POLYFIT PROGRAM USERS' GUIDE

A Statistical Polynomial
Curvefitting Program

Version 2.0
Professional Fortran 77

Developed by

U.S. DEPARTMENT OF TRANSPORTATION

TRANSPORTATION SYSTEMS CENTER

KENDALL SQUARE

CAMBRIDGE, MA. 02142

PREPARED FOR

FEDERAL AVIATION ADMINISTRATION

OFFICE OF ENVIRONMENT AND ENERGY

WASHINGTON, D.C. 20591

January 1988

POLYFIT PROGRAM USERS' GUIDE

1.0 DESCRIPTION

POLYFIT is a user friendly program written in Microsoft Fortran for the IBM PC/AT and compatibles to perform polynomial regression analysis. This program allows the user to create and modify XY files (ie, sample data containing from 5 to 50 XY pairs, X is the independent variable, Y is the dependent variable), or use an existing XY file as data input to the program. Menus are displayed to simplify program use. The program was designed to do polynomial curve fits for the given data set for orders 1, 2, and 3. A summary and plot of the fit for a selected order (1, 2, or 3) is provided, as well as the option of hardcopy (i.e. : printer output) if desired.

Confidence intervals are computed and displayed as part of the summary of the chosen order of curve fit for the given sample data.

A flow diagram of the program is included in Appendix A.

2.0 USE

2.1 Startup

To start the POLYFIT program, load the "TSC STAT PK" bootable disk into the "A" disk drive and boot up the computer (ie., turn on the computer or hit the CTRL-ALT-DEL keys simultaneously). A list of data files from the disk will be displayed, followed by a pause. All data files are assumed to have an extension of ".DAT". To input a file name, just type the name without the extension (eg, "DATA1"). The Main menu will be displayed as follows:

MAIN MENU

TYPE 0 - TO STOP
1 - TO CREATE AN XY FILE
2 - TO USE AN EXISTING XY DATA FILE
3 - TO MODIFY AN EXISTING XY FILE

TYPE DECISION:

2.2 Data Input

2.2.1 Create a New File

Type "1" from the Main menu to create a new file. A file name prompt will appear on the screen. The file name chosen must not already exist on the disk in the directory currently in use. The program will reprompt if this condition is not met. After inputting a valid file name, the user is prompted to input the number of observations (ie, X,Y pairs). The number of input pairs must be between 5 and 50, otherwise the program will reprompt.

The user is then asked to input the actual X, Y values. Each X value must be followed by a comma then a Y value followed by a RETURN/ENTER. If there is an error in data input an error message will appear and the data must be re-entered

The program displays the X, Y pairs along with an index number on the screen. Only 14 X, Y pairs will be shown at any time. To view additional data, hit the Enter/Return key to display the next consecutive 14 X, Y pairs.

2.2.2 Use an Existing File

Type "2" from the Main menu to use an existing data file. A file name prompt will appear on the screen. The file name entered must already exist on the disk or the program will reprompt.

After inputting a valid file name, the program opens the file and reads in the number of X, Y pairs and assigns an index number. Data is outputed to the screen and automatically sorted in increasing order on X.

2.3 Data Edit

Type "3" from the Main menu to modify an existing file. An Edit menu is displayed on the screen as follows:

```
                    EDIT MENU

TYPE 0 - TO STOP FILE EDIT
     1 - TO ADD X, Y PAIRS
     2 - TO DELETE X, Y PAIRS
     3 - TO MODIFY X, Y PAIRS

TYPE DECISION:
```

2.3.1 Edit an Existing File

A file name prompt will appear on the screen. The file to be edited must already exist in the directory. The program will reprompt if this condition is not met.

After inputting a valid file name, the program opens the

file and reads in the number of X, Y pairs. This data is output to the screen, and automatically sorted in increasing order of X.

2.3.1.1 Add an X, Y Pair

Type "1" from the Edit menu to add X, Y pairs. The program checks to see that the data file does not exceed 49 pairs. If the file does exceed 49 pairs an error message is displayed and the program pauses and then redisplays the data and the Edit menu. If the data file does not exceed 49 pairs, a prompt for a new X, Y pair is displayed. Each X value typed must be followed by a comma, then a Y value followed by a RETURN/ENTER. If there is an error in the data input, it must be retyped.

