

INTERNAL REPORT 53

PARAMETER ESTIMATION MADE EASIER

ABSTRACT

The Coniferous Forest Biome is committing itself to the synthesis of the collected data. The computer programs documented in the following three internal reports have had a very useful track record in the synthesis of data and the computer implementation of models.

METHOD DESCRIPTION

The nonlinear least squares parameter estimation algorithm provides a method by which parameters of any model can be estimated from a data set consisting of dependent and independent variables (see Draper and Smith 1966 for detailed discussion). The University of Washington Computer Center version of biomedical program MBDX85 was modified to facilitate computer approximation of the partial derivatives required by the following algorithm. The nonlinear regression method can be illustrated by the following example:

Given a model

$$Y = \alpha \sin(\beta x) + \gamma z \quad (1)$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are parameters to be estimated;  $x$  and  $z$  are independent variables,  $Y$  is the dependent variable. Then the statistical notation, equation (1), becomes

$$F(\bar{X}_i; \bar{P}) = p_1 \sin(p_2 x_i) + p_3 z_i$$

where  $\bar{P} = [p_1, \dots, p_m]^T$   $m = 3$

and  $\bar{X}$  is the set of parameters to be estimated from the set of  $n$  observations of the variables  $y$ ,  $x$ , and  $z$  to form the matrix

$$\bar{X} = \begin{bmatrix} y_1, x_1, z_1 \\ y_2, x_2, z_2 \\ \vdots \\ y_n, x_n, z_n \end{bmatrix}$$

and  $\bar{X}_i$  is the  $i$ th row of  $\bar{X}$ . The algorithm seeks  $\bar{P} = [p_1, \dots, p_m]^T$

---

NOTE:  $[p_1, \dots, p_m]^T = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix}$

such that the sum of squared residuals

$$R(\bar{X}; \bar{P}) = \sum_{i=1}^n (\text{observed}_i - \text{predicted}_i)^2 = \sum_{i=1}^n (y_i - F[\bar{X}_i; \bar{P}])^2$$

is minimum. The partial derivative of  $R(\bar{P}; \bar{X})$  with respect to each parameter is evaluated as

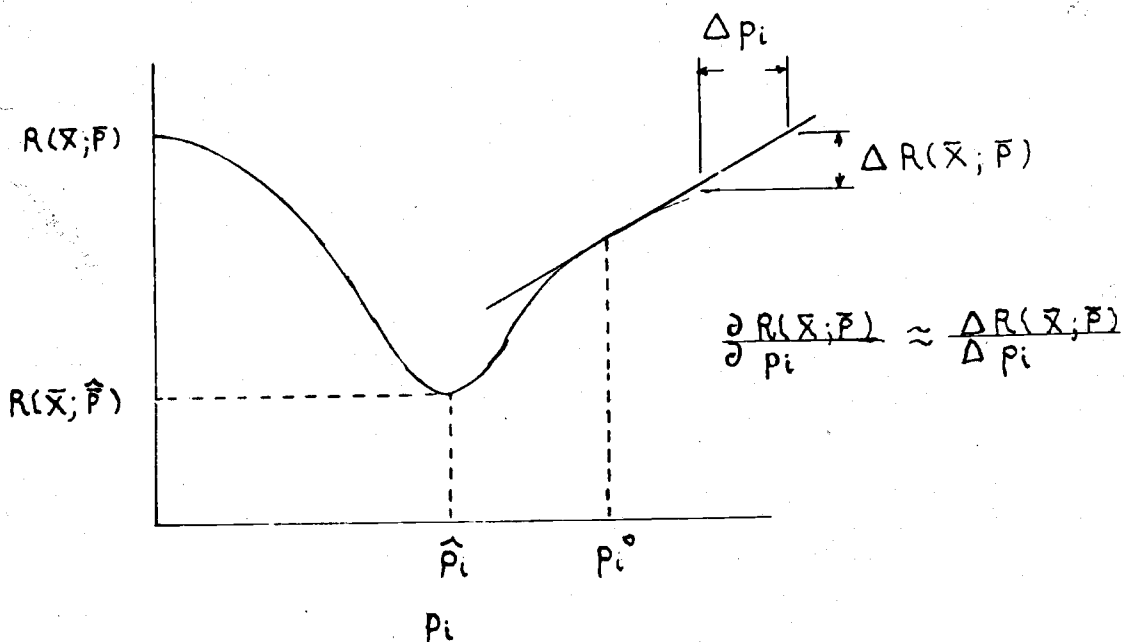
$$\frac{\partial R(\bar{X}; \bar{P})}{\partial p_i} = -2 \sum_{i=1}^n (y_i - F[\bar{X}_i; \bar{P}]) \frac{\partial F(\bar{X}_i; \bar{P})}{\partial p_i}$$

