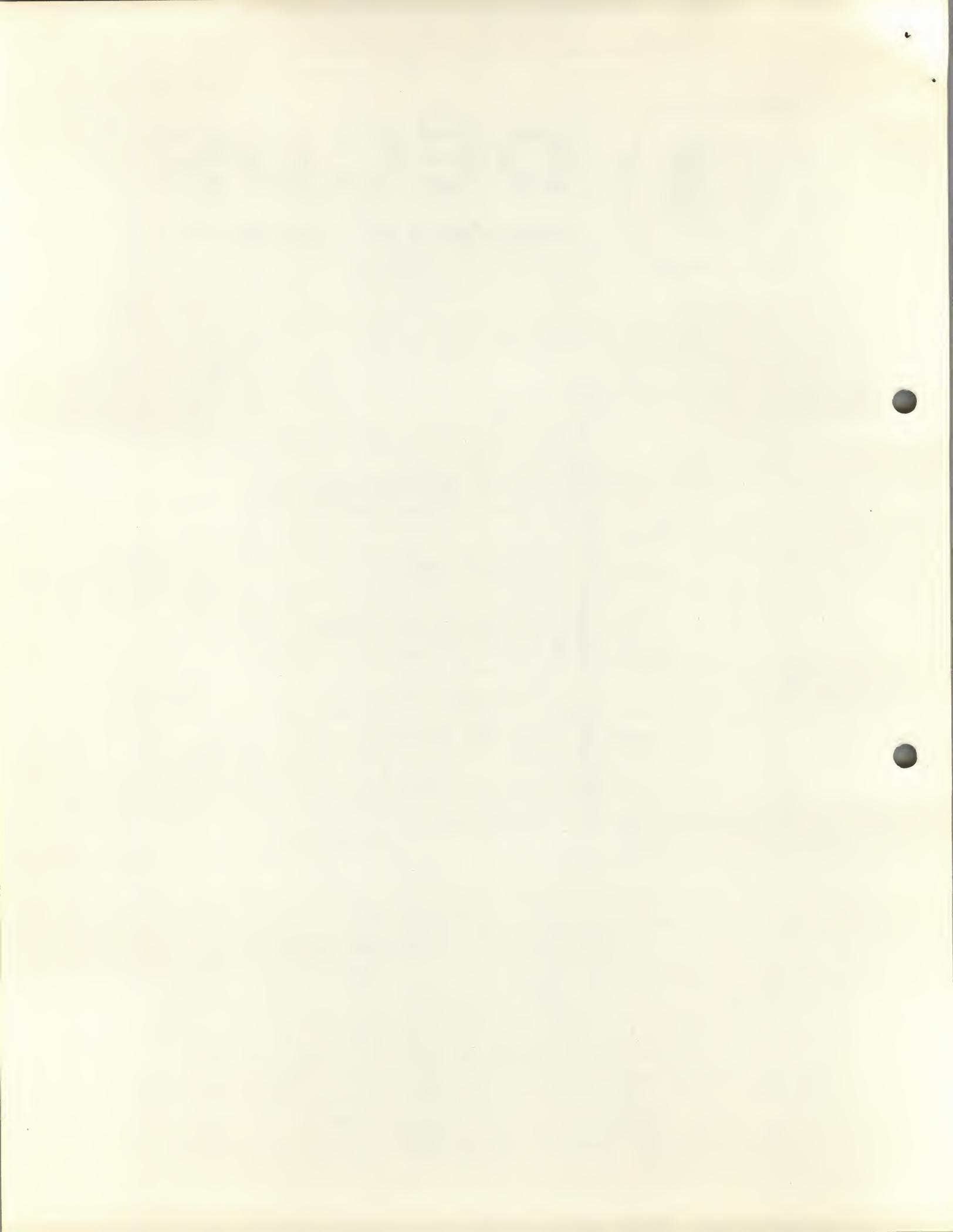




DECUS

PROGRAM LIBRARY

DECUS NO.	8-327
TITLE	CLAN (CLUSTER ANALYSIS) AND GRMN (CALCULATE GROUP MEANS)
AUTHOR	J. N. R. Jeffers
COMPANY	Submitted by: A. J. P. Gore The Nature Conservancy Grange-over-Sands Lancashire, England
DATE	September 25, 1969
SOURCE LANGUAGE	FORTRAN D



CLAN (CLUSTER ANALYSIS) AND GRMN (CALCULATE GROUP MEANS)

DECUS Program Library Write-up

DECUS No. 8-327

CLAN - CLUSTER ANALYSIS

SUMMARY

This program performs a cluster analysis on data for which the nearest neighbor to each individual point has been calculated. The algorithm groups the points into clusters which contain a pair of mutually nearest points and all points which refer to points already included in the groups as their nearest neighbors.

TAPES REQUIRED

Form of program tape - The program is written in the PDP-8 FORTRAN-D language, and is in the source language.

Form of data tape - The data input tape for this program is the output tape from the NNAN program. It contains, for each point, the identifying number of the point, the identifying number of the nearest neighbor, and the distance to the nearest neighbor.

OPERATING INSTRUCTIONS

.FORT	Source program in high-speed reader
*OUT-S:CLAN	
*	
*IN-R:	
* ↑	
↑	Data tape in high-speed reader
*READY	
↑	

The program will request the entry of the number of sets (points), and these should be entered on the teletype and terminated by "Return". If an output tape is required, switch on low-speed punch before typing "Return".

