Computational Statistics

Editors:
Herbert Büning · Peter Naeve

Offprint

Walter de Gruyter · Berlin · New York 1981

Inhalt / Contents

continued on inside back cover

Statistical Computing with a Text Editor

Seppo Mustonen

1. INTRODUCTION

It is quite common that when writing a research report containing numerical tables the output from the computer cannot be used as such, but the results have to be retyped manually. This may happen even if the computer output is well designed, since the needs of the user may change during the reporting phase.

One solution to this problem is to have the possibility of saving results in output files which can subsequently be treated as text files using a text editing program. Parallel to this approach we have developed another solution to these editorial problems within the interactive statistical system SURVO 76 operating on the desk computer Wang 2200VP, (Mustonen 1980a). Our solution has been realized as a new extensive SURVO 76 module called EDITOR and it is closely connected to other parts of the system.

SURVO 76 EDITOR works as a normal text editor, but it also includes several numerical and statistical operations thus forming a small scale statistical operating system. The main purpose of EDITOR is to lessen the burden of a statistician in data management and report writing. EDITOR can be used for various tasks encountered in statistical data processing like

1) input and editing unformatted data,
2) saving data in SURVO 76 files,
3) editing SURVO 76 files and results,
4) manipulating lists and multiway tables,
5) arithmetical and statistical computations,
6) data analysis.

Working in editing mode differs considerably from conversational interactive data processing, as used for instance in the majority of SURVO 76 modules. Since both the data and the operations or commands are displayed and handled together within the frame of an edit field, the user has an extremely close contact with his work and he has the possibility to control each detail of the computational and editorial activities better than in other forms of interactive analysis. An editorial operating system demands more experience from the user than a simple conversational system, but at the same time gives more scope for imagination. Editorial mode is no substitute for conversational mode, in general. There are, however, several forms of interactive data analysis and management where editorial mode is more natural than conversational use. For example, problems in analysis of variance and log-linear models for contingency tables can be easily handled in EDITOR (Mustonen 1980b).

It may be reasonable to have different interactive working principles available in the same interactive statistical system and let the user select the devices most suitable for a particular problem. In the SURVO 76 system we try to provide many of the statistical operations both in conversational and editorial mode.

In this paper we shall describe principles and applications of the editorial approach. The main emphasis will be on statistical and numerical operations. Pure text processing activities will not be discussed in this context. A comprehensive account of SURVO 76 EDITOR is given in Mustonen (1980b).

When working with EDITOR the whole editing process is controlled by the normal keyboard keys and 32 programmable 'soft keys' (F-keys) which are used for simple text editing. For more complicated tasks various editing operations are available.

All the information is represented in an edit field which consists, for example, of 100 columns and 250 rows. The field is always partially visible on the CRT which is like a window to the field. The user can easily scroll the text on the CRT to any direction by pressing certain F-keys with arrows indicating the direction. The editing operations are also typed in the edit field and they can be treated as normal text. Any operation is activated by moving the cursor to the corresponding line and by pressing key CONTINUE. Whenever needed the contents of the edit field (tables, text and operations) can be saved in an edit file by activating a SAVE operation.

The edit field is like a notebook for the user, but it is much more flexible, since text and data in that notebook can be worked upon by editing operations and the results of these operations can be directed to any part of the field. Since the editing operations themselves are typed among the text and data, the user can place them as he likes, usually near the object of operation.

If the user likes he can put the operations on adjacent lines and carry them out step by step as an editing program, but usually this is unnecessary. On the contrary, it is typical that during the editing process the field is filled with a mixture of text, data and operations, and the user scratches unessential ingredients when needed.

When starting a new job with EDITOR the upper left side corner of a 100x100 edit field is displayed on the 24x80 screen:

Disp.1

```
    1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 1  *_
 2  *
 3  *
 4  *
 5  *
 6  *
 7  *
 8  *
 9  *
10  *
11  *
12  *
13  *
14  *
15  *
16  *
17  *
18  *
19  *
20  *
21  *
22  *
23  *
```

The size of the field can be altered by using a redimensioning operation REDIM m,n. Observe that there is no strict correspondence between the size of the edit field and the size of a page in the final report. It is typical to have rather large edit fields and let EDITOR automatically split the fields into pages when the document is printed with a PRINT operation.