and must be zero at the minimum of  $R$ , or  $R(\bar{X}; \hat{\bar{P}})$ .

The algorithm uses the steepest descent approach starting with an initial guess  $\bar{P}^0 = [p_1^0, p_2^0, p_3^0]^T$  and calculates

$$\frac{\partial R(\bar{X}; \bar{P}^0)}{\partial p_i} = -2 \sum_{i=1}^n (y_i - F[\bar{X}_i; \bar{P}^0]) \frac{\partial F(\bar{X}_i; \bar{P}^0)}{\partial p_i}$$

giving the tangential slope of the residual graph with respect to each parameter  $p_i$ .



The sign of  $\partial R(\bar{X}; \bar{P}^0) / \partial p_i$  determines whether to decrease or increase the initial guess. The algorithm uses an  $m$ -dimensional Taylor series linearization to find the next approximation  $p_i^1$  of the best estimate for each parameter  $p_i$ . Then each  $p_i^1$  is substituted for each initial guess  $p_i^0$ . The procedure is repeated until no further reduction in the residual sum of squares occurs along the path of steepest descent. The algorithm corrects for overshooting the best estimate  $\bar{P}$  by checking the new residual function

$$R(\bar{P}^1; \bar{X}) \quad \bar{P}^1 = [p_1^1, p_2^1, \dots, p_m^1]^T$$

to determine if a reduction in the residual sum of squares has occurred. If  $R(\bar{P}^1; \bar{X})$  is not less than  $R(\bar{P}^0; \bar{X})$ , then the value for the parameters  $\bar{P}^1$  is changed to

$$\bar{P}^1 = \bar{P}^0 + (\bar{P}^0 + \bar{P}^1)/2$$

This halving is repeated ten times before the program assumes that no further reduction in the residual sum of squares occurs along the path of steepest descent.

Systems of nonlinear equations can be difficult to differentiate in obtaining the partials with respect to the parameters, that is, in finding  $\partial F(\bar{X}_i; \bar{P}^0) / \partial p_i$ . Of course these partials are easy to approximate:

$$\partial F(\bar{X}_i; \bar{P}^0) / \partial p_i = [F(\bar{X}_i; \bar{P}^0 + d_j) - F(\bar{X}_i; \bar{P}^0 - d_j)] / 2d_j$$

where  $\bar{P}^0 + d_j = [p_1^0, \dots, p_j^0 + d_j, p_{j+1}^0, \dots, p_m^0]^T$  and setting  $d_j$  small compared with  $p_j^0$  and  $p_j$  (Draper and Smith 1966).

#### PROGRAM DESCRIPTION AND EXAMPLE

The program is only a modification of program BMDX85, Nonlinear Least Squares. The program obtains a weighted least-squares fit of a user-specified function  $F$  to data values of the dependent and independent variables. The program iterates as described under Method Description. The program avoids singularity problems by iterating with a subset of the parameters when singularities occur. The program iteratively finds new values of the parameters that reduce the error mean square of the fit. The program halves the distance between the new parameter vector and the old parameter vector if the new parameter vector yields a higher error mean square than the old. In addition, the user places upper and lower bounds on each parameter. The iteration is performed on the boundary when the best fit appears to be outside the specified parameter range.

