COMPUTER PROGRAM
704-G1-L2020

T1, T2, T3 Title Cards

Three title cards for the identification of the ouput. A T1 must be in columns 1 and 2 of the first card. A new T3 card is required for each subsequent analysis which uses the same data set.

J1 - First Job Card

Field[1]	Variable	Value	Description
1.	NVAL		Total number of values (variables) to be read for each observation and stored for use in this and subsequent analyses. This is not necessarily the number of variables to be included in the first analysis which will be specified by the TR card. Dimensioned for 18 variables.
2.	NCOMB		Number of new variables to be computed by combining two or more variables and constants. All combinations for subsequent analyses which use the same data must be performed in first analysis. Dimensioned for five combinations. Sum of NVAL and NCOMB must not exceed 18.
3.	IPRNT		Data list options.
		0	Provides list of input data for each job.
		1	Suppresses listing of input data.

[1] The standard card format for most HEC programs consists of columns 1 2 reserved for the card code (Field 0) and the remaining columns used for program specification or data. Columns 3-8 are Field 1 and the remaining 72 columns are divided into nine 8-column fields.

J1 Card (continued)

Field	Variable	Value	Description
4.	IFRMT		Format of data on the cards.
		0	Data in standard format of (A2,A6,9F8.0) for card identification, observation identification, and data, respectively. The format for the second card, if required, is (8X,9F8.0).
		1	Special format will be read on FT card.
5.	IDELE		Deletion options.
		0	Automatic deletion of each variable with lowest partial determination coefficient, in turn, until only one independent variable remains.
		1	Suppresses automatic deletion.
6-10.			Not used.

NM - Variable Name Card

Field	Variable	Value	Description
1.	DELTA		Increment which will be added to all data to meet certain constraints when logarithmic, square root, or reciprocal transformations are made. The increment should be zero unless the following constraints are violated:

Transform	Constraint
Logarithmic	"data" + DELTA >0
Square root	"data" + DELTA ≥0
Reciprocal	"data" + DELTA ≠0

Field	Variable	Value	Description
2-10.	ANAMA		Alphanumeric name of each variable read in, NVAL (J1.1) names. Order must correspond to order of data as read. If more than one card is needed, the name for the 10th variable is placed in field 2 of the second card, etc.

FT - Special Format Card

Supply this card only if IFRMT (J1.4) is positive.

Field	Variable	Value	Description
1-10.	IFMT		Format of data to be read from cards. The first field must read the data card code (the DT) in A2 format and the second field must provide for an alphanumeric identification of the observation in A format and NVAL (J1.1) floating point fields, e.g., (A2, A4, 10X, 10F6.0). The second A field cannot exceed six columns. NOTE: Parentheses must be provided in specification.

CJ - Combination Job Card[1]

Field	Variable	Value	Description
1.	NOP		Number of operation codes necessary to compute a new variable by combining one or more variables. Dimensioned for 20 operation codes.
2.	NCØN		Number of constants necessary to compute a new variable. Dimensioned for five constants.
3.	ANAMA		Alphanumeric name of new variable being computed.
4-10.			Not used.

CO - Combination Operations Card[1]

Supply this card only in conjunction with CJ card.

Field	Variable	Value	Description
1-10.	IOP		Operation to be performed, NOP (CJ.1) values.

[1] Provide NCOMB (J1.2) sets of CJ, CO, and, if required, CC cards.

CO Card (continued)

The sequence of operations necessary to obtain a desired computed variable must be specified by "operation codes." These operations are performed by using three operating registers, argument (ARG), accumulator (ACC) and storage (STORE). The accumulator has been initialized to zero prior to computations. Any constants necessary for the arithmetic will be entered on the CC card in the sequence necessary. The final value in the accumulator is used as the value for the new variable.

Operation	Operation Code	
Variable→ARG	-vn	Variable number, preceded by minus sign, to use next where "vn" is order number as read in or combined. This variable becomes the argument.
ARG+ACC→ACC	1	Add argument to accumulator.
ACC-ARG→ACC	2	Subtract argument from accumulator.
ARG(ACC)→ACC	3	Multiply argument times accumulator.
ACC/ARG→ACC	4	Divide accumulator by argument.
ACC^{ARG}→ACC	5	Raise accumulator to power of argument.
ACC→STORE then 0.→ACC	6	Store accumulator while further arithmetic is done. Accumulator is then reset to zero.
STORE→ARG	7	Retrieve stored value. This value now becomes the argument.
Constant→ARG	8	Use the next constant on the CC card as the argument.