In display 1 representing an empty edit field, the cursor is blinking in the first position and the user may start typing text, data values and editing operations. The system can in many respects be operated like a normal typewriter and the text appears continuously on the screen.

Typical text editing tasks are accomplished simply by F-keys. We thus have special keys for insertion and deletion of characters and lines, for moving the cursor, etc.
The leading principle in all editorial activities is that minimum effort (i.e. minimum number of touches with the keyboard) is needed for them.

2. EXAMPLE ON DATA ANALYSIS

Assume now that we have typed the following small data set in the edit field:

Disp.2

1: SURVO 76 EDITOR (C)1979 S.Mustonen (100x100)

Yearly consumption of certain beverages per inhabitant
in some European countries in 1974-1978:

	Coffee (kg)	Tea (kg)	Beer (l)	Wine (l)	Spirits (l)
Denmark	11.4	0.43	116.3	11.5	1.7
England	1.5	3.42	118.3	5.7	1.6
Finland	12.0	0.17	55.2	8.7	2.9
France	5.3	0.11	45.8	101.8	2.5
Germany	5.8	0.17	148.0	22.9	3.0
Holland	8.7	0.54	81.5	11.2	2.9
Ireland	0.2	3.78	127.9	4.5	2.1
Italy	3.5	0.06	14.2	99.1	2.0
Norway	9.0	0.19	45.1	3.2	1.8
Spain	2.5	0.03	47.6	71.8	2.8
Sweden	11.7	0.34	56.0	8.6	3.0
Switzerland	8.7	0.25	70.9	44.8	2.0

If we like to sort this data set in ascending order with respect to 'Wine consumption', this will be carried out by a SORT operation which may be typed on any editor line (here on line 19) in the form

Disp.3

```
   1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
1  *
2  *          Yearly consumption of certain beverages per inhabitant
3  *          in some European countries in 1974-1978:
4  *             Coffee  Tea    Beer  Wine  Spirits
5  *               (kg)  (kg)   (l)   (l)   (l)
6  * Denmark      11.4   0.43  116.3  11.5   1.7
7  * England       1.5   3.42  118.3   5.7   1.6
8  * Finland      12.0   0.17   55.2   8.7   2.9
9  * France        5.3   0.11   45.8 101.8   2.5
10 * Germany       5.8   0.17  148.0  22.9   3.0
11 * Holland       8.7   0.54   81.5  11.2   2.9
12 * Ireland       0.2   3.78  127.9   4.5   2.1
13 * Italy         3.5   0.06   14.2  99.1   2.0
14 * Norway        9.0   0.19   45.1   3.2   1.8
15 * Spain         2.5   0.03   47.6  71.8   2.8
16 * Sweden       11.7   0.34   56.0   8.6   3.0
17 * Switzerland   8.7   0.25   70.9  44.8   2.0
18 *                                 11111
19 *SORT 6,17,18_
20 *
```

SORT has three parameters. The two first parameters 6,17 refer to the lines which are to be sorted and the last parameter 18 is also a line number indicating the line with the sort key which is given as a mask '11111' in the column corresponding to 'Wine'.

Now in order to start sorting we have to activate the SORT operation by pressing the key CONTINUE when the cursor is on line 19. When CONTINUE is pressed EDITOR always tries to interpret the contents of the current editor line as an editing operation. If the line corresponds to a valid operation, it is immediately executed. Otherwise an error message is displayed. In this case activation with CONTINUE leads to the following display

Disp.4

```
   1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
1  *
2  *          Yearly consumption of certain beverages per inhabitant
3  *          in some European countries in 1974-1978:
4  *             Coffee  Tea    Beer  Wine  Spirits
5  *               (kg)  (kg)   (l)   (l)   (l)
6  * Norway        9.0   0.19   45.1   3.2   1.8
7  * Ireland       0.2   3.78  127.9   4.5   2.1
8  * England       1.5   3.42  118.3   5.7   1.6
9  * Sweden       11.7   0.34   56.0   8.6   3.0
10 * Finland      12.0   0.17   55.2   8.7   2.9
11 * Holland       8.7   0.54   81.5  11.2   2.9
12 * Denmark      11.4   0.43  116.3  11.5   1.7
13 * Germany       5.8   0.17  148.0  22.9   3.0
14 * Switzerland   8.7   0.25   70.9  44.8   2.0
15 * Spain         2.5   0.03   47.6  71.8   2.8
16 * Italy         3.5   0.06   14.2  99.1   2.0
17 * France        5.3   0.11   45.8 101.8   2.5
18 *                                 11111
19 *SORT 6,17,18_
20 *
```

