NOAA Eastern Region Computer Programs
and Problems NWS ERCP - No. 31

DEPARTMENT OF COMMERCE
UNITED STATES OF AMERICA

Correlation and Regression Equation Program - REGRS

Hugh M. Stone
National Weather Service Eastern Region
Garden City, New York

Scientific Services Division
Eastern Region Headquarters
May 1985

U.S. DEPARTMENT OF COMMERCE / National Oceanic and Atmospheric Administration / National Weather Service

NOAA TECHNICAL MEMORANDUM

National Weather Service, Eastern Region Computer Programs and Problems

The Eastern Region Computer Programs and Problems (ERCP) series is a subset of the Eastern Region Technical Memorandum series. It will serve as the vehicle for the transfer of information about fully documented AFOS application programs. The format ERCP - No. 1 will serve as the model for future issuances in this series.

1 An AFOS version of the Flash Flood Checklist. Cynthia M. Scott, March 1981. (PB81 211252).

2 An AFOS Applications Program to Compute Three-Hourly Stream Stages. Alan P. Blackburn, September 1981. (PB82 156886).

3 PUPPY (AFOS Hydrologic Data Reporting Program). Daniel P. Provost, December 1981. (PB82 199720).

4 Special Search Computer Program. Alan P. Blackburn, April 1982. (PB83 175455).

5 Conversion of ALEMBIC$ Workbins. Alan P. Blackburn, October 1982. (PB83 138313).

6 Real-Time Quality Control of SAOs. John A. Billet, January 1983. (PB83 166082).

7 Automated Hourly Weather Collective from HRR Data Input. Lawrence Cedrone, January 1983 (PB83 167122).

8 Decoders for FRH, FTJ and FD Products. Cynthia M. Scott, February 1983. (PB83 176057).

9 Stability Analysis Program. Hugh M. Stone, March 1983. (PB83 197947).

10 Help for AFOS Message Comp. Alan P. Blackburn, May 1983. (PB83 213561).

11 Stability and Other Parameters from the First Transmission RAOB Data. Charles D. Little, May 1983. (PB83 220475).

12 TERR, PERR, and BIGC: Three Programs to Compute Verification Statistics. Matthew R. Peroutka, August 1983. (PB84 127521).

13 Decoder for Manually Digitized Radar Observations. Matthew R. Peroutka, June 1983. (PB84 127539).

14 Slick and Quick Data Entry for AFOS Era Verification (AEV) Program. Alan P. Blackburn, December 1983. (PB84 138726).

15 MDR--Processing Manually Digitized Radar Observations. Matthew R. Peroutka, November 1983. (PB84 161462)

16 RAWP: Stability Analysis Program. Hugh M. Stone, February 1984.(PB84 1614

17 ZONES. Gerald G. Rigdon, March 1984. (PB84 174325)

18 Automated Analysis of Upper Air Soundings to Specify Precipitation Type. Joseph R. Bocchieri and Gerald G. Rigdon, March 1984. (PB84 174333)

(Continued on Inside Rear Cover)

EASTERN REGION COMPUTER PROGRAMS AND PROBLEMS - No. 31

Correlation and Regression Equation Program - REGRS

Hugh M. Stone
National Weather Service Eastern Region
Garden City, New York

Scientific Services Division
Eastern Region Headquarters
May 1985

Correlation and Regression Equation Program - REGRS

Hugh M. Stone
Scientific Services Division
National Weather Service Eastern Region
Garden City, New York

I. Introduction

In local research studies, there is frequently a need to compute correlation coefficients between some variable, which is to be predicted, and various other variables which may serve as predictors. After an appropriate database has been collected, this program computes linear correlation coefficients between the dependent variable to be predicted and all the possible predictor variables, and also computes a multiple correlation coefficient using all the predictor variables together and partial correlations for each of the predictors. In addition, a linear regression equation is obtained relating the predictors to the dependent variable.

The program also has the capability of going through a screening process, whereby one variable at a time is dropped from the regression equation, eliminating those with lowest partial correlation coefficients first, until only one predictor variable remains. A new multiple correlation coefficient and new regression equation is computed as variables are dropped. A choice can then be made concerning which variables to keep in the final regression equation. This is the so-called backwards elimination process (Draper and Smith, 1966). The screening process may be disregarded and a regression equation may be obtained using any specified subset of the original set of predictors.

Point biserial correlation coefficients (Walker and Lev, 1953) may also be obtained from this program. The point biserial correlation coefficient measures the relationship between a continuous variable, e.g. energy index, and a binary variable, e.g. occurrence or non-occurrence of severe weather. To compute this type of coefficient, the dependent variable is assigned a value of zero (Ø) for a non-occurrence of an event and a value of one (1) for an occurrence. The same equations that are used for computation of correlation coefficients for a continuous dependent variable are then also used for the binary valued dependent variable.

Although these are standard statistical procedures, a generalized program that accomplishes all the above tasks has not previously been widely available to the field offices.

II. Methodology and Software Structure

A. Methodology

The program operates on data stored in an RDOS file called IDXCR1. The format of data storage in this file is arbitrary and determined by the user; however, the first line <u>must</u> consist of exactly 80 characters, which are read in the format (40A2). This first line is for identification purposes and may contain whatever the user wishes or may be left blank (80 spaces). The file may consist of data that has been automatically extracted from the AFOS database and written to IDXCR1, or it may be data that has been collected manually and typed into the file, or it may be a mixture of automatically extracted and manually input data.

The format for reading the IDXCR1 file is specified at runtime. An example of a small portion of an IDXCR1 file is shown in figure 1, and consists of six records containing nine variables each. The width of each field for each variable on the first line of the record is six characters and a field width of 9 characters on the second line of the record. The appropriate format in this case would be (5F6.0, F6.2, F6.0, F6.4/F9.0), but since all the variables contain a decimal point, an alternate simpler format for reading would be (8F6.0/F9.0). The parentheses must be included when specifying a format.

Since this program is for research purposes only and will not be run very often, it is designed to operate from the Dasher. When the program is started a series of questions appear on the Dasher, which must be answered:

1. No. of variables per record in IDXCR1?

2. Format for reading data in IDXCR1?
   Up to 80 characters may be typed in, including parentheses.

3. Which of the variables are independent (predictor) variables and which one is the dependent (predictand) variable?
   A series of one digit integers corresponding to the position of each variable in the record must be typed in, with "2" denoting the dependent variable, "1" for each of the predictor variables and "0" denoting not to use that variable in the computation. For example, if the first variable in the record of figure 1 is considered the dependent variable, and variables number 3, 4, 5, 7, and 9 are considered the independent variables, the appropriate entry would be 201110101. Variables 2, 6 and 8 will not be used in the computation. The dependent variable does <u>not</u> have to be first, e.g. an entry of the form 102110101 would be acceptable.

4. Type dependent variable name or ID.
   This provides additional identification on the output or may be left blank. Up to 80 characters may be typed in.