If the program has already been compiled onto the disk, it may be called into core as follows:

.FOSL	
*IN-S:CLAN	
*	
*OPT-	
* ↑	
*READY	Data tape in high-speed reader
↑	

OUTPUT

The program prints the identification numbers of the points in groups. Each group is terminated by -1 and separated by five line-feeds. The output tape from this program, produced simultaneously on the low-speed punch, may be used as one of the input tapes for the group-mean program, GRMN.

STORAGE AND LIMITATIONS

Normal for FORTRAN-D. For the 4K version, the total number of points must not exceed 100. Larger sets can be computed by using the 8K version of FORTRAN.

METHOD

Starting from an arbitrary point, the program traces the chain from each point to its nearest neighbor, until it finds two points which are mutually nearest neighbors. The remaining points are then searched to find all other points which refer to any points already in the chain as their nearest neighbor, and extends the search until the group is complete.

```

*L
C   PROGRAM SEGMENT CLAN
    DIMENSION L(100),M(100),K(100)
    TYPE 100
100  FORMAT (/, "ENTER NO OF SETS",/,/,/)
    ACCEPT 101, N2
    DO 20 I=1, N2
    READ 2, 102, L(I), M(I), V
20   CONTINUE

    TYPE 104
    N4=N5=0
9    N6=0
    N1=N4
    N0=N5
    IF (0-M(1+N0) ) 10, 10, 1
10   K(1+N6)=L(1+N0)
    N6=N6+1
4    K(1+N6)=M(1+N0)
    N6=N6+1
    N7=M(1+N0)
    M(1+N0)=-1
    N0=N7-1
    DO 3 I=1, N6
    IF (M(1+N0)-K(I) ) 3, 2, 3
3    CONTINUE

    GO TO 4
2    M(1+N0)=-1
    DO 7 I=1, N6
    N1=N0=0
6    IF (M(1+N0)-K(I) ) 5, 11, 5
11   K(1+N6)=L(1+N0)
    M(1+N0)=-1
    N6=N6+1
5    N1=N1+1
    N0=N0+1
    IF (N1-N2) 6, 7, 6
7    CONTINUE

    DO 8 I=1, N6
    TYPE 101, K(I)
8    CONTINUE
    N8=-1
    TYPE 101, N8
    TYPE 104

```

```
1      N4=N4+1
      N5=N5+1
      IF (N4-N2) 9,12,9
101    FORMAT (I)
102    FORMAT (I,I,E)
104    FORMAT (/,/././././)
12     END
```

*

GRMN - CALCULATE GROUP MEANS

SUMMARY

This program calculates the means of groups of points selected from a larger matrix of points and variables. It is intended for the calculation of means of multivariate data which have previously been subjected to cluster analysis, so as to prepare the data for the next stage of the clustering process.

TAPES REQUIRED

Form of program tape - The program is written in the PDP-8 FORTRAN-D language, and is in the source language.

Form of data tapes - Two data tapes are required:

(a) A data tape containing the values of each variable, in a standard order, for each point of the original set. The output tape from the CVAL or GRMN programs will be the usual source of these data.

(b) A data tape containing the row numbers of the points to be included in each group. Groups must be terminated by -1. The output tape from the CLAN program will be the usual source of this data tape.

OPERATING INSTRUCTIONS

.FORT	Source program tape in
*OUT-S:GRMN	high-speed reader
*	
*IN-R:	
* ↑	
↑	Data tape in high-speed
*READY	reader

The program will request the entry of the number of variables and sets in the original data matrix, and the number of groups to be computed. These numbers should be entered in sequence on the teletype, and terminated by "Return". After the first data tape has been read, the computer will pause for the second data tape to be inserted in the high-speed reader. Switch on low-speed punch and type ↑ to continue.

If the program has already been compiled onto the disk, it may be called into core as follows:

.FOSL	
*IN-S:GRMN	
*	
*OPT-	
* ↑	
*READY	First data tape in high-speed
↑	reader

The operation of the program then follows as usual.

OUTPUT

The program prints the means for each variable of the new groups. An output tape, produced simultaneously on the low-speed punch, is suitable for re-input to the NNAN program or for input to other multivariate or multiple regression programs.

STORAGE AND LIMITATIONS

Normal for FORTRAN-D.

For the 4K version, the number of variables times the number of points on the original data tape must not exceed 260, and the number of variables must not exceed 16.

METHOD

Simple arithmetic.

```

L
C;   PROGRAM SEGMENT GRMN1
      DIMENSION V(250),VAR(16)
      TYPE 100
100  FORMAT (/, "ENTER NOS OF VARIABLES, SETS AND GROUPS",/,/)
      ACCEPT 101, N5, N6, N7
101;  FORMAT(I, I, I)
      NO=N5*N6
      DO 10 I=1, N0
      READ 2, 102, V(I)
102  FORMAT (E)
10;  CONTINUE
      PAUSE
      NG=0
5;   DO 20 I=1, N5
      VAR(I)=0.0
20;  CONTINUE
      N4=0
3    READ 2, 101, N1
      IF (N1) 1, 2, 2
2    N2=N1*N5-(N5-1)
      DO 30 I=1, N5
      VAR (I)=VAR(I)+V(N2)
      N2=N2+1
30;  CONTINUE
      N4=N4+1
      GO TO 3
1;   VO=N4
      J=0
      DO 4 I=1, N5
      VAR(I)=VAR(I)/VO
      J=J+1
      IF (J-4) 7, 7, 8
8    J=0
      TYPE 103
103  FORMAT (/)
7    TYPE 102, VAR(I)
4;   CONTINUE
      TYPE 103
      NG=NG+1
      IF(N7-NG) 6, 6, 5
6;   STOP
      END

```

*