The lines are now in an ascending order according to 'Wine'. To display them in descending order we simply change the SORT operation to a -SORT operation by inserting a "-" sign

```
19    *-SORT 6,17,18
```

and reactivation of line 19 by CONTINUE leads now to

Disp.5

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 1    *
 2    *          Yearly consumption of certain beverages per inhabitant
 3    *          in some European countries in 1974-1978:
 4    *             Coffee  Tea    Beer  Wine  Spirits
 5    *              (kg)   (kg)   (l)   (l)   (l)
 6    * France       5.3    0.11   45.8  101.8  2.5
 7    * Italy        3.5    0.06   14.2  99.1   2.0
 8    * Spain        2.5    0.03   47.6  71.8   2.8
 9    * Switzerland  8.7    0.25   70.9  44.8   2.0
10    * Germany      5.8    0.17  148.0  22.9   3.0
11    * Denmark     11.4    0.43  116.3  11.5   1.7
12    * Holland      8.7    0.54   81.5  11.2   2.9
13    * Finland     12.0    0.17   55.2   8.7   2.9
14    * Sweden      11.7    0.34   56.0   8.6   3.0
15    * England      1.5    3.42  118.3   5.7   1.6
16    * Ireland      0.2    3.78  127.9   4.5   2.1
17    * Norway       9.0    0.19   45.1   3.2   1.8
18    *                                  11111
19    *-SORT 6,17,18
20    *
```

As an example of how typical data analytic problems are solved in EDITOR we try to estimate a linear regression model where 'Wine' is the dependent variable. To have a constant term in the model we first create a new column of plain 1's and this takes place by a SET operation of the form

Disp.6

```
15    * England      1.5    3.42  118.3   5.7   1.6
16    * Ireland      0.2    3.78  127.9   4.5   2.1
17    * Norway       9.0    0.19   45.1   3.2   1.8
18    *                                                  1
19    *SET 6,17,18_
20    *
```

saying that the non-blank characters on line 18 ought to be inserted on lines 6-17. After the activation of this operation we have

Disp.7

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 1    *
 2    *          Yearly consumption of certain beverages per inhabitant
 3    *          in some European countries in 1974-1978:
 4    *             Coffee  Tea    Beer  Wine  Spirits
 5    *               (kg)  (kg)   (l)   (l)   (l)
 6    * France        5.3   0.11   45.8 101.8  2.5       1
 7    * Italy         3.5   0.06   14.2  99.1  2.0       1
 8    * Spain         2.5   0.03   47.6  71.8  2.8       1
 9    * Switzerland   8.7   0.25   70.9  44.8  2.0       1
10    * Germany       5.8   0.17  148.0  22.9  3.0       1
11    * Denmark      11.4   0.43  116.3  11.5  1.7       1
12    * Holland       8.7   0.54   81.5  11.2  2.9       1
13    * Finland      12.0   0.17   55.2   8.7  2.9       1
14    * Sweden       11.7   0.34   56.0   8.6  3.0       1
15    * England       1.5   3.42  118.3   5.7  1.6       1
16    * Ireland       0.2   3.78  127.9   4.5  2.1       1
17    * Norway        9.0   0.19   45.1   3.2  1.8       1
18    *                                                  1
19    *SET 6,17,18_
```

In statistical operations our data set can now be referred to by a name COUNTRIES by inserting (here on line 19) a DATA specification

```
19    *DATA COUNTRIES,6,17,4
```

where 6-17 are the lines for the data values and 4 refers to the line telling the names of the variables (columns). Thus our data set ready for statistical analysis looks like