5. Simple correlation only?
   Response is "Y" yes or "N" no.
   Simple correlation means that all correlation coefficients are computed along with the full regression equation, but <u>no</u> screening is done to determine regression equations with fewer variables.

An example of the Dasher printout with questions and answers is given in Figure 2.

B. Software Structure

The file IDXCR1 is read the first time and the matrix needed to obtain the normal equations for the regression coefficients is derived then the system of equations is solved. Mean values of all the variables are computed.

IDXCR1 is read for the second time and the regression equation derived previously is used to estimate the predictand variable for each observation in the file. The multiple correlation coefficient is given by the formula,

$$RM = \left(\frac{\Sigma(Y'-\overline{Y})^2}{\Sigma(Y-\overline{Y})^2}\right)^{\frac{1}{2}} \tag{1}$$

where Y is the value of the dependent variable
Y' is estimate of Y from the regression equation
$\overline{Y}$ is mean value of Y,
and sums are taken over the entire sample in file IDXCR1.

The standard deviation of Y is computed from the formula,

$$SY = \left(\frac{\Sigma(Y-\overline{Y})^2}{M}\right)^{\frac{1}{2}} \tag{2}$$

where M is the number of observations.

The simple correlation coefficients for each predictor variable are computed from the product-moment formula,

$$R = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\left[\Sigma(X-\overline{X})^2 \cdot \Sigma(Y-\overline{Y})^2\right]^{\frac{1}{2}}} \tag{3}$$

where X is the value of one of the predictor variables and $\overline{X}$ is the mean value of X. All of the above formulas may be found in Spiegel (1961).

If the backward elimination process is used to determine regression equations with fewer predictors, one variable is removed from the system and a new regression equation is obtained, then that variable is restored and a second is removed, with a new regression equation computed. The process continues until a regression equation is obtained for each of the subsystems with one variable eliminated. A new matrix does not have to be computed for each of the subsystems. Solutions are obtained by eliminating one row and one column systematically from the original matrix.

As an example of this process, consider equation (7) in the Appendix, which is the linear system for four predictor variables X_1, X_2, X_3, and X_4. The solution of the entire system yields the coefficients C_0, C_1, ...C_4 for the complete regression equation. Now to obtain the solution for the system with variable X_1 eliminated, all we need to do is remove the second row and the second column from the original system and solve the reduced system. Likewise, if variable X_2 is eliminated, remove the third row and third column and solve again. The process continues until each one has been removed in turn.

Now that we have regression equations for each of the partial sets of variables, we may compute partial correlation coefficients for each variable. The partial correlation coefficient measures the correlation between a dependent variable and one particular independent variable when all other variables involved are kept constant, i.e. when the effects of all other variables are removed. The multiple correlation coefficient for each of the partial sets of variables must be computed prior to the partial correlation computation and this will be denoted RMP_i; the "i" subscript denotes that it is the multiple correlation coefficient with the ith independent variable removed. Recall that RM denotes the multiple correlation coefficient with all independent variables accounted for. The equation for the partial correlation coefficient may now be written,

$$RP_i = \left[1 - \frac{1-(RM)^2}{1-(RMP_i)^2}\right]^{\frac{1}{2}} \qquad (4)$$

This equation is given with different notation in Ezekiel (1941).

The variable having the lowest partial correlation coefficient is now permanently eliminated from the system. In the example mentioned previously, equation (7) in the Appendix, one row and one column have now been permanently removed from the system. At this point, the entire process of elimination is repeated, but this time only three independent variables appear in equation (7), and one at a time these are removed until the system is permanently reduced to two variables. Eventually, only one variable remains, and the computation is finished. All of the correlation coefficients and regression equation coefficients for the original system of variables and each of the reduced subsystems are written to file OUTR.

A sample of the output in file OUTR, when the screening process is eliminated, is shown in Figure 3. In this example, the dependent variable is number 2 in a larger IDXCR1 file and the computation uses independent (predictor) variables 16, 20, 21, and 29; all other variables are ignored. Simple correlation coefficients from the product-moment formula (3) are given first, followed by the multiple correlation coefficient using all four variables and finally the regression equation coefficients are given with the constant term printed on first line followed by the coefficient for each of the four predictor variables.

If the screening process is used, additional information is written to OUTR. A sample of OUTR using the same data is shown in Figure 4. In this case, multiple correlation coefficients (MCC) are given using the original set of four variables (N=4) and each of the subsets of reduced variables (N=3,2,1). The fourth column (COEF) gives the regression equation coefficients using the variables indicated in column 1 (VRBL). For N=4, these are, of course, the same as those shown in Figure 3. Column 2 (MPC) gives the multiple correlation coefficient for the subset when the variable indicated in column 1 is eliminated from the computation. The third column PCC is the partial correlation coefficient for each of the variables in column 1, and is computed from MPC (RMP) and MCC (RM) using equation 4.

The variable having the lowest partial correlation (variable 16) is permanently eliminated; three variables 20, 21, and 29 remain. The computation is repeated for the three remaining variables and this time variable 21 has the lowest PCC and is eliminated. Finally one variable, number 29, remains, which also has the highest simple correlation with the dependent variable 2.

It can be seen that MCC increases from .35803 to .41050 when variable 20 is combined with variable 29 and represents a significant improvement. When variable 21 is included MCC increases only slightly to .41574 and when variable 16 is included the increase is even less. Therefore, the regression equation involving only variables 20 and 29 would give almost as good a prediction as including the four variables. The correlation (MCC=.41050) is not very high, and additional predictor variables other than 16 and 21 would need to be considered to develop a more useful prediction equation.

The software structure with associated subroutines and loadline is shown in figure 5. The main program REGRS calls several subroutines which will be briefly described:

SYSL

This subroutine was extracted from the codeof Subroutine LSTSQ, which is in the CRH.LB (Schwein, 1983). It solves a system of up to 40 linear equations in 40 variables.

REDUC

Reduces a matrix by eliminating one row and one column, then compacts the remaining elements.

RPMIN

Tests a set of partial correlation coefficients RP to find the minimum value, which will be eliminated, and adjusts indices of the remaining variables for printout.

TDIF

Computes elapsed time of the computation, from beginning and ending time.

III. Cautions and Restrictions

The main problem in using the backward elimination process for screening the independent variables is that care must be taken to assure that the variables are truly independent. If a variable is chosen that is too closely related to one or more of the other independent variables, there is a slight possibility that the linear system generated to determine the regression coefficients is unsolvable, i.e. the matrix of the coefficients is singular with determinant equal to zero. This situation is indicated by the message "singular matrix" written to both the Dasher and output file OUTR, and the computation will terminate.

Double precision arithmetic is used to minimize the chances of encountering a singular matrix, and it will probably never occur, if the independent variables are carefully selected. If a singular matrix does occur, one of the variables must be eliminated. Usually a reconsideration of the independent variables will reveal two that are very similar; elimination of one of them will probably solve the problem. If need be, the trial and error method can be used, eliminating one at a time until the program does run. However, the problem will probably not arise, since double precision makes a singular matrix a rare occurrence. Use of double precision arithmetic limits the number of independent variables to 39, but this is probably more than sufficient for most applications.