CC - Combination Constants Card(1)

Supply this card only if **NCON** (CJ.2) is positive.

Field	Variable	Value	Description
1-5.	CNST		Supply NCON (CJ.2) values.
6-10.			Not used.

(1) Provide NCOMB (J1.2) sets of CJ, CO, and, if required, CC cards.

TR - Analysis Specification Card

This card is provided for each analysis.

Field	Variable	Value	Description
1.	IDEP		Number identifying the order number of the dependent variable in the total array of values read in, i.e., 3 if third value in the array is the desired dependent variable.
2-6.	IFUNC		Variable deletion or transformation code for each variable. NVAL (J1.1) plus NCOMB (J1.2) values in special (I2) format, i.e., the code for the first variable read would be in column 10, the code for the second variable in column 12, etc.
			Operation
		0	Delete from current analysis
		1	No transformation
		2	Square root
		3	Logarithmic, base 10
		4	Reciprocal
7.	NOBR		The number of observations to be used in this analysis. The program will count the number of observations (NOBS) as the data are read; therefore, may be left blank if all observations are to be used in analysis. If positive, the first NOBR observations will be used, but must be less than or equal to NOBS. Dimensioned for 500 observations.
8.	IRES		Residual-prediction options. Provides a tabulation of the observed, calculated, the difference between the observed and calculated, and, if the log transformation of the dependent variable, the ratio of the observed to the calculated. The computation for a multiple linear regression analysis will be performed when $IRES \geq 0$. Predetermined regression parameters will be read (RP card) when IRES= -1.

TR Card (continued)

Field	Variable	Value	Description
		-1	Regression parameters to be read in with no analyses.
		0	Suppresses residual--prediction routine.
		>0	Positive number indicating maximum number of independent variables to be included in equation. Routine will operate for each deletion thereafter.
9-10.			Not used.

RP - Regression Parameters Card

Supply this card only if IRES (TR.8) is -1.

Field	Variable	Value	Description
1.	AA		Regression constant.
2.	BB		Regression coefficients. Supply same number of coefficients as there are independent variables (one less than the number of IFUNC (TR.2) values that are greater than zero) in the order corresponding to data.

DT - Input Data Cards

Input data need only be provided for first analysis. Subsequent analyses can use same data provided all necessary data are read in first analysis. The data may be in the following format or the format specified by the FT card. Dimensioned for 500 observations.

Field	Variable	Value	Description
1.	XID		An alphanumeric identification for the observation.
2.	X		Value for the first variable.

DT Card (continued)

Field	Variable	Value	Description
3-10.	X		Value for succeeding variables, NVAL (J1.1) values. If more than one card is needed, the value for the 10th variable is placed in field 2 of the second card, etc. The identification (XID) can be repeated in field 1 to assist in identifying each card.