Output from the program includes:

1. Parameter values and residual mean square after each iteration
2. Asymptotic standard deviations and correlations for the estimated parameters after the last iteration
3. For each case after the last iteration:
  - a. Original data
  - b. Value of the function
  - c. Residual
  - d. Standard deviation of the prediction
4. Analysis of variance table
5. Plots of the estimated function with respect to the independent variables (optional)
6. Plots of the residuals with respect to the dependent, independent, and weighting (if weights are used) variables (optional)

The limitations of the program are: (1) total number of observations  $n(t + 1)$  must be less than or equal to 15,000, where  $t$  is the number of variables and  $n$  is the number of data points, (2)  $n$  must be less than or equal to 1000 for residual plots analysis, and (3) the number of parameters must be less than or equal to 100.

## Deck Structure

Computer input must be arranged according to the following example:

```
JØBCARD,T100,CM75000,P5
ACCØUNT(12345678,PASSWRD)
ATTACH,NLIN,ID=HAMERLY.
FØRTRAN.
CØPYBF(NLIN,LGØ)
LGØ
(end of record card)
      SUBROUTINE FUN ( F,D,P,X)
C
C   PURPOSE - FUN EVALUATES THE FUNCTION F AND THE PARTIALS
C   IN ARRAY D.  THIS EXAMPLE DOES NOT CALCULATE THE
C   PARTIALS SINCE THE OPTION TO ALLOW THE PROGRAM TO
C   APPROXIMATE THE PARTIALS IS GIVEN IN THE PROBLM CARD.
C   YOU MAY WANT TO SUPPLY A SET OF PARTIALS.
C   VOCABULARY
C   F - VALUE OF THE FUNCTION RETURNED
C   D(1) - PARTIAL DERIVATIVE OF F WITH RESPECT TO PARAMETER
C         P(1)
C   P(1) - 1TH PARAMETER
C   X(1) - 1TH VARIABLE READ BY THE FORMAT CARD
C
      DIMENSION D(1),P(1),X(1)
      E = P(1)*SIN (P(2)*X(2)) +P(3)*X(3)
C   AND OPTION
      (AND OPTIONALLY IF THE PARTIALS ARE NOT ESTIMATED)
      D(1) = SIN (P(2)*X(2))
      D(2) = P(1)*CØ S(P(2)*X(2)*X(2))
      D(3) = X(3)
      RETURN
      END
(end of record card)
(PROBLM CARD)
(format card)
(data cards)
(MAXMUM CARD)
(MINMUM CARD)
(PARAM CARD)
FINISH
(end of job card)
```

Problem Card

Columns	<u>Description</u>
1 - 6	PROBLM
7 - 12	Alphanumeric identification
13 - 16	Total number of variables read (not observations)
17 - 20	Index of the dependent variable
21 - 26	Number of observations, i.e., the number of rows in X matrix
27 - 30	Number of parameters
31 - 36	Tolerance T (Dixon 1970, p. 182). If blank, assumed to be 0.00001. This is used in testing for singularity.
37 - 42	Convergence criterion C (Dixon 1970, p. 183). If blank, C is assumed to be 0.00001. Each iteration must reduce the mean square error by at least this fraction or the process will be assumed to be converged.
43 - 46	Maximum number of iterations. This is assumed to be 100 if left blank. Enter a negative number to skip parameter estimation if only a residual analysis is desired.
47 - 50	Number of variable format cards (must be fewer than 9).
51 - 54	Alternate tape number for data input. Blank assumes data are on cards. For tape input, place 2 in this field and name the tape ALTTAPE.
55 - 57	YES if the alternate input tape is to be rewound.
58 - 61	Index of the weight variable. If blank, equal weights of unity are assumed.
62 - 63	$I$ , an integer where the partials of the function with respect to the parameters are evaluated: $\partial F / \partial p_j = [F(p_j + d) - F(p_j - d)] / 2d, \text{ where } d = 0.5 \times 10^{-I}$ Leave blank if the partials are calculated in subroutine FUN.
64 - 65	NO if plots of the function and residuals with respect to the variables are not desired. TE if plots of the function and residuals are to be made a convenient size for teletype or xeroxing. BLANK for full page line printer plots.

### Data cards

The input data are arranged by case. Each case contains the independent variables, dependent variables, and the weight variable in the positions indicated on the PROBLM card.

### MAXMUM card

This card sets the upper limit of each parameter.