Disp.8

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 1    *
 2    *          Yearly consumption of certain beverages per inhabitant
 3    *          in some European countries in 1974-1978:
 4    *             Coffee  Tea    Beer  Wine  Spirits Constant
 5    *               (kg)  (kg)   (l)   (l)   (l)
 6    * France        5.3   0.11   45.8 101.8  2.5       1
 7    * Italy         3.5   0.06   14.2  99.1  2.0       1
 8    * Spain         2.5   0.03   47.6  71.8  2.8       1
 9    * Switzerland   8.7   0.25   70.9  44.8  2.0       1
10    * Germany       5.8   0.17  148.0  22.9  3.0       1
11    * Denmark      11.4   0.43  116.3  11.5  1.7       1
12    * Holland       8.7   0.54   81.5  11.2  2.9       1
13    * Finland      12.0   0.17   55.2   8.7  2.9       1
14    * Sweden       11.7   0.34   56.0   8.6  3.0       1
15    * England       1.5   3.42  118.3   5.7  1.6       1
16    * Ireland       0.2   3.78  127.9   4.5  2.1       1
17    * Norway        9.0   0.19   45.1   3.2  1.8       1
18    *                                                  1
19    *DATA COUNTRIES,6,17,4_
```

Observe that we have also added the label 'Constant' above the last column.
To compute the parameters of a linear regression model a REGRAN operation is entered:

Disp.9

```
 15   * England        1.5   3.42  118.3   5.7   1.6        1
 16   * Ireland        0.2   3.78  127.9   4.5   2.1        1
 17   * Norway         9.0   0.19   45.1   3.2   1.8        1
 18   *              XXXX   XXXX  XXXXX YYYYY   XXX        X     -123.1
 19   *DATA COUNTRIES,6,17,4
 20   *REGRAN COUNTRIES,18,21_
 21   *
```

REGRAN has three parameters: the name of the data set (COUNTRIES), number of the line specifying the model (18) and number of the the first line to be used for the results (21). The model is specified (on line 18) by placing Y's to the columns of the dependent variable (Wine) and X's to the columns of the independent variables (Coffee, Tea, Beer, Spirits and Constant). Furthermore, as an option, it is possible to obtain the residuals as a new column by giving a numerical format as an image (-123.1). Activation of the REGRAN line 20 gives the results

Disp.10

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
  1   *
  2   *           Yearly consumption of certain beverages per inhabitant
  3   *           in some European countries in 1974-1978:
  4   *              Coffee  Tea    Beer   Wine  Spirits Constant
  5   *              (kg)    (kg)   (l)    (l)   (l)
  6   * France         5.3   0.11   45.8  101.8   2.5        1       35.4
  7   * Italy          3.5   0.06   14.2   99.1   2.0        1        5.7
  8   * Spain          2.5   0.03   47.6   71.8   2.8        1      -15.3
  9   * Switzerland    8.7   0.25   70.9   44.8   2.0        1       10.9
 10   * Germany        5.8   0.17  148.0   22.9   3.0        1       -7.3
 11   * Denmark       11.4   0.43  116.3   11.5   1.7        1       12.5
 12   * Holland        8.7   0.54   81.5   11.2   2.9        1       -6.8
 13   * Finland       12.0   0.17   55.2    8.7   2.9        1        0.5
 14   * Sweden        11.7   0.34   56.0    8.6   3.0        1        2.9
 15   * England        1.5   3.42  118.3    5.7   1.6        1       -1.7
 16   * Ireland        0.2   3.78  127.9    4.5   2.1        1        1.3
 17   * Norway         9.0   0.19   45.1    3.2   1.8        1      -38.0
 18   *              XXXX   XXXX  XXXXX YYYYY   XXX        X      -12.1
 19   *DATA COUNTRIES,6,17,4
 20   *REGRAN COUNTRIES,18,21_
 21   *REGRESSION ANALYSIS: REGRESSAND:'Wine'  DATA:'COUNTRIES'
 22   *   TOTAL VARIANCE=     1401.623 DF= 11
 23   *   RESIDUAL VARIANCE=   480.221 DF=  7  R↑2=0.7819
 24   * VARIABLE      REGR.COEFF.  STD.DEVIATION      T
 25   * Coffee         -7.689164       2.05534    -3.741
 26   * Tea           -22.694462       8.28630    -2.738
 27   * Beer           -0.268177       0.20304    -1.320
 28   * Spirits        -7.087441      13.62378    -0.520
 29   * Constant      139.644319      37.98259     3.676
```

The great advantage of the editorial approach is that we have all the ingredients (data, results and operations) on the same level and immediately available for new modifications and operations. For example, it is easy to alter the data values or the model. Here it seems to be reasonable to simplify the model by dropping the variable 'Spirits'. This is done by cancelling the X's for 'Spirits' on the image line 18 and by reactivating the REGRAN line 20.