There is an additional problem with the backward elimination process. Running times with regression screening varies approximately exponentially with the number of independent variables. Computing with a data file of approximately 39000 bytes, the following run times were obtained as the number of independent variables N was changed:

N	Time
6	2.8 minutes
11	7.1 "
17	19.4 "
23	51.9 "

The equation, Time = $e^{.171N}$, fits the above data fairly well. The program has not been run for more than 23 variables, but the equation predicts a runtime of about 13 hours, if all 39 variables are run on this size data file. This obviously puts a practical limitation on the number of independent variables that can be handled.

The variation of runtimes with the size of the data file IDXCR1 is approximately linear, so this does not present a serious problem. The only limitation on size of the data sample IDXCR1 is storage space available on the disk.

The first line of the IDXCR1 file must consist of exactly 80 characters. Any deviation from this may cause an error in reading the records in the file.

IV. References

Draper, N.R. and Smith, H., 1966: Applied Regression Analysis, John Wiley & Sons, 407 pp., 167-169.

Ezekiel, M., 1941: Methods of Correlation Analysis, John Wiley & Sons, 531 pp., 214-215.

Schwein, T.F., 1983: Meteorological Applications Library (CRH.LB), NOAA Central Region Computer Programs and Problems, NWS CRCP - No. 11, National Weather Service, Kansas City, MO.

Spiegel, M.R., 1961: Schaum's Outline of Theory and Problems of Statistics, Schaum Publishing Co., 359 pp., 243-244.

Walker, H.M. and Lev, J., 1953: Statistical Inference, Henery Holt and Co., 510 pp., 262-263.

Acknowledgement

Thanks to Fred Zuckerberg for assistance in locating the appropriate statistical literature, upon which, this work is based. Thanks also to Fortune Vilcko for the typing of this manuscript.

V. Program Information and Procedures for Installation and Execution

Correlation and Regression Equation Program - REGRS

Part A: PROGRAM INFORMATION AND INSTALLATION PROCEDURES

PROGRAM NAME: REGRS.SV

AAL ID:
REV NO.: 01.00

FUNCTION: The program computes linear correlation coefficients between a dependent variable and several independent (predictor) variables. Both single and multiple correlation coefficients are calculated along with coefficients for a regression equation. May also do screening by backward elimination process to find "best" regression equation.

PROGRAM INFORMATION:

Development Programmer:
Hugh Stone
Location: ERH Garden City, NY
Phone: FTS 649-5443
Language: DG FORTRAN IV/5.20
Save File Creation Date: 4/26/85
Running Time: Linear variation with size of the data file IDXCR1 and approximately exponential variation with the number of dependent variables.
For IDXCR1 file of 39000 bytes:
Runtime: 11 variables = 7.1 minutes
23 variables = 51.9 minutes
39 variables $\doteq$ 13 hours (estimated)

Maintenance Programmer:
Hugh Stone

Type: Standard

Disk Space:

REGRS.SV	39 RDOS Blocks
OUTR	Variable (2 to 50 blocks usually)
IDXCR1	Variable (about 25 blocks minimum, 100-200 blocks more likely)

PROGRAM REQUIREMENTS

Program Files:

Name	Comments
REGRS.SV	

Data Files;

Name	DP Location	Read/Write	Comments
IDXCR1	DPØ	R	First line must contain 80 characters before data begins.
OUTR	DPØ	W	Output File

AFOS Products: None

LOAD LINE

RLDR REGRS SYSL REDUC RPMIN TDIF UTIL.LB FORT.LB

PROGRAM INSTALLATION

1. REGRS.SV should be on DPØ or DPØF with link to DPØ.
2. File OUTR will be on DPØ, unless linked to DPØF.
3. IDXCR1 should be on DPØ or DPØF with link to DPØ.

Correlation and Regression Equation Program - REGRS

PART B: PROGRAM EXECUTION AND ERROR CONDITIONS

PROGRAM NAME: REGRS.SV

AAL ID:
REV NO.: 01.00

PROGRAM EXECUTION

1. At Dasher type: REGRS

2. Five responses will be requested:

   a) No. of variables per record in IDXCR1?

   b) Format for reading data in IDXCR1?
      Type in format including parentheses.

   c) Specify dependent variable and independent variables, in IDXCR1, using the code:
      2=dependent, 1=independent, 0=do not use.
      For example, 02011101 denotes
      dependent variable = 2
      independent variables = 4, 5, 6, 8
      variables 1, 3, 7 are not used.

   d) Type dependent variable name or ID.
      (optional).

   e) Simple correlation only?
      Response is 'Y' yes or 'N' no.
      Y denotes no screening process.
      N denotes screening to be done.

3. Computer will type "Computation in Progress" and beginning time.

4. When finished output will be in file OUTR, which may be printed.

ERROR CONDITIONS

The only error condition is caused by using independent variables that are not truly independent, i.e. one variable may be very similar to or a function of some other variable or variables. If this occurs, the message "Singular Matrix" will appear both on the Dasher and in file OUTR and the computation will terminate. The solution to the problem is that one of the variables must be eliminated.

DASHER MESSAGE	MEANING
1. SINGULAR MATRIX	One of the independent variables is not really independent, and must be eliminated.

```
HOUR=12  PBS      MDRMIN=6 1111111   0  0 0   0.  0.                    XXXXXXXXXXXX
  -13.    1.    9.   11.    11.    .11    0. .0042
34012389.
    3.   12.    1.   23.    23.    .23    1. .0012
   97129.
   14.    2.    3.  109.  1090.   1.09    1. .0083
    2791.
  -10.   14.    5.   97.    97.    .97    0. .0033
 234901.
    7.    8.    3.    9.     9.    .09    1. .0025
   34918.
  -12.    5.    2.  115.   115.   1.15    0. .0060
12922567.
```

Fig. 1. Sample of part of IDXCR1 file. First line must consist of eighty characters and may be used for identification. Six data records follow containing 9 variables, which may be read using FORMAT (8F6.0/F9.0).

```
REGRS
STRUCTURE OF 'IDXCR1' FILE:
TYPE NO. OF VARIABLES PER RECORD IN IDXCR1, FORMAT (I2)
09
KKZ =  9
TYPE (FORMAT FOR READING DATA IN IDXCR1), USE (***)
(8F6.0/F9.0)
TYPE '0'=DON'T USE, OR '1'=USE, FOR EACH INDEPENDENT VARIABLE
 AND '2' FOR DEPENDENT VARIABLE IN FILE 'IDXCR1', FORMAT (69I1)
201110101
TYPE DEPENDENT VARIABLE NAME OR ID,  FORMAT(40A2)
TEST VRBL
SIMPLE CORRELATION ONLY?...TYPE 'Y' OR 'N'
.....
COMPUTATION IN PROGRESS  18 34 16
MULTIPLE CORRELATION COEFFICIENTS
     5.    4     3     2     1R
```