After all valid pairs are entered, the data is sorted and the number of observations are increased accordingly. The program then redisplays the complete data, followed by the Edit menu.

2.3.1.2 Delete an X, Y Pair

Type "2" from the Edit menu to delete X, Y pairs. The program checks to see that the data file exceeds 5 pairs. If the file does not exceed 5 pairs an error message is displayed and the program pauses. The data and the Edit menu are redisplayed.

If the data file does exceed 5 pairs, a prompt for the index number of the X, Y pair to delete is displayed. An invalid index number will result in an error message. A valid input number causes the program to delete the associated X, Y pair, and decreases the number of observations by 1. The data and the Edit menu are redisplayed.

2.3.1.3 Modify an X, Y Pair

Type "3" from the Edit menu to modify X, Y pairs. A prompt for the index number of the X, Y pair to be modified is displayed. If the number typed is invalid, an error message is displayed. A valid index number causes the program to prompt for a new X, Y pair. All data that is improperly entered, must be retyped.

When a valid pair is entered, the old X, Y pair associated with the index number is replaced, and the observation pairs and index numbers are resorted. The program then redisplays this data, followed by the Edit menu.

2.3.2 Terminate File Edit

Type "0" from the Edit menu to stop file edit. The file is transferred to the disk with the latest changes, and control returns to the Main menu.

2.4 Data Output

2.4.1 Output Selection

The polynomial regression analysis is performed for orders 1, 2, and 3. The program suggests an order of polynomial that best fits the plot of the data set. The program achieves this by selecting the degree where the correlation coefficient is closest to the value of one and where the residual variance produces the largest decrease over it's predecessor. This will insure that the variance about the reqression and the difference between observed and calculated values will be kept to a minimum. The user is cautioned to evaluate this suggested value by examining the statistical data and plotted curves, provided by the program, to see if the degree suggested indeed produces the best fit.

A Summary and Plot menu is displayed as follows:

```
        SUMMARY AND PLOT MENU
TYPE 0 - TO STOP
     1 - FOR SUMMARY & PLOT OF POLY. OF ORDER 1
     2 - FOR SUMMARY & PLOT OF POLY. OF ORDER 2
     3 - FOR SUMMARY & PLOT OF POLY. OF ORDER 3
     4 - TO CHOOSE A NEW DATA SET
TYPE DECISION:
```

2.4.2 Summary And Plot Display

Type "1", "2", or "3" from the Summary and Plot menu to get a summary data tabulation and plot of the respective order polynomial. The program will display least square coefficients, correlation coefficient, coefficient of determination, standard error of estimate, and residual variance of Y, on the top part of the screen followed by the index number, X (independent), Y (dependent), Y (calculated) and the confidence interval in columnar form. The program continues by generating the plot of the regression curve superimposed over the actual data points on the screen anywhere between 8 - 16 equidistant divisions of the X range. Observation of the fitted curve will help determine if the degree inputted best fits the data and that obvious erroneous data points has not forced a higher degree polynomial than necessary.

Type "4" from the Summary and Plot menu to choose a new data set. Choosing a new data set will take the user back to the Main menu.

2.4.3 Summary And Plot Hardcopy

After the summary data tabulation and plot is performed, a HARDCOPY menu is displayed as follows:

```
                HARDCOPY MENU

          TYPE 0 - TO CONTINUE
               1 - FOR HARDCOPY
               2 - CALC CONF INT FOR ADDITIONAL X VALUES
          TYPE DECISION:
```

Type "0" from this menu to return to the Summary and Plot menu. Type "1" from the Hardcopy menu to send a copy of the summary data tabulation and plot of the previously chosen polynomial to the printer anywhere between 20-40 equidistant divisions of the X range.

Type "2" from the HARDCOPY menu to calculate confidence intervals for additional X values. The program prompts the user for the number of X values. The program prompts to input the X values. If there is an error in data input a prompt will appear and the data must be retyped. The confidence intervals for these additional X values are computed and displayed on the screen. The program returns to the HARDCOPY menu. At this point, type "1" to obtain a hardcopy of the summary data tabulation, a plot of the previously chosen polynomial, and a table of confidence intervals for additional X values.