Columns	<u>Description</u>
1 - 6	MAXMUM
7 - 12	Upper limit for the first parameter
13 - 18	Upper limit for the second parameter
⋮	
67 - 72	Upper limit for the eleventh parameter Additional cards must be used for additional parameters using the same format on each card and giving 11 values per card until all the limits are specified.

### MINMUM card

This card specifies the lower limit of each parameter

Columns	<u>Description</u>
1 - 6	MINMUM
7 - 12	Lower limit for the first parameter
13 - 18	Lower limit for the second parameter
⋮	
67 - 72	Lower limit for the eleventh parameter. Add additional cards for additional parameters as described under the MAXMUM card.

### PARAM card

This card specifies the initial estimates of each parameter.

Columns	<u>Description</u>
1 - 6	PARAM
7 - 12	Initial estimate for the first parameter

- 13 - 18 Initial estimate for the second parameter  
:  
:  
67 - 72 Initial estimate for the eleventh parameter  
Add additional cards for additional parameters as  
described under the MAXMUM card.

#### REFERENCES

- DIXON, W. J. (ed.) 1970. BMD:Biomedical computer programs-- X-series supplement. Univ. Calif. Publ. Autom. Comput. no. 3. Univ. California Press, Berkeley. 260 p.
- DRAPER, N. R., and H. SMITH. 1966. Applied regression analysis. Wiley, New York. 407 p.



SAMPLE OUTPUT

BMDX85-NONLINEAR LEAST SQUARES  
USING A STEPWISE GAUSS-NEWTON METHOD  
C.S.U. STAT LAB REVISED 20 JUNE 1970 FROM  
BMDX85 - NONLINEAR LEAST SQUARES - REVISED OCT. 72 BY EODIE

PROBLM CODE

NUMBER OF VARIABLES	3
INDEX OF THE DEPENDENT VARIABLE	1
INDEX OF THE WEIGHTING VARIABLE	-0
NUMBER OF CASES	17
NUMBER OF PARAMETERS	3
TOLERANCE	.000010
EPSILON	.000010
MAXIMUM NUMBER OF ITERATIONS	100
NUMBER OF VARIABLE FORMAT CARDS	1
ALTERNATE INPUT TAPE NUMBER	5
REWIND OPTION	NO

PARTIALS ARE ESTIMATED

$DF/DP(I) = (F(P(I)+DELP) - F(P(I)-DELP)) / (2.*DELP)$

DELP = 5.00000000E-04

GRAPHS OF THE FUNCTION AND RESIDUALS: YES.

INPUT FORMAT (3F10.0)

MINIMA -1.0000E+01 -1.0000E+01 -1.0000E+01

MAXIMA 1.0000E+01 1.0000E+01 1.0000E+01

ITERATION ERROR  
MEAN

NO.	SHF	SQUARE	PARAMETERS		
0	0	5.9501E+01	5.0000E+00	1.3000E+00	5.0000E+00
1	0	1.5375E-01	8.4104E-01	1.2818E+00	1.0825E+00
2	0	6.9466E-02	8.4353E-01	1.1359E+00	1.0634E+00
3	0	1.8566E-02	1.0459E+00	9.9939E-01	1.0156E+00
4	0	7.6428E-03	1.0713E+00	1.0365E+00	1.0206E+00
5	0	7.6354E-03	1.0740E+00	1.0362E+00	1.0213E+00
6	0	7.6354E-03	1.0740E+00	1.0362E+00	1.0213E+00
7	0	7.6354E-03	1.0740E+00	1.0362E+00	1.0213E+00
8	0	7.6354E-03	1.0740E+00	1.0362E+00	1.0213E+00
9	7	7.6354E-03	1.0740E+00	1.0362E+00	1.0213E+00
10	6	7.6354E-03	1.0740E+00	1.0362E+00	1.0213E+00