Fig. 2. Sample of Dasher printout, when REGRS was run using IDXCR1 file shown partially in Fig. 1 above. Underlined characters were typed from the keyboard. Parentheses must be included in the format specification. 'N' was typed in response to the last question and indicates program to do screening, eliminating variables from the regression equation. Numbers on bottom line are printed one at a time as computation progresses through the screening process. Numbers 18 34 16 indicate starting time (hr, min, sec) of computation.

```
REGRESSION ANALYSIS -- BACKWARD ELIMINATION PROCESS
  HOUR=12  PBS      MDRMIN=6 0111111   0  0 0   0.  0.
  DEPENDENT VARIABLE NO.  2    SVR WX 2ND SIX HRS
  YMEAN =     0.93834D -1      SY =     0.29160D  0
  YMAX  =     0.10000D  1    YMIN =     0.00000D  0
  INDEPENDENT VRBLS =  4     NO. OBSERVATIONS =    373.
  02000000000000010001100000001

SIMPLE CORRELATION COEFFICIENTS
 VRBL          R        VRBL           R        VRBL          R
  16     0.20791         20     0.17418          21     0.04254
  29     0.35803

MULTIPLE CORRELATION COEF, MCC = 0.41736

REGRESSION EQUATION COEFFICIENTS :
VRBL            COEF
             -0.19360D  0
  16         -0.19400D -3
  20          0.49716D -2
  21          0.11352D -2
  29          0.61662D -3

ELAPSED TIME :    1 MIN  5 SEC
```

Fig. 3. Sample of OUTR file, without the variable elimination process.
A guide to the various lines of output:

Line 2...identification only, read from first line of IDXCR1 file.

Line 3...indicates that dependent variable 'Y' is no. 2 in IDXCR1 file. Second half of the line is optional identification of variable no. 2, and is typed in from Dasher keyboard at beginning of the run.

Line 4...YMEAN = mean value of Y, SY = standard deviation of Y.

Line 5...YMAX = maximum value of Y in the sample, YMIN = minimum value. In this case Y is binary, assuming only values of '0' or '1', and all correlation coefficients are of the point biserial type.

Line 6...indicates that 4 independent variablesare being used and IDXCR1 contains a sample of 373 observations.

Line 7...a duplication of the variable specification line typed on Dasher at beginning of the run. The dependent variable is indicated by the number '2' and independent variables by '1'. Those variables indicated '0' are ignored during the computation. In this case there are 29 variables in each record of IDXCR1.

Simple correlation coefficients between Y and the various independent variables are given in the next section.
MCC is the multiple correlation between Y and all the independent variables.
Regression equation coefficients are given last, with the constant term on the unlabeled line at the top.
Elapsed time of the computation is given on the last line.

```
REGRESSION ANALYSIS -- BACKWARD ELIMINATION PROCESS
  HOUR=12  PBS     MDRMIN=6 0111111   0  0 0   0.  0.
  DEPENDENT VARIABLE NO.  2   SVR WX 2ND SIX HRS
  YMEAN =     0.93834D -1      SY =     0.29160D  0
  YMAX  =     0.10000D  1   YMIN =     0.00000D  0
  INDEPENDENT VRBLS =  4    NO. OBSERVATIONS =    373.
  020000000000000100011000000001

SIMPLE CORRELATION COEFFICIENTS
 VRBL          R        VRBL          R        VRBL          R
  16      0.20791        20      0.17418        21      0.04254
  29      0.35803

MULTIPLE CORRELATION AND REGRESSION COEFFICIENTS

N =  4
    MCC =  0.41736
VRBL          MPC           PCC           COEF
                                      -0.19360D  0
  16        0.41574       0.04036    -0.19400D -3
  20        0.39063       0.15963     0.49716D -2
  21        0.41408       0.05735     0.11352D -2
  29        0.28407       0.31890     0.61662D -3

N =  3
    MCC =  0.41574
VRBL          MPC           PCC           COEF
                                      -0.17374D  0
  20        0.39023       0.15575     0.48110D -2
  21        0.41050       0.07217     0.13686D -2
  29        0.18067       0.38070     0.56782D -3

N =  2
    MCC =  0.41050
VRBL          MPC           PCC           COEF
                                      -0.14552D  0
  20        0.35803       0.21506     0.58853D -2
  29        0.17418       0.37748     0.53807D -3

N =  1
    MCC =  0.35803
VRBL          MPC           PCC           COEF
                                      -0.42248D -1
  29        0.00000       0.35803     0.51689D -3

ELAPSED TIME :   3 MIN  0 SEC
```

Fig. 4. Sample of OUTR file, when the elimination process is used. Format is the same as in Fig. 3, except intermediate results are printed out for each subset of independent variables. PCC denotes partial correlation coefficient. MPC denotes the multiple correlation when the variable listed in the first column is removed. The variable with minumum PCC is removed, and the process continues until only one remains. In this case, variable 29 and 20 combined yield a fairly good correlation (MCC) with little additional gain when variables 21 and 16 are included.

MAIN PROGRAM

REGRS

SUBROUTINE

SYSL
REDUC
RPMIN
TDIF

LOAD LINE

```
RLDR REGRS SYSL REDUC RPMIN TDIF
  UTIL.LB FORT.LB
```

Fig. 5. Software Structure and Load Line for program REGRS.

Appendix

Derivation of Normal Equations for Linear Regression Coefficients

A linear equation of the form,

$$Y' = C_0 + C_1X_1 + C_2X_2 + \ldots\ldots + C_nX_n \tag{1}$$

is needed, where Y' is the predictand variable and $X_1, X_2, \ldots X_n$ are the predictor variables. The coefficients $C_0, C_1, C_2, \ldots C_n$ need to be determined from a set of "m" observations of Y and the predictors $X_1, X_2, \ldots X_n$. The regression equation (1) may be written in a more compact form:

$$Y_i' = \sum_{j=0}^{n} C_jX_{ij} \quad , \text{ where } X_{io}=1 \tag{2}$$

and may be considered an estimate of the predictand Y for each of the observations $i=1,\ldots m$.

The coefficients C_j are to be selected so that the sum of the squares of the errors E is minimized:

$$E = \sum_{i=1}^{m} (Y_i - Y_i')^2 \tag{3}$$

Substituting (2) into (3):

$$E = \sum_{i=1}^{m} (Y_i - \sum_{j=0}^{n} C_jX_{ij})^2$$

$$E = \sum_{i=1}^{m} \left[Y_i^2 - 2Y_i\sum_{j=0}^{n} C_jX_{ij} + (\sum_{j=0}^{n} C_jX_{ij})^2\right] \tag{4}$$

The error E is a function of the coefficients $C_0, C_1, \ldots C_n$ and in order for E to be a minimum, the following condition must be met:

$$\frac{\partial E}{\partial C_k} = 0 \quad \text{for all } k = 0,1,\ldots n,$$

or from (4) :

$$\frac{\partial E}{\partial C_k} = \sum_{i=1}^{m} \left[0 - 2Y_i X_{ik} + 2\left(\sum_{j=0}^{n} C_j X_{ij}\right) X_{ik} \right] = 0$$

$$\boxed{\begin{array}{l} \therefore \displaystyle\sum_{j=0}^{n} C_j \sum_{i=1}^{m} X_{ij} X_{ik} = \sum_{i=1}^{m} Y_i X_{ik} \\ \text{for all } K = 0,1,\ldots n \text{ and } X_{io}=1 \text{ as in (2)} \end{array}} \qquad (5)$$

Equation (5) defines the so-called normal equations for the regression equation (1) and represents linear system of n+1 equations in n+1 variables, and may be written in matrix notation as,

$$A \cdot C = D \qquad (6)$$

where A is a symmetric matrix having $\Sigma X_j X_k$ as the element of the k row and j column (subscript i eliminated for simplicity), C is a column matrix with elements C_j, and D is a column matrix with elements ΣYX_k.