Disp.11

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 15   * England        1.5   3.42  118.3   5.7   1.6        1     -1.7   0.3
 16   * Ireland        0.2   3.78  127.9   4.5   2.1        1      1.3  -0.4
 17   * Norway         9.0   0.19   45.1   3.2   1.8        1    -38.0 -33.8
 18   *              XXXX   XXXX  XXXXX YYYYY               X          -12.1
 19   *DATA COUNTRIES,6,17,4
 20   *REGRAN COUNTRIES,18,31_
 21   *REGRESSION ANALYSIS: REGRESSAND:'Wine'  DATA:'COUNTRIES'
 22   *   TOTAL VARIANCE=     1401.623 DF= 11
 23   *   RESIDUAL VARIANCE=   480.221 DF=  7  R↑2=0.7819
 24   * VARIABLE      REGR.COEFF.  STD.DEVIATION     T
 25   * Coffee          -7.689164       2.05534   -3.741
 26   * Tea            -22.694462       8.28630   -2.738
 27   * Beer            -0.268177       0.20304   -1.320
 28   * Spirits         -7.087441      13.62378   -0.520
 29   * Constant       139.644319      37.98259    3.676
 30   *
 31   *REGRESSION ANALYSIS: REGRESSAND:'Wine'  DATA:'COUNTRIES'
 32   *   TOTAL VARIANCE=     1401.623 DF= 11
 33   *   RESIDUAL VARIANCE=   436.439 DF=  8  R↑2=0.7735
 34   * VARIABLE      RGR.COEFF.   STD.DEVIATION     T
 35   * Coffee          -7.636371       1.95702   -3.902
 36   * Tea            -21.002946       7.26597   -2.890
 37   * Beer            -0.289111       0.18972   -1.523
 38   * Constant       122.855589      19.09494    6.433
 39   *
```

Observe that in order to preserve the previous results we have altered the first line of the results from 21 to 31 and moved the image of the residuals some steps to the right.

Before taking a printout of the results obtained we may edit them as we like, for instance, by removing technical lines and inserting new text as labels and comments.

To have a printout on paper the simplest way is to type PRINT 2,39 and activate this operation by pressing CONTINUE and we get a list of lines 2-39 (without line numbers and other control information) on paper. The edit field can also be saved on disk for subsequent use by typing a SAVE <file name> operation on any editor line and by activating this line. Similarly any edit file may be called to the edit field by a LOAD <file name> operation. EDITOR also provides means for combining edit fields.

3. NUMERICAL COMPUTATIONS IN THE EDIT FIELD

When working with statistical data in the edit field it is natural to expect simple arithmetic operations to be readily available. As a matter of fact, standard calculations enriched with various mathematical, statistical and user-defined functions can easily be carried out in the edit field.

The arithmetic expressions are written in the edit field in normal fashion and they are evaluated by typing a "=" sign after the expression itself and by pressing the activation key CONTINUE. The value obtained will then be displayed after the "=" sign.

For instance, to compute the mean of the numbers 12, 17 and 25, we type

Disp.12

```
      1  SURVO 76 EDITOR   (C)1979 S.Mustonen   (100x100)
 1    *
 2    *
 3    *          (12+17+25)/3=_
 4    *
```

and activate by pressing CONTINUE. This leads to the display

Disp.13

```
      1  SURVO 76 EDITOR   (C)1979 S.Mustonen   (100x100)
 1    *
 2    *
 3    *          (12+17+25)/3=18
 4    *
```

In order to avoid repeating various constants and variables appearing in the expressions it is possible to use symbolic notation. For instance, the preceding example could be written as follows:

Disp.14