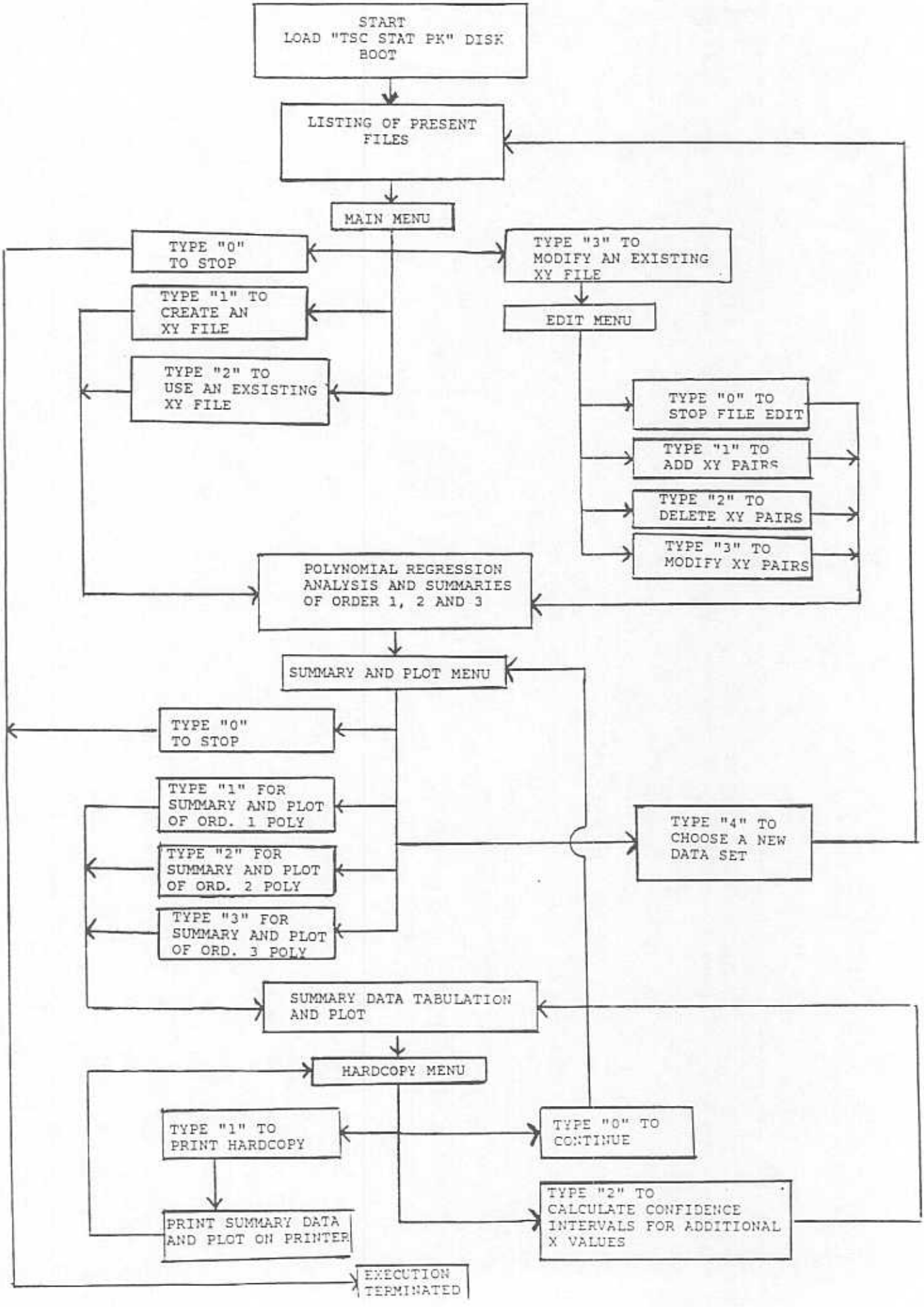
2.5 Polynomial Regression

After data is input, the regression analysis is performed on the sample data for polynomials of order 1, 2, and 3 with the results of each case displayed on the screen. See the program listings for the method used. Output examples are provided in

Appendix B.

2.6 Program Termination

When "0" is typed from the Main or Summary and Pl
all files connected to the program will be closed and
will return to the DOS operating system prompt.