For example, if there were four predictors (n=4) in equation (5), then equation (6) would be,

$$\begin{pmatrix} m & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \Sigma X_4 \\ \Sigma X_1 & \Sigma X_1{}^2 & \Sigma X_1 X_2 & \Sigma X_1 X_3 & \Sigma X_1 X_4 \\ \Sigma X_2 & \Sigma X_1 X_2 & \Sigma X_2{}^2 & \Sigma X_2 X_3 & \Sigma X_2 X_4 \\ \Sigma X_3 & \Sigma X_1 X_3 & \Sigma X_2 X_3 & \Sigma X_3{}^2 & \Sigma X_3 X_4 \\ \Sigma X_4 & \Sigma X_1 X_4 & \Sigma X_2 X_4 & \Sigma X_3 X_4 & \Sigma X_4{}^2 \end{pmatrix} \cdot \begin{pmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma YX_1 \\ \Sigma YX_2 \\ \Sigma YX_3 \\ \Sigma YX_4 \end{pmatrix} \qquad (7)$$

The solution of the above system will provide the coefficients C_j required for the regression equation (1). The system will have a solution provided that the matrix of coefficients A is not singular, i.e. determinant of A is non-zero, and this will be true, if all the predictor variables are independent.

```
C       PROGRAM REGRS                                    REV 01.00
C       APR 1985              STONE, H. M.               ERH SSD/FTS 649-5443
C       FORTRAN IV/ REV 5.20  DG ECLIPSE (S230)   RDOS/REV 7.20
C       LOAD LINE:  RLDR REGRS SYSL REDUC RPMIN TDIF UTIL.LB FORT.LB
C       PURPOSE
C           COMPUTES LINEAR CORRELATION COEFFICIENTS BETWEEN A DEPENDENT
C           VARIABLE AND SEVERAL INDEPENDENT (PREDICTOR) VARIABLES WHICH
C           ARE READ FROM FILE 'IDXCR1'.  BOTH SINGLE AND MULTIPLE
C           CORRELATION COEFFICIENTS ARE CALCULATED ALONG WITH
C           COEFFICIENTS FOR A REGRESSION EQUATION USING ALL THE PREDICTOR
C           VARIABLES.
C           FORMAT FOR READING 'IDXCR1' IS INPUT AT RUNTIME.
C           MAY ALSO DO SCREENING BY BACKWARD ELIMINATION PROCESS TO FIND
C           'BEST' REGRESSION EQUATION.  ALL INDEPENDENT VARIABLES ARE
C           USED IN THE INITIAL EQUATION, THEN ONE VARIABLE AT A TIME IS
C           DROPPED, COMPUTING PARTIAL CORRELATION COEFFICIENT FOR EACH.
C           THE ONE WITH MINIMUM CORRELATION IS ELIMINATED AND PROCEDURE
C           CONTINUES UNTIL ONLY ONE REMAINS.
C       CHANNELS/FILES
C           CHANNEL 20...IDXCR1
C           CHANNEL 21...OUTR
C       EXITS
C           STOP 'SINGULAR MATRIX '
C
C     FORMAT STATEMENT 65, 66 & 80 MAY BE CHANGED, IF PARAMETER L1 IS CHANGED.
        COMPILER DOUBLE PRECISION
C     L = NO. OF INDEPENDENT VARIABLES, L1 = L+1, L2 = L+2, L3 = L+11
C     L = 39 IS MAXIMUM ALLOWABLE, WHEN USING DOUBLE PRECISION ARITHMETIC
        PARAMETER L=39, L1=40, L2=41, L3=50
        DIMENSION A(L1,L2),B(L1,L2),CO(0:L),C(L,0:L),X(0:L1),
     1  TP(L1),RMP(L1),M(L),RP(L),T(40),CS(0:L),ITITLE(40),
     2  X1(L3),KX(L3),KV(L),KT(L),FT(80),XBAR(L),TOP1(L),BOT1(L),R(L),
     3  ITAR(3),ISAR(3),ITOT(3)
C       DIMENSION D(L1)    ;   FOR ERROR TEST
        ISTOP=0
        CALL DFILW("OUTR",IER)
        CALL CFILW("OUTR",2,IER)
        CALL OPEN (20,"IDXCR1",1,IER)
        IF (IER.NE.1) TYPE "OPEN 20, IER = ",IER
        CALL OPEN (21,"OUTR",3,IER)
        IF (IER.NE.1) TYPE "OPEN 21, IER = ",IER
        TYPE "STRUCTURE OF 'IDXCR1' FILE:"
        TYPE "TYPE NO. OF VARIABLES PER RECORD IN IDXCR1, FORMAT (I2)"
        READ (11,1) KKZ
1       FORMAT (I2)
        WRITE (10,69) KKZ
69      FORMAT (1H ,"KKZ = ",I2)
        TYPE "TYPE (FORMAT FOR READING DATA IN IDXCR1), USE (***)"
        READ (11,32) FT(1)
32      FORMAT (S80)
        TYPE "TYPE '0'=DON'T USE, OR '1'=USE, FOR EACH INDEPENDENT VARIABLE"
        TYPE " AND '2' FOR DEPENDENT VARIABLE IN FILE 'IDXCR1', FORMAT (68I1)"
        READ (11,65) (KX(I),I=1,L3)
65      FORMAT (68I1)  ;  CHANGE THIS FORMAT, IF 'L3' IS CHANGED
        NI=0
        DO 70 I=1,L3    ;    DETERMINE WHICH VARIABLES TO USE
        KX1=KX(I)+1
        GO TO (70,72,73), KX1
73      KY=I   ;   NUMBER OF DEPENDENT VARIABLE
        GO TO 70
```

```
72      NI=NI+1    ;    COUNT # OF INDEPENDENT VARIABLES
        KV(NI)=I   ;    NUMBER EACH INDEPENDENT VARIABLE
        KT(NI)=I   ;    NUMBER ID FOR PRINTER OUTPUT
70      CONTINUE
        KLAST=MAX0(KV(NI),KY)
        TYPE "TYPE DEPENDENT VARIABLE NAME OR ID.  FORMAT(40A2)"
        READ (11,45) (T(I),I=1,40)
45      FORMAT (40A2)
        TYPE "SIMPLE CORRELATION ONLY?...TYPE 'Y' OR 'N'"
        CALL GCHAR (ISIMPLE,IER)
        CALL TIME (ITAR,IER)    ;    START TIME
        TYPE "....."
        WRITE (10,99) (ITAR(I),I=1,3)   ;    WRITE START TIME
99      FORMAT (" COMPUTATION IN PROGRESS ",3I3)
        DO 27 I=1,NI
27      M(I)=I
        NI1=NI+1
        NI2=NI+2
        DO 4 I=1,NI1
        DO 4 J=1,NI2
4       A(I,J)=0.
        X(0)=1.
        YMIN=1.E+50
        YMAX=1.E-50
        READ (20,62) (ITITLE(IT),IT=1,40)
62      FORMAT (40A2)
3       READ (20,FT,END=13) (X1(I),I=1,KKZ)
C
        Y=X1(KY)
        IF (Y.LT.YMIN) YMIN=Y    ;    FIND MINIMUM VALUE OF Y
        IF (Y.GT.YMAX) YMAX=Y    ;    FIND MAXIMUM VALUE OF Y
        DO 67 I=1,NI
67      X(I)=X1(KV(I))   ;   SET X'S
        X(NI1)=Y
C   FIRST LINE OF MATRIX
        DO 5 J=1,NI2
5       A(1,J)=A(1,J)+X(J-1)
C   SUBSEQUENT LINES OF MATRIX
        DO 6 I=2,NI1
        I1=I-1
        DO 7 J=I,NI1
7       A(I,J)=A(I,J)+X(I1)*X(J-1)
C   AUGMENTED MATRIX
6       A(I,NI2)=A(I,NI2)+X(I1)*Y
        GO TO 3
13      CONTINUE
        REWIND 20
C   FILL LOWER LEFT OF SYMMETRIC MATRIX
        DO 8 I=2,NI1
        I1=I-1
        DO 8 J=1,I1
8       A(I,J)=A(J,I)
        AM=A(1,1)    ;    NO. OF OBSERVATIONS
        YBAR=A(1,NI2)/A(1,1)    ;    MEAN VALUE OF DEPENDENT VARIABLE