ED - End of Data Card

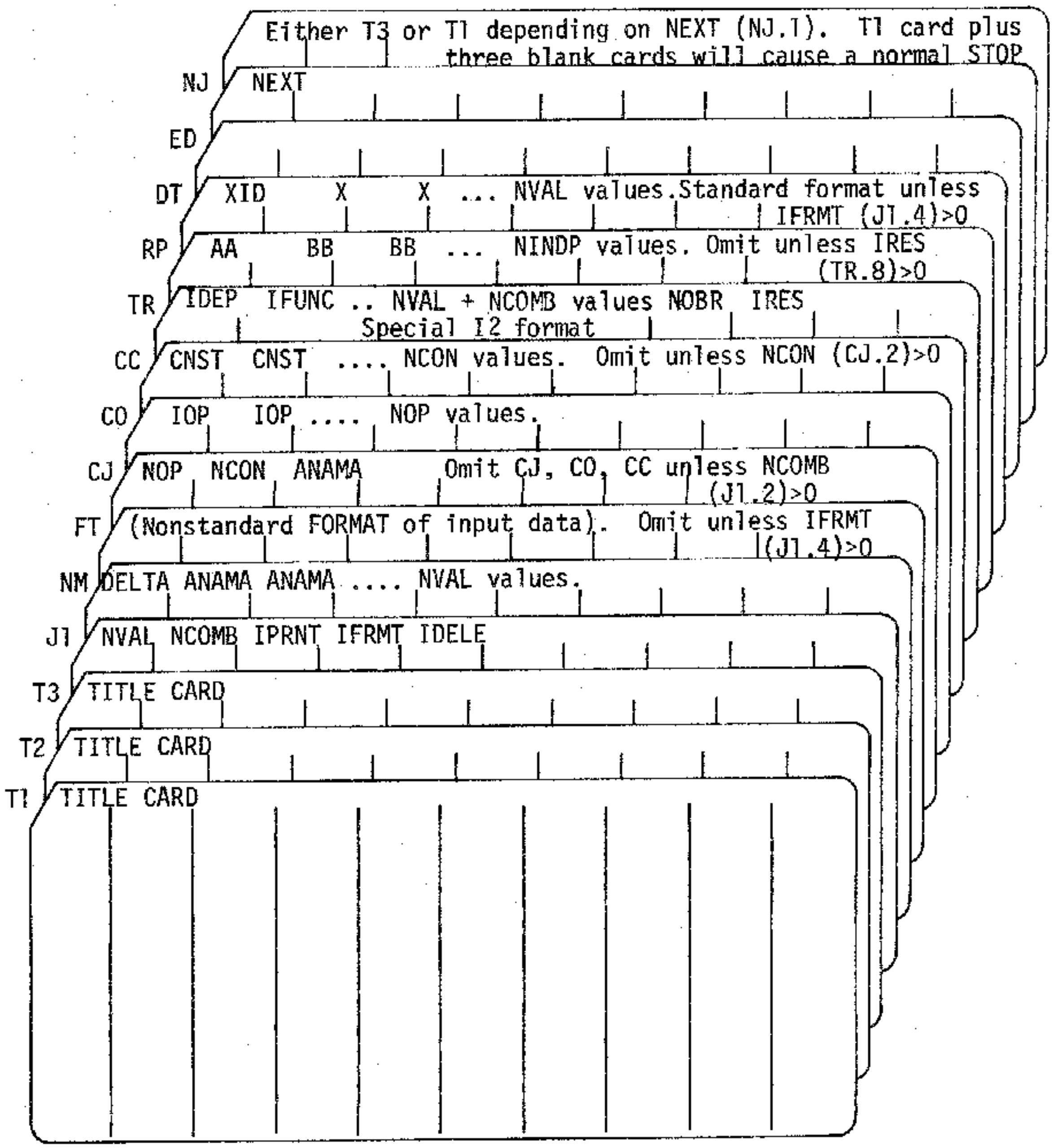
Supply the same number of ED cards as required to supply one observation on DT cards.

Field	Variable	Value	Description
0.	ID		The letters ED <u>must</u> appear in columns 1 and 2.
1-10.			Not used.

NJ - Next Job Card

Field	Variable	Value	Description
1.	NEXT		This variable indicates what type of job follows according to the following values:
		0	Begin a new job. Branches to beginning of program and reads T1, T2, and T3 cards for a new job or for a normal stop.
		1	Different analysis with same input data. Branches to read T3, TR, and, if required, RP cards. (No data cards provided.)

T1 card, plus three blank cards, will cause a normal STOP.


Either T3 or T1 depending on NEXT (NJ.1). T1 card plus three blank cards will cause a normal STOP
NJ
NEXT
ED
DT
XID X X ... NVAL values. Standard format unless IFRMT (J1.4)>0
RP
AA BB BB ... NINDP values. Omit unless IRES (TR.8)>0
TR
IDEP IFUNC .. NVAL + NCOMB values NOBR IRES
Special I2 format
CC
CNST CNST .... NCON values. Omit unless NCON (CJ.2)>0
CO
IOP IOP .... NOP values.
CJ
NOP NCON ANAMA Omit CJ, CO, CC unless NCOMB (J1.2)>0
FT
(Nonstandard FORMAT of input data). Omit unless IFRMT (J1.4)>0
NM
DELTA ANAMA ANAMA .... NVAL values.
J1
NVAL NCOMB IPRNT IFRMT IDELE
T3
TITLE CARD
T2
TITLE CARD
T1
TITLE CARD