ASYMPTOTIC STANDARD DEVIATIONS OF THE PARAMETERS

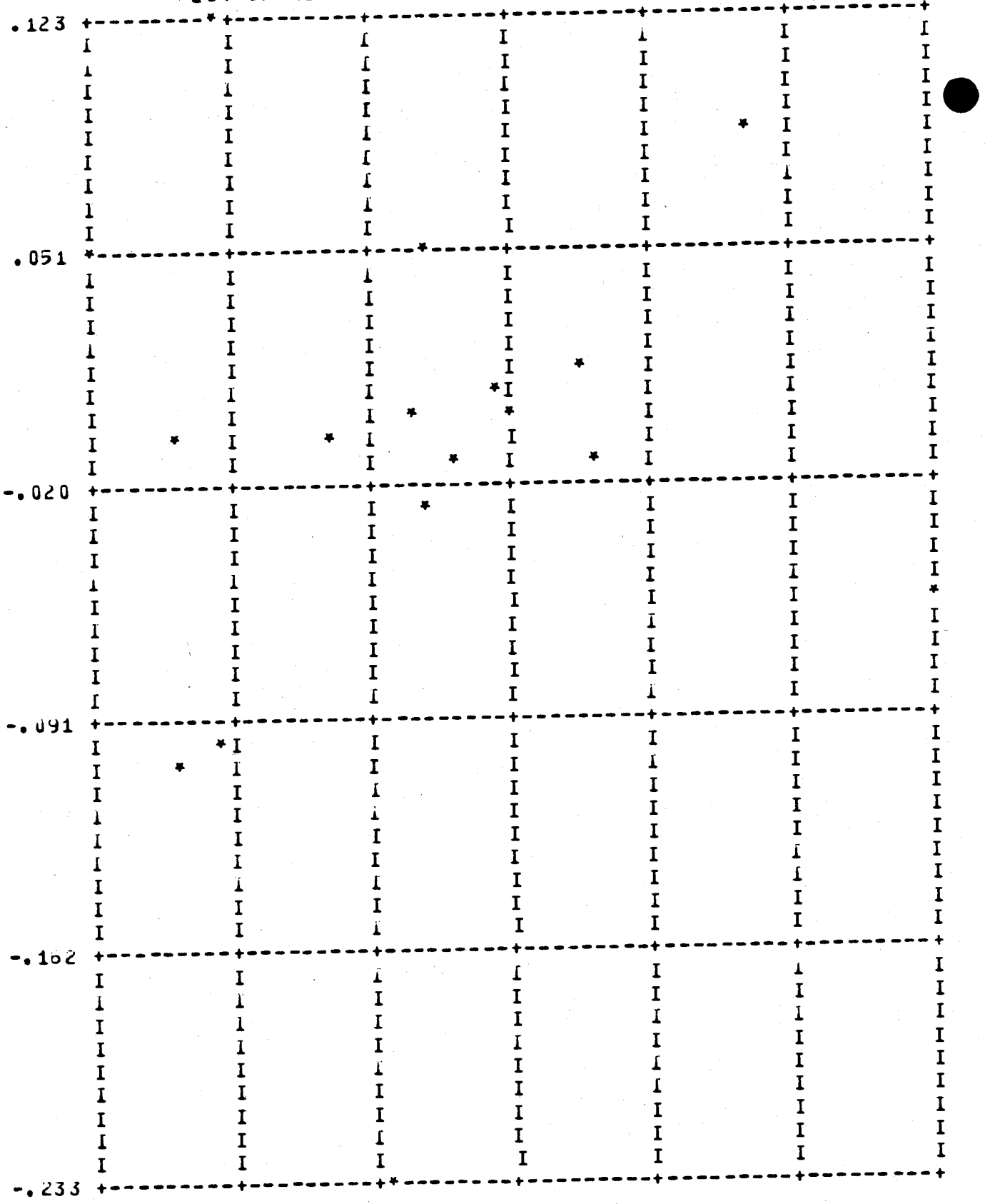
2.9901E-02 8.8245E-03 1.2344E-02

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

	1	2	3
1	1.00000	-.14980	-.05049
2	-.14980	1.00000	.32442
3	-.05049	.32442	1.00000