        DO 74 I=1,NI
74      XBAR(I)=A(1,I+1)/A(1,1)   ;   MEAN VALUE OF INDEPENDENT VARIABLES
        DO 15 I=1,NI1
        DO 15 J=1,NI2
15      B(I,J)=A(I,J)    ;    SAVE ORIGINAL MATRIX IN 'B'
        CALL SYSL(A,NI1,CO,IL,$85)    ;    SOLVE SYSTEM
```

```
      GO TO 86   ;   NORMAL
85    WRITE (10,87) IL
87    FORMAT (1X,"IL = ",I2," SINGULAR MATRIX")
      DO 108 I=0,NI
108   CO(I)=0.
      ISTOP=1
86    DO 56 I=0,NI
56    CS(I)=CO(I)    ;    SAVE ORIGINAL COEFFICIENTS
C
      TOP=0.
      BOT=0.
      DO 75 I=1,NI
      TOP1(I)=0.
75    BOT1(I)=0.
      READ (20,62) (ITITLE(IT),IT=1,40)
29    READ (20,FT,END=10) (X1(I),I=1,KKZ)
      Y=X1(KY)
      DO 68 I=1,NI
68    X(I)=X1(KV(I))   ;   SET X'S
      YE=CO(0)
      DO 9 I=1,NI
9     YE=YE+CO(I)*X(I)
      YEB=YE-YBAR
      YB=Y-YBAR
      TOP=TOP+YEB*YEB
      BOT=BOT+YB*YB
      DO 76 I=1,NI
      XB=X(I)-XBAR(I)
      TOP1(I)=TOP1(I)+XB*YB
76    BOT1(I)=BOT1(I)+XB*XB  ;  FOR SIMPLE COREL COEF
      GO TO 29
10    CONTINUE
      REWIND 20
      RM=(TOP/BOT)**.5   ;   MULTIPLE COREL COEF
      SY=(BOT/AM)**.5   ;   STANDARD DEVIATION OF Y
      DO 77 I=1,NI
      BT=BOT1(I)*BOT
77    R(I)=TOP1(I)/BT**.5  ;  SIMPLE COREL COEFFICIENT
C
C   WRITE TITLE INFORMATION
      WRITE (21,46)
46    FORMAT (1H ,"REGRESSION ANALYSIS -- BACKWARD ELIMINATION PROCESS")
      WRITE (21,63) (ITITLE(I),I=1,40)  ;  WRITE ID FROM IDXCR1 FILE
63    FORMAT (3X,40A2)
      WRITE (21,47) KY,(T(I),I=1,40)
47    FORMAT (3X,"DEPENDENT VARIABLE NO. ",I2,3X,40A2)
      WRITE (21,59) YBAR,SY
59    FORMAT (3X,"YMEAN = ",E14.5,5X,"SY = ",E14.5)
      WRITE (21,60) YMAX,YMIN
60    FORMAT (3X,"YMAX  = ",E14.5,3X,"YMIN = ",E14.5)
      WRITE (21,48) NI,AM
48    FORMAT (3X,"INDEPENDENT VRBLS = ",I2,4X,"NO. OBSERVATIONS = ",F6.0)
      WRITE (21,66) (KX(I),I=1,KLAST)
66    FORMAT (3X,68I1)  ;  CHANGE THIS FORMAT, IF 'L3' IS CHANGED
C
      WRITE (21,79)
79    FORMAT (1H0,"SIMPLE CORRELATION COEFFICIENTS"/2X,3("VRBL",10X,"R",6X))
      WRITE (21,80) (KV(I),R(I),I=1,NI)
80    FORMAT (19(3X,3(I2,4X,F8.5,7X)/))
      IF (ISTOP.EQ.0) GO TO 106    ;    NORMAL
```

```
      WRITE (21,107) IL
107   FORMAT (1H0,"IL = ",I2,"  SINGULAR MATRIX")
      GO TO 91   ;  STOP COMPUTATION
106   IF (ISIMPLE.EQ.89) GO TO 94  ;  SIMPLE CORRELATION ONLY
      KA=MOD(NI,3)
      IF (KA.NE.0) WRITE (21,81)
81    FORMAT (1H )  ;  BLANK LINE
      WRITE (10,83)
83    FORMAT (1H ,"MULTIPLE CORRELATION COEFFICIENTS")
      WRITE (21,92)
92    FORMAT (" MULTIPLE CORRELATION AND REGRESSION COEFFICIENTS")
C
C
C   MAIN LOOP
      DO 44 KK=1,NI
      KI=NI+1-KK
      NI1=KI+1
      NI2=KI+2
      N=KI-1
C
      KA=MOD(KK,13)
      IF (KA.EQ.0) WRITE (10,81)   ;  BLANK LINE
      CALL TYPED(KI)  ;  TO MONITOR PROGRESS
C
C   ELIMINATE ONE LINE & ONE COLUMN FROM ORIGINAL MATRIX FOR EACH OF THE
C   INDEPENDENT VARIABLE IN TURN
      DO 14 K=1,KI   ;   ELIMINATE ONE VARIABLE AT A TIME
      CALL REDUC (K,KI,NI1,A,B)  ;   REDUCES MATRIX
      CALL SYSL (A,KI,CO,IL,$88)  ;   SOLVE EACH SUBSYSTEM
      GO TO 89
88    WRITE (10,90) KK,K,IL,KT(IL)
      WRITE (21,90) KK,K,IL,KT(IL)
90    FORMAT (1H0,"KK = ",I2,"  K = ",I2,"  KT(",I2,") = ",I2,
     1 "  SINGULAR MATRIX")
      GO TO 91   ;  STOP COMPUTATION
89    DO 11 I=0,N
11    C(K,I)=CO(I)   ;   SAVE SOLUTION TO EACH SYSTEM
14    CONTINUE
C
      DO 28 I=1,KI
28    TP(I)=0.
      READ (20,62) (ITITLE(IT),IT=1,40)
30    READ (20,FT,END=12) (X1(I),I=1,KKZ)
      DO 71 I=1,NI
71    X(I)=X1(KV(I))   ;   SET X'S
      DO 24 K=1,KI
      YE=C(K,0)
      DO 25 J=1,N
      JK=M(J)
      IF (J.GE.K) JK=M(J+1)
25    YE=YE+C(K,J)*X(JK)
      YEB=YE-YBAR
24    TP(K)=TP(K)+YEB*YEB
      GO TO 30
12    CONTINUE
      REWIND 20
      VAR=1.-RM*RM   ;   VARIANCE
      DO 110 K=1,KI
      RMP(K)=(TP(K)/BOT)**.5   ;   MULTIPLE COREL COEF FOR EACH SUBSYSTEM
      IF (RMP(K).LT.RM) GO TO 26   ;   NORMAL
```

```
      WRITE (10,109) KI,K,RMP(K),RM
109   FORMAT (1X,"KI = ",I2,2X,"RMP(",I2,") = ",E20.8,3X,"RM = ",E20.8,
     1 /1X,"CORRECTION")
      RP(K)=0.
      GO TO 110
26    RP(K)=(1.-VAR/(1.-RMP(K)*RMP(K)))**.5   ;  PARTIAL COREL COEF
110   CONTINUE
      WRITE (21,49) KI
49    FORMAT (1H0,"N = ",I2)
      WRITE (21,58) RM
58    FORMAT (5X,"MCC =  ",F7.5)
C   INSERT ERROR TEST HERE
      WRITE (21,50)
50    FORMAT (1H ,"VRBL",10X,"MPC",10X,"PCC",10X,"COEF")
      WRITE (21,51) CS(0)
51    FORMAT (33X,E14.5)
      DO 52 I=1,KI
52    WRITE (21,55) KT(I),RMP(I),RP(I),CS(I)
55    FORMAT (3X,I2,7X,F7.5,7X,F7.5,E14.5)
      CALL RPMIN (KI,RP,M,JELIM,KT)  ;  FIND MINIMUM RP AND ELIMINATE IT.
      RM=RMP(JELIM)   ;   RESET MULTIPLE COREL COEF
C   RECONSTRUCT MATRIX WITH MINIMUM RP VARIABLE REMOVED
      CALL REDUC (JELIM,KI,NI1,B,B)
      DO 57 I=0,N
57    CS(I)=C(JELIM,I)  ;  SAVE SOLUTION FOR SYSTEM WITH RP MIN REMOVED
44    CONTINUE
      GO TO 91
C   MULTIPLE CORRELATION COEF & REGRESSION COEFS WITHOUT ELIMINATION PROCESS
94    WRITE (21,93) RM
93    FORMAT (1H0,"MULTIPLE CORRELATION COEF, MCC = ",F7.5/)
      WRITE (21,95)
95    FORMAT (" REGRESSION EQUATION COEFFICIENTS : "/" VRBL",11X,"COEF")
      WRITE (21,96) CS(0)
96    FORMAT (12X,E14.5)
      DO 97 K=1,NI
97    WRITE (21,98) KT(K),CS(K)
98    FORMAT (3X,I2,7X,E14.5)
91    CALL TIME (ISAR,IER)  ;  FINISH TIME
      CALL TDIF (ISAR,ITAR,ITOT)   ;   GET ELAPSED TIME
      WRITE (21,84) (ITOT(I),I=2,3)
84    FORMAT (1H0,"ELAPSED TIME : ",I3," MIN",I3," SEC")
      CALL CLOSE (20,IER)
      CALL CLOSE (21,IER)
      STOP
      END
```