```
      1  SURVO 76 EDITOR   (C)1979 S.Mustonen   (100x100)
 1    *
 2    *
 3    *          X=12, Y=17, Z=25
 4    * arithmetic mean   (X+Y+Z)/3=_
```

If the expression on line 4 is activated, EDITOR is capable of finding the definitions of X,Y and Z and inserting their values in the expression.

The definitions of variables and constants in the edit field can be nested, i.e. a variable may be defined as a function of other variables. This property enables the construction of general arithmetic computation schemes.
For instance, to compute values of the standard normal distribution function according to a well-known polynomial approximation, the following scheme will be sufficient:

Disp.15

```
      1  SURVO 76 EDITOR   (C)1979 S.Mustonen   (100x100)
12    *
13    *  For X>0, the standard normal distribution function F(X)
14    *  is approximated by
15    *  F=1-f*(b1*t+b2*t↑2+b3*t↑3+b4*t↑4+b5*t↑5) ,
16    *  where
17    *  f=(1/sqr(2*pi))*exp(-X↑2/2) (density function) and
18    *  t=1/(1+r*X) , r=0.2316419
19    *  b1=.31938153 ,     b2=-.356563782
20    *  b3=1.781477937 ,   b4=-1.821255978
21    *  b5=1.330274429 ,
22    *  pi=3.14159265359 .
23    *
24    *  To compute F(X) enter X on line 26 and activate F
25    *
26    *  X=_   F=
27    *
```

The scheme needs no further explanations. The reader immediately sees the essential details and EDITOR can interpret all the components provided that everything is unambiguously defined. Thus, inserting X=3 on line 26 and activating F on the same line gives the result

Disp.16

```
 24   *   To compute F(X) enter X on line 26 and activate F
 25   *
 26   *   X=3   F=0.9986500327
 27   *
```

Observe that this computation scheme included nested definitions as F depends on t which depends on r, etc. Some elementary functions like square root and exponential function were used, too.

In addition to the standard functions several statistical functions are readily available. For example, the distribution function, the density function and the inverse distribution function of the normal distribution N(u,s↑2) are N.F(u,s↑2,x), N.f(u,s↑2,x) and N.G(u,s↑2,y), respectively. Hence the preceding scheme could be replaced simply by the function call N.F(0,1,X)=_. Since the arguments may be functions, we may generate random normal deviates in the edit field as the values of the function N.G(0,1,rnd(1)).

The SURVO 76 EDITOR computation schemes can be considered simple programs involving arithmetic operations in a free form. Since the schemes and the results may easily be edited by normal editing operations, the results will be printed on paper precisely in the way the user likes. The possibility to combine computation schemes with other activities provided by EDITOR strongly increases the versatility of the system.

More advanced examples of the possibilities offered by the editorial approach are presented in the following schemes.

Disp.17

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 52   *
 53   *   The roots of the equation a*z↑2+b*z+c=0 are computed as follows:
 54   *
 55   *   Let the roots be z1=x1+i*y1, z2=x2+i*y2
 56   *   and the discriminant=b↑2-4*a*c.
 57   *
 58   *   if discriminant)=0 then
 59   *   x1=(-b+sqr(discriminant))/(2*a),  y1=0
 60   *   x2=(-b-sqr(discriminant))/(2*a),  y2=0   endif
 61   *
 62   *   if discriminant<0  then
 63   *   x1=-b/(2*a),   y1=sqr(-discriminant)/(2*a)
 64   *   x2=x1,         y2=-y1                    endif
 65   *
 66   *   Let a=1, b=4 and c=13.
 67   *   Then
 68   *       x1:=-2              y1:=3
 69   *       x2:=-2              y2:=-3
 70   *       discriminant:=-36
 71   *
```

The scheme in display 17 illustrates the possibility of using conditional definitions (if-then-endif structure) and multiple activations. If an expression tailed with := is activated, all the expressions having the same tail will be activated simultaneously.