PLOT OF RESIDUALS WITH RESPECT TO VARIABLE 1



RESIDUALS

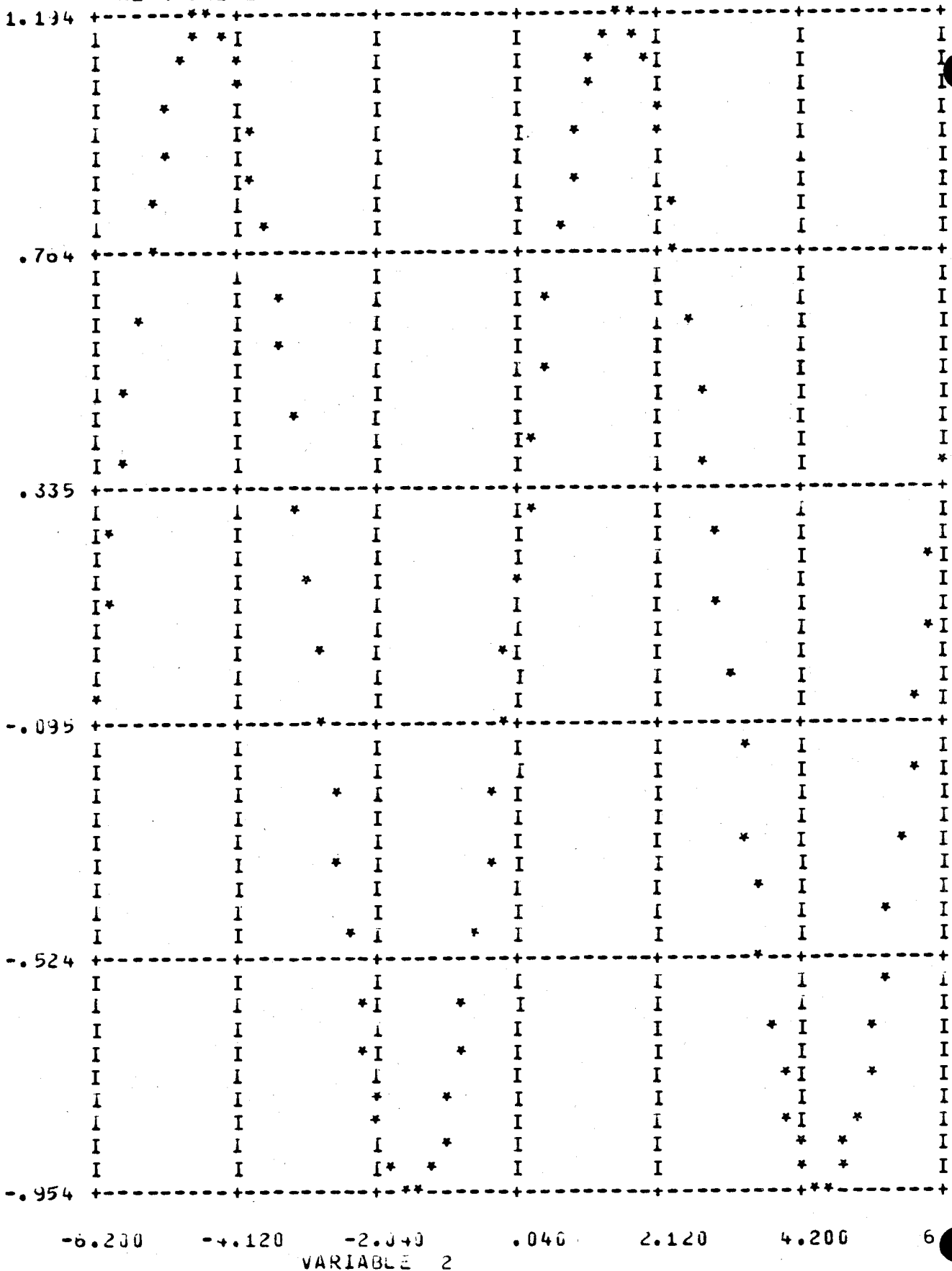
-3.050    -1.720    -.390    .940    2.270    3.600    4.930  
 VARIABLE 1