*

*

```
      COMPILER DOUBLE PRECISION
      PARAMETER L=39, L1=40, L2=41
      SUBROUTINE SYSL (A,N1,CO,IL,Q)
C                          REV 01.00
C     APR 1985             STONE, H. M.          ERH SSD/FTS 649-5443
C     FORTRAN IV/ REV 5.20 DG ECLIPSE (S230)     RDOS/REV 7.20
C     PURPOSE
C        SOLVES LINEAR SYSTEM OF 'N' EQUATIONS IN 'N' UNKNOWNS.
C     ARGUMENT LIST
C        A               - AUGMENTED MATRIX OF COEFFICIENTS OF THE SYSTEM
C        N1              - NUMBER OF EQUATIONS (OR UNKNOWNS)
```

```
C         CO                 - SOLUTION OF SYSTEM
C         IL                 - NO. OF THE EQUATION, WHERE DIVISION BY ZERO
C                              MAY OCCUR; SINGULAR MATRIX, NO SOLUTION
C         Q                  - ABNORMAL RETURN (SINGULAR MATRIX)
C
      INTEGER Q
      DIMENSION A(L1,L2),CO(0:L)
      N2=N1+1
      DO 4 I=1,N1
      ADIV=A(I,I)
      IL=I
      IF (ADIV.EQ.0.) RETURN Q  ;  PREVENTS DIVISION BY ZERO
      DO 1 J=1,N2
1     A(I,J)=A(I,J)/ADIV
      DO 3 K=1,N1
      IF(K.EQ.I) GO TO 3
      ADIV=A(K,I)
      DO 2 J=1,N2
2     A(K,J)=A(K,J)-ADIV*A(I,J)
3     CONTINUE
4     CONTINUE
      DO 5 J=1,N1
5     CO(J-1)=A(J,N2)
      RETURN
      END
```