Disp.18

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
  1   *
  2   *     Generating data according to a given regression model:
  3   *
  4   *     Y=a0+a1*X1+a2*X2+eps,
  5   *     where eps=N.G(0,s↑2,rnd(1)), s=3,
  6   *     a0=100, a1=10 and a2=1.
  7   *
  8   *COMP 11,22,10,9_
  9   *    XXXXXXXX        XXX         123.12
 10   *       X1            X2           Y
 11   * 1  -0.35449         68
 12   * 2  -0.49315        110
 13   * 3   0.48293         33
 14   * 4  -0.55747         86
 15   * 5   1.94759        136
 16   * 6   1.21945         30
 17   * 7  -0.07194         94
 18   * 8   1.54428        144
 19   * 9   0.07833        150
 20   *10  -1.06143         89
 21   *11  -0.84473        130
 22   *12  -0.52181         84
 23   *
```

In display 18 we are using a COMP operation which enables construction of tables of functions. When COMP on line 8 is activated the Y values will be computed for the X1 and X2 values on lines 11-22. The auxiliary lines 9 and 10 indicate that Y is the function to be computed. EDITOR finds the definition of Y on line 4 and further explanations (definition of eps and values of s,a0,a1 and a2) on lines 5 and 6. On the basis of this information the Y column is created and displayed according to the format on the line 9:

Disp.19

```
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
  8   *COMP 11,22,10,9_
  9   *    XXXXXXXX        XXX         123.12
 10   *       X1            X2           Y
 11   * 1  -0.35449         68         165.65
 12   * 2  -0.49315        110         195.23
 13   * 3   0.48293         33         137.54
 14   * 4  -0.55747         86         179.11
 15   * 5   1.94759        136         252.33
 16   * 6   1.21945         30         141.12
 17   * 7  -0.07194         94         200.35
 18   * 8   1.54428        144         263.96
 19   * 9   0.07833        150         252.53
 20   *10  -1.06143         89         172.77
 21   *11  -0.84473        130         218.60
 22   *12  -0.52181         84         177.53
 23   *
```

To "check" this result we can easily perform a REGRAN operation (after inserting a column C of 1's as a constant term):

```
Disp.20
      1  SURVO 76 EDITOR    (C)1979 S.Mustonen    (100x100)
 1    *
 2    *     Generating data according to a given regression model:
 3    *
 4    *     Y=a0+a1*X1+a2*X2+eps,
 5    *     where eps=N.G(0,s↑2,rnd(1)), s=3,
 6    *     a0=100, a1=10 and a2=1.
 7    *
 8    *COMP 11,22,10,9
 9    *    XXXXXXXX       XXX        123.12
10    *       X1           X2           Y        C
11    * 1  -0.35449         68         165.65     1        2.76
12    * 2  -0.49315        110         195.23     1       -8.09
13    * 3   0.48293         33         137.54     1        0.13
14    * 4  -0.55747         86         179.11     1        0.52
15    * 5   1.94759        136         252.33     1       -4.77
16    * 6   1.21945         30         141.12     1       -1.66
17    * 7  -0.07194         94         200.35     1        8.23
18    * 8   1.54428        144         263.96     1        3.44
19    * 9   0.07833        150         252.53     1        2.69
20    *10  -1.06143         89         172.77     1       -3.08
21    *11  -0.84473        130         218.60     1       -0.72
22    *12  -0.52181         84         177.53     1        0.54
23    *    XXXXXXXX       XXX        YYYYYY     X      -12.12
24    *DATA TEST,11,22,10
25    *REGRAN TEST,23,26_
26    *REGRESSION ANALYSIS: REGRESSAND:'Y'  DATA:'TEST'
27    *   TOTAL VARIANCE=     1821.113 DF= 11
28    *   RESIDUAL VARIANCE=    21.808 DF=  9  R↑2=0.9902
29    * VARIABLE      REGR.COEFF.  STD.DEVIATION     T
30    * X1              11.378313       1.44394     7.880
31    * X2               1.000256       0.03536    28.281
32    * C               98.905259       3.64851    27.108
33    *
```

REFERENCES

Mustonen S.(1980), Interactive analysis in SURVO 76, Proceedings of Computational Statistics (Ed. M.M.Barritt, D.Wishart), Physica-Werlag, Wien, 253-259.

Mustonen S.(1980), SURVO 76 Editor, a new tool for interactive statistical computing, text and data management, Research report No.19, Dept.of Statistics, University of Helsinki.