THE PARAMETERS BEING USED ARE  
 1.0740    1.0362    1.0213  
 MEAN VECTOR  
 1 .111853    .360000    .117647



PLOT OF THE FUNCTION WITH RESPECT TO VARIABLE 2  
 MEAN VALUES SUBSTITUED FOR THE OTHER VARIABLES

FUNCTION



THE PARAMETERS BEING USED ARE

1.0740

1.0362

1.0213

MEAN VECTOR

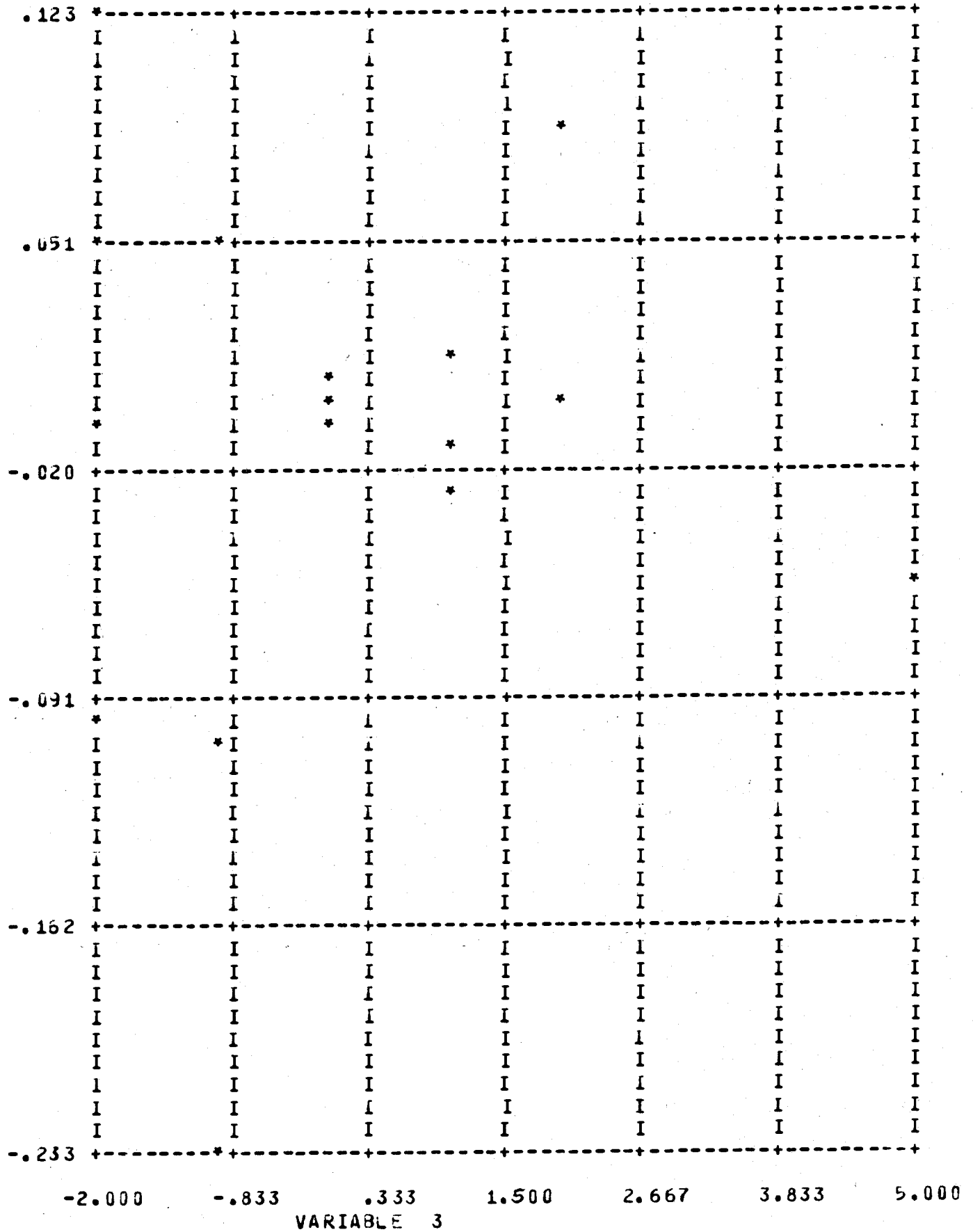
1 .111853

.360000

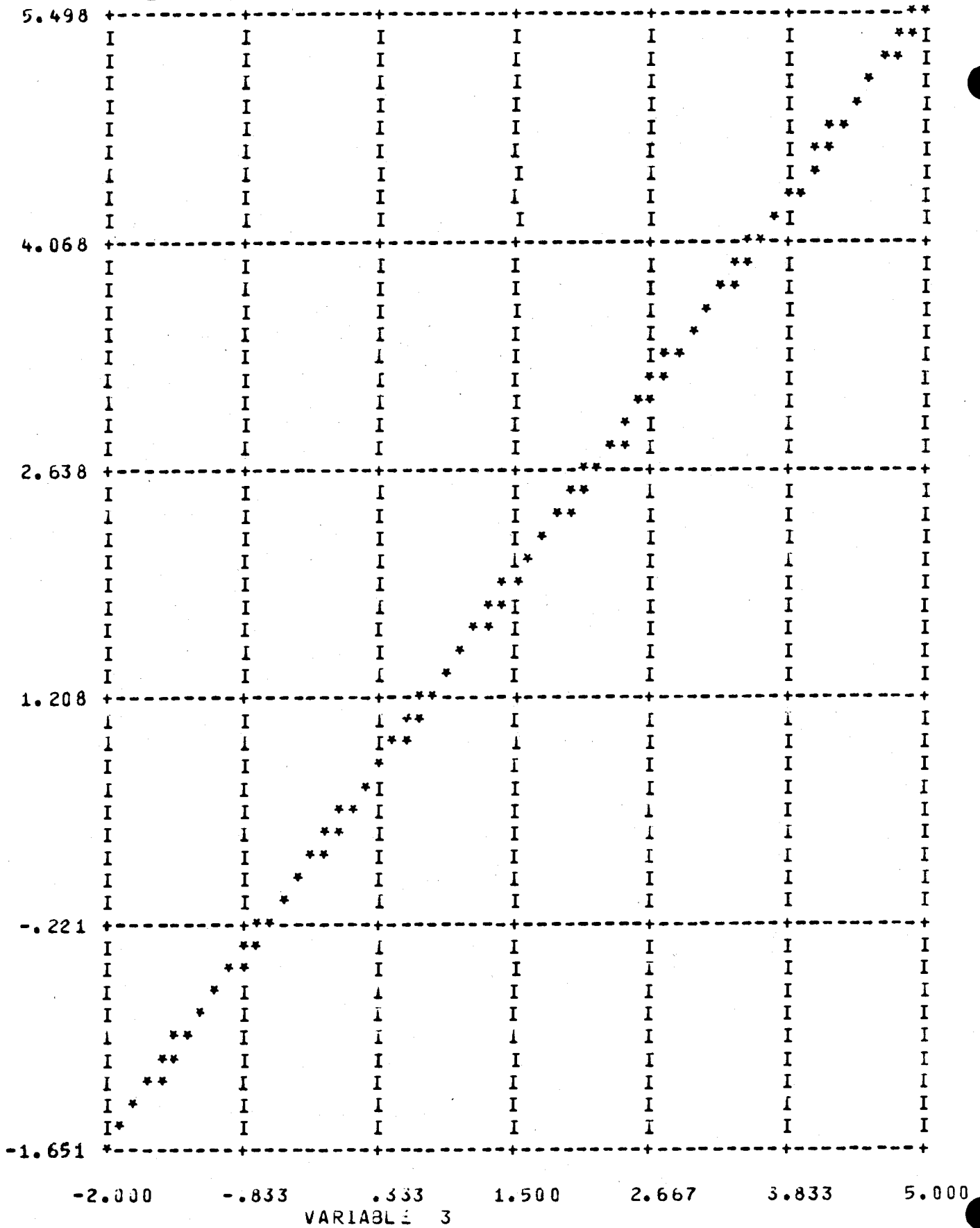
.117647

PLOT OF RESIDUALS WITH RESPECT TO VARIABLE 3

RESIDUALS



PLOT OF THE FUNCTION WITH RESPECT TO VARIABLE 3  
 MEAN VALUES SUBSTITUED FOR THE OTHER VARIABLES



FUNCTION