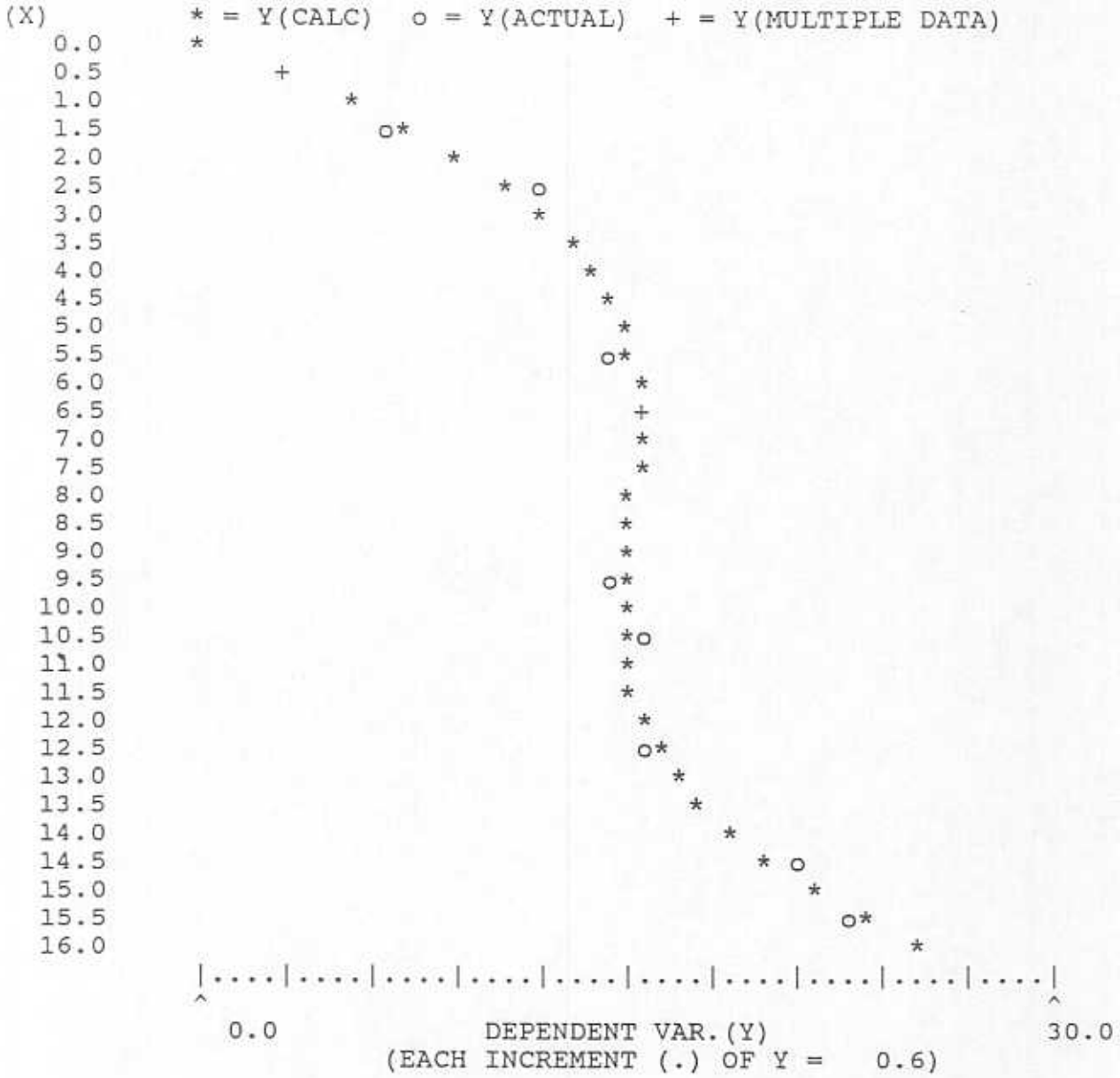

START
LOAD "TSC STAT PK" DISK
BOOT
LISTING OF PRESENT FILES
MAIN MENU
TYPE "0" TO STOP
TYPE "1" TO CREATE AN XY FILE
TYPE "2" TO USE AN EXSISTING XY FILE
TYPE "3" TO MODIFY AN EXISTING XY FILE
EDIT MENU
TYPE "0" TO STOP FILE EDIT
TYPE "1" TO ADD XY PAIRS
TYPE "2" TO DELETE XY PAIRS
TYPE "3" TO MODIFY XY PAIRS
POLYNOMIAL REGRESSION ANALYSIS AND SUMMARIES OF ORDER 1, 2 AND 3
SUMMARY AND PLOT MENU
TYPE "0" TO STOP
TYPE "1" FOR SUMMARY AND PLOT OF ORD. 1 POLY
TYPE "2" FOR SUMMARY AND PLOT OF ORD. 2 POLY
TYPE "3" FOR SUMMARY AND PLOT OF ORD. 3 POLY
TYPE "4" TO CHOOSE A NEW DATA SET
SUMMARY DATA TABULATION AND PLOT
HARDCOPY MENU
TYPE "1" TO PRINT HARDCOPY
PRINT SUMMARY DATA AND PLOT ON PRINTER
TYPE "0" TO CONTINUE
TYPE "2" TO CALCULATE CONFIDENCE INTERVALS FOR ADDITIONAL X VALUES
EXECUTION TERMINATED

APPENDIX B

```
            POLYFIT

A:\>ECHO OFF
--------------------------------
VOLUME IN DRIVE A IS TSC STAT PK
--------------------------------

SAMPLE   DAT

TYPE  0 - TO STOP
      1 - TO CREATE AN XY DATA FILE
      2 - TO USE AN EXISTING XY DATA FILE
      3 - TO MODIFY AN EXISTING XY FILE

TYPE DECISION: 2

FILE NAME: SAMPLE

NUMBER OF OBSERVATIONS:          10

READING DATA FILE: SAMPLE  .DAT

  I     X(INDEP)       Y(DEP)
  1        0.500         3.000
  2        1.500         7.000
  3        2.500        12.500
  4        5.500        14.500
  5        6.500        16.000
  6        9.500        14.500
  7       10.500        16.000
  8       12.500        16.000
  9       14.500        21.000
 10       15.500        23.000
Execution suspended : HIT ENTER KEY TO CONTINUE...


ORDER OF POLY = 1    ORDER OF POLY = 2    ORDER OF POLY = 3

LEAST SQ. COEFF.     LEAST SQ. COEFF.     LEAST SQ. COEFF.
B(1) = 0.65785E+01   B(1) = 0.53994E+01   B(1) = 0.49717E+00
B(2) = 0.98374E+00   B(2) = 0.15004E+01   B(2) = 0.57930E+01
                     B(3) =-0.32669E-01   B(3) =-0.71398E+00
                                          B(4) = 0.28146E-01

COEFF. OF DET.       COEFF. OF DET.       COEFF. OF DET.
   0.81611              0.83077              0.98047

CORR. COEFF.         CORR. COEFF.         CORR. COEFF.
   0.90339              0.91146              0.99019

STD ERROR OF EST.    STD ERROR OF EST.    STD ERROR OF EST.
   2.68456              2.75315              1.01026

RES. VAR. OF Y       RES. VAR. OF Y       RES. VAR. OF Y
   7.20688              7.57981              1.02062

 POLYNOMIAL OF ORDER   3   IS SUGGESTED FOR THE FOLLOWING MENU
Execution suspended : HIT ENTER KEY TO CONTINUE...
```

```
TYPE 0 - TO STOP
     1 - FOR SUMMARY & PLOT OF POLY. OF ORDER 1
     2 - FOR SUMMARY & PLOT OF POLY. OF ORDER 2
     3 - FOR SUMMARY & PLOT OF POLY. OF ORDER 3
     4 - TO CHOOSE A NEW DATA SET

TYPE DECISION: 3


ORDER OF POLY = 3

            LEAST SQ. COEFF.:
      0.49717E+00       0.57930E+01      -0.71398E+00       0.28146E-01

      CORR. COEFF.     COEFF. OF DET.   STD ERROR OF EST.    RES. VAR. OF Y
        0.99019           0.98047           1.01026            1.02062