*

*

```
      COMPILER DOUBLE PRECISION
      PARAMETER L1=40, L2=41
      SUBROUTINE REDUC (K,KI,NI1,A,B)
C                           REV 01.00
C     APR 1985              STONE, H. M.        ERH SSD/FTS 649-5443
C     FORTRAN IV/ REV 5.20  DG ECLIPSE (S230)   RDOS/REV 7.20
C     PURPOSE
C        REDUCES MATRIX B TO MATRIX A, BY ELIMINATING THE K+1 ROW AND
C        K+1 COLUMN OF MATRIX B.
C     ARGUMENT LIST
C        K                   - INDICATES ROW & COLUMN NO. OF B TO BE DELETED
C        KI                  - NO. OF ROWS IN MATRIX 'A'
C        NI1                 - NO. OF COLUMNS IN MATRIX 'A'
C        A                   - REDUCED MATRIX
C        B                   - ORIGINAL MATRIX
C
      DIMENSION A(L1,L2),B(L1,L2)
      DO 1 I=1,KI
      IM1=I-1
      I1=I
      IF (IM1.GE.K) I1=I+1
      DO 1 J=1,NI1
      JM1=J-1
      J1=J
      IF (JM1.GE.K) J1=J+1
      A(I,J)=B(I1,J1)
1     CONTINUE
      RETURN
      END
```

*

*

```
      COMPILER DOUBLE PRECISION
      PARAMETER L=39
      SUBROUTINE RPMIN (KI,RP,M,JELIM,KT)
C                           REV 01.00
C     APR 1985              STONE, H. M.          ERH SSD/FTS 649-5443
C     FORTRAN IV/ REV 5.20  DG ECLIPSE (S230)     RDOS/REV 7.20
C     PURPOSE
C         FINDS MINIMUM VALUE IN RP ARRAY FOR ELIMINATION AND
C         MAINTAINS LIST OF ORIGINAL VARIABLE NUMBERS 'KT' FOR THE
C         REMAINING VARIABLES
C     ARGUMENT LIST
C         KI                - NUMBER OF VARIABLES
C         RP                - PARTIAL CORRELATION COEFFICIENTS
C         M                 - INDEX USED IN MAIN PROGRAM
C         JELIM             - INDEX NO. OF THE LOWEST RP VALUE
C         KT                - ARRAY OF ORIGINAL VARIABLE NUMBERS FOR THE
C                             REMAINING VARIABLES, AFTER ONE ELIMINATED
C
      DIMENSION RP(L),M(L),KT(L)
      RMIN=RP(1)
      JELIM=1
      DO 1 I=1,KI
      IF (RP(I).GE.RMIN) GO TO 1
      RMIN=RP(I)
      JELIM=I
1     CONTINUE
      KI1=KI-1
C   RESET M VALUES, ELIMINATING ONE WITH MINIMUM PARTIAL COREL COEF (RP)
      DO 2 I=1,KI1
      IF (I.LT.JELIM) GO TO 2
      M(I)=M(I+1)
      KT(I)=KT(I+1)
2     CONTINUE
      RETURN
      END
```

*

*

```
      SUBROUTINE TDIF (ISAR,ITAR,ITOT)
C                           REV 01.00
C     JUN 1984              STONE, H. M.          ERH   FTS 649-5443
C     FORTRAN IV/ REV 5.20  DG ECLIPSE (S230)     RDOS/REV 7.20
C     PURPOSE
C         COMPUTES RUNNING TIME OF PROGRAM.
C         SUBROUTINE TIME MUST BE CALLED AT BEGINNING AND END OF
C         MAIN PROGRAM.
C     ARGUMENT LIST
C         ISAR              - FINISH TIME (HR,MIN,SEC)
C         ITAR              - START TIME (HR,MIN,SEC)
C         ITOT              - ELAPSED TIME. ITOT(2)=MIN, ITOT(3)=SEC
C
      DIMENSION ISAR(3),ITAR(3),ITOT(3)
      DO 1 I=1,3
1     ITOT(I)=ISAR(I)-ITAR(I)
      IF (ITOT(2).GE.0) GO TO 2
      ITOT(1)=ITOT(1)-1
      ITOT(2)=ITOT(2)+60
```

```
2       IF (ITOT(3).GE.0) GO TO 3
        ITOT(2)=ITOT(2)-1
        ITOT(3)=ITOT(3)+60
3       ITOT(2)=ITOT(2)+60*ITOT(1)   ;    CONVERT HRS TO MINS
        RETURN
        END
```

Eastern Region Computer Programs and Problems (Continued)

19 Verification of Asynchronous Transmissions. Lawrence Cedrone, March 1984. (PB84 189885)

20 AFOS Hurricane Plotter. Charles Little, May 1984. (PB84 199629)

21 WARN - A Warning Formatter. Gerald G. Rigdon, June 1984. (PB84 204551)

22 Plotting TDL Coastal Wind Forecasts, Paula Severe, June 1984 (Revised) (PB84-220789)

23 Severe Weather Statistics STADTS Decoder (SWX) and Plotter (SWY), Hugh M. Stone, June 1984. (PB84-213693)

24 WXR, Harold Opitz, August 1984. (PB84-23722)

25 FTASUM: Aviation Forecast Summaries, Matthew Peroutka, August 1984. (PB85-112977)

26 SAOSUM: A Short Summary of Observations. Matthew Peroutka, October 1984. (PB85-120384)

27 TRAJ - Single Station Trajectory Plot, Tom Niziol, December 1984. (PB85-135002)

28 VIDTEX, Gerald G. Rigdon, February 1985 (PB85-175669/AS)

29 ISENTROPIC PLOTTER, Charles D. Little, February 1985 (PB85-175651/AS)

30 CERR: An Aviation Verification Program, M. Peroutka, April 1985. (PB85-204824/AS)

NOAA SCIENTIFIC AND TECHNICAL PUBLICATIONS

The National Oceanic and Atmospheric Administration was established as part of the Department of Commerce on October 3, 1970. The mission responsibilities of NOAA are to assess the socioeconomic impact of natural and technological changes in the environment and to monitor and predict the state of the solid Earth, the oceans and their living resources, the atmosphere, and the space environment of the Earth.

The major components of NOAA regularly produce various types of scientific and technical information in the following kinds of publications:

PROFESSIONAL PAPERS — Important definitive research results, major techniques, and special investigations.

CONTRACT AND GRANT REPORTS — Reports prepared by contractors or grantees under NOAA sponsorship.

ATLAS — Presentation of analyzed data generally in the form of maps showing distribution of rainfall, chemical and physical conditions of oceans and atmosphere, distribution of fishes and marine mammals, ionospheric conditions, etc.

TECHNICAL SERVICE PUBLICATIONS — Reports containing data, observations, instructions, etc. A partial listing includes data serials; prediction and outlook periodicals; technical manuals, training papers, planning reports, and information serials; and miscellaneous technical publications.

TECHNICAL REPORTS — Journal quality with extensive details, mathematical developments, or data listings.

TECHNICAL MEMORANDUMS — Reports of preliminary, partial, or negative research or technology results, interim instructions, and the like.

Information on availability of NOAA publications can be obtained from:

ENVIRONMENTAL SCIENCE INFORMATION CENTER (D822)
ENVIRONMENTAL DATA AND INFORMATION SERVICE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

6009 Executive Boulevard
Rockville, MD 20852