  I     X(INDEP)       Y(DEP)      Y(CALC)     CONF INT
  1        0.500        3.000        3.219        1.615
  2        1.500        7.000        7.675        1.084
  3        2.500       12.500       10.957        1.072
  4        5.500       14.500       15.443        1.219
  5        6.500       16.000       15.715        1.122
  6        9.500       14.500       15.225        1.055
  7       10.500       16.000       15.189        1.136
  8       12.500       16.000       16.322        1.153
  9       14.500       21.000       20.187        1.131
 10       15.500       23.000       23.566        1.650
Execution suspended : HIT ENTER KEY TO CONTINUE...


(X)         * = Y(CALC)   o = Y(ACTUAL)   + = Y(MULTIPLE DATA)
   0.0      *    o
   1.0           * o
   2.0                *    o
   3.0                *
   4.0                   *
   5.0                   o*
   6.0                     +
   7.0                     *
   8.0                    *
   9.0                    o*
  10.0                     *o
  11.0                     *
  12.0                      +
  13.0                        *
  14.0                           *   o
  15.0                             * o
  16.0                                    *
          |....|....|....|....|....|....|....|....|....|....|
          ^                                                   ^
            0.0            DEPENDENT VAR.(Y)                 30.0
                    (EACH INCREMENT (.) OF Y =   0.6)

Execution suspended : HIT ENTER KEY TO CONTINUE...
```

```
TYPE 0 - TO CONTINUE
     1 - FOR HARDCOPY
     2 - CALC CONF INT FOR ADDITIONAL X VALUES

TYPE DECISION: 2

NUMBER OF X VALUES(15 MAX): 5

INPUT X VALUES:
?
5
10
13
15

  I     X(INDEP)      Y(CALC)     CONF INT
  1        2.000        9.452        1.035
  2        6.000       15.631        1.174
  3       10.000       15.175        1.096
  4       13.000       16.980        1.113
  5       15.000       21.738        1.321
Execution suspended : HIT ENTER KEY TO CONTINUE...


TYPE 0 - TO CONTINUE
     1 - FOR HARDCOPY
     2 - CALC CONF INT FOR ADDITIONAL X VALUES

TYPE DECISION: 1

 ORDER OF POLY = 3

              LEAST SQ. COEFF.:
        0.49717E+00        0.57930E+01       -0.71398E+00        0.28146E-01

       CORR. COEFF.     COEFF. OF DET.   STD ERROR OF EST.     RES. VAR. OF Y
          0.99019            0.98047            1.01026             1.02062


   I     X(INDEP)       Y(DEP)      Y(CALC)     CONF INT
   1        0.500        3.000        3.219        1.615
   2        1.500        7.000        7.675        1.084
   3        2.500       12.500       10.957        1.072
   4        5.500       14.500       15.443        1.219
   5        6.500       16.000       15.715        1.122
   6        9.500       14.500       15.225        1.055
   7       10.500       16.000       15.189        1.136
   8       12.500       16.000       16.322        1.153
   9       14.500       21.000       20.187        1.131
  10       15.500       23.000       23.566        1.650
```


(X) * = Y(CALC) o = Y(ACTUAL) + = Y(MULTIPLE DATA)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
0.0
DEPENDENT VAR.(Y)
30.0
(EACH INCREMENT (.) OF Y = 0.6)

```
  I     X(INDEP)      Y(CALC)     CONF INT
  1        2.000        9.452        1.035
  2        6.000       15.631        1.174
  3       10.000       15.175        1.096
  4       13.000       16.980        1.113
  5       15.000       21.738        1.321

Execution suspended : HIT ENTER KEY TO CONTINUE...


TYPE 0 - TO CONTINUE
     1 - FOR HARDCOPY
     2 - CALC CONF INT FOR ADDITIONAL X VALUES

TYPE DECISION: 0

TYPE 0 - TO STOP
     1 - FOR SUMMARY & PLOT OF POLY. OF ORDER 1
     2 - FOR SUMMARY & PLOT OF POLY. OF ORDER 2
     3 - FOR SUMMARY & PLOT OF POLY. OF ORDER 3
     4 - TO CHOOSE A NEW DATA SET

TYPE DECISION: 0

Execution terminated : 0


C:\FORTRAN1>
```

References

Draper, N.R. and H. Smith, Wiley Series in Probablity and Mathematical Statistic, (John Wiley & Sons, Inc. New York) 1966

IBM Application Program, "Programmer's Manual," System/360 Scientific Subroutine Package Version III, Program Number 360A-CM-03X, (International Business Machines Corporation) 1970.

IBM Application Program, "System Manual," System/360 Scientific Subroutine Package Vertion III, Program Number 360A-CM-03X, (International Business Machines Corporation) 1968.

Walpole, Ronald E. and Raymond H. Myers, Probability and Statistics for Engineers and Scientists Second Edition, (Macmillan Publishing Co., Inc., New York) 1978.