DANIEL F. MERRIAM, Editor

# FORTRAN IV PROGRAM FOR Q-MODE CLUSTER ANALYSIS ON DISTANCE FUNCTION WITH PRINTED DENDROGRAM

By

**JAMES M. PARKS**
Lehigh University

in cooperation with the
American Association of Petroleum Geologists
Tulsa, Oklahoma

**COMPUTER CONTRIBUTION 46**
State Geological Survey
The University of Kansas, Lawrence
1970

## Editor's Remarks

This program by J.M. Parks, "FORTRAN IV program for Q-mode cluster analysis on distance function with printed dendrogram", is one of several COMPUTER CONTRIBUTIONS concerned with classification. Others are indicated on the following table.

The program will be made available on magnetic tape for a limited time by the Geological Survey for $15.00 (US). If punched cards are required an extra $10.00 is necessary to cover handling and postage. For a complete list of the available publications and programs write Editor, COMPUTER CONTRIBUTIONS, Kansas Geological Survey, The University of Kansas, Lawrence, Kansas 66044.

METHOD

| Statistical Procedures | Contouring and Graphic Output | Time Trend | Trend Analysis | Harmonic Analysis | Classification | Simulation | Colloquia and Symposia | Groundwater Applications | Other |
|---|---|---|---|---|---|---|---|---|---|
| CC2 | CC15 | SDP12 | SDP3* | SDP24 | SDP4* | CC1* | CC7 | SDP39 | B170-3 |
| CC30 | CC23 | CC19 | SDP11 | CC5 | SDP9* | CC9 | CC12 | SDP42 | SDP28 |
| CC35 | CC36 | | SDP14 | CC16 | SDP13 | CC13 | CC18 | SDP43 | SDP39 |
| CC39 | CC37 | | SDP26 | CC29 | SDP15 | CC24 | CC22 | SDP45 | SDP42 |
| CC41 | CC44 | | B171 | | SDP23 | CC26 | CC40 | | SDP43 |
| | | | CC3 | | SDP27 | | | | SDP45 |
| | | | CC10 | | CC4 | | | | CC8 |
| | | | CC11 | | CC6 | | | | CC14 |
| | | | CC27 | | CC17 | | | | CC25 |
| | | | CC28 | | CC20 | | | | CC33 |
| | | | CC32 | | CC21 | | | | CC34 |
| | | | | | CC31 | | | | CC43 |
| | | | | | CC38 | | | | CC45 |
| | | | | | CC42 | | | | |
| | | | | | CC46 | | | | |

* Active Member, ° Associate Member, + Junior Member, American Association of Petroleum Geologists

# FORTRAN IV PROGRAM FOR Q-MODE CLUSTER ANALYSIS ON DISTANCE FUNCTION WITH PRINTED DENDROGRAM

by

James M. Parks

## INTRODUCTION

In statistical literature, "classification" means the assignment of elements into a priori defined classes (Rohlf, 1963, p. 3). In taxonomy (biology and paleontology), geology, oceanography and this paper, classification is the act or result "of putting similar objects into an unknown number of distinct categories, with the objects in each category being more similar to each other than to the objects in all other categories" (Parks, 1966, p. 703). Various techniques of cluster analysis (also termed "numerical taxonomy", Sokal and Sneath, 1963) have been used to find "natural" classifications inherent in the data (some recent examples are Sokal and Sneath, 1963; Parks, 1964, 1966; and Bonham-Carter, 1965,1967). Factor analysis (Imbrie and Purdy, 1962), principal components analysis (McCammon, 1966), canonical analysis (Oxnard, 1969), multiple component analysis (McCammon, 1968a) and optimizing an objective function (Ward, 1963) also have been used for this purpose. Most of these methods produce essentially similar results if applied to the same data (Parks, 1966, 1969).

Early versions of the computer program described in this paper were written in MAD language for an IBM 704 computer while I was with The Pure Oil Company Research Center. Later at the Union Oil Company of California Research Center I rewrote the early version in FORTRAN for an IBM 1620 utilizing disk storage for the similarity matrix. A subsequent version was written in FORTRAN for an IBM 360/30 at the Union Research Center, in which the similarity matrix did not have to be stored at all. A Univac 1108 was used for the first successful attempts at combining an R-mode principal components analysis with a Q-mode cluster analysis. At Lehigh University I rewrote these programs for a GE 225 computer, utilizing a segment-link-chain technique so that the entire program did not have to occupy memory at one time, but the relatively small memory of the GE 225 imposed a limit of 200 samples. Finally the program has been rewritten in FORTRAN IV for Lehigh's CDC 6400 computer, using overlays to achieve a 1000 sample-200 variable capacity on a 65k memory machine (with 4 scratch tapes). I wish to thank the Computing Center of Lehigh University for providing unsponsored research computer time and helpful advice.

## METHOD

Cluster analysis usually begins with a Q-mode sample-by-sample comparison on some measurements of similarity. The resulting N x N similarity matrix then is searched repeatedly for the pairs or groups of samples exhibiting the closest similarities to derive a dendrogram that depicts the sample relationships in a simple two-dimensional diagram.

Several measures of similarity have been used--product-moment correlation coefficient, cosine-theta, various types of matching coefficients and the simple distance function. Some of these can only be used on continuous variates, some only on presence-absence data, and some only on multistate coded data and integer counts. In many real-world problems, data exist in more than one of these modes; only the simple distance function seemingly is capable of handling such mixed-mode data (Parks, 1968).

The simple distance function is a Euclidean measurement and presupposes that the coordinates on which the measurements are based are orthogonal. In any given study using a large number of variables it is almost inevitable that many variables are correlated and therefore not orthogonal. Before the Q-mode cluster analysis is performed, it is necessary to transform the raw variables to an uncorrelated orthogonal set of variables. This is done by performing an R-mode factor or principal components analysis on the variables and then computing factor scores or component measurements for each sample. An added advantage of the R-mode principal components analysis is a reduction in the number of (transformed) variables used in the Q-mode cluster analysis.

The usual R-mode factor analysis is performed on a matrix of product-moment correlation coefficients, but these are inadequate for mixed-mode data. For the program described in this paper, an R-mode matrix of the one-complement (1.0-D(I,J)) of the normalized distance function is used (the correlation coefficient matrix is available as an option).

Most available factor analysis or eigenvalue-eigenfactor computer programs are limited by the number of variables they can handle. A special principal components analysis routine was written for this program utilizing a Householder-type algorithm that requires only half of a symmetric matrix.

This is accomplished by storing the upper half of the matrix row-wise in a linear array, and retrieving elements with an ISYMM function given row and column indices--this allows a medium-sized 65k computer to handle up to a 200 variable R-mode analysis.

The Q-mode cluster analysis algorithm used in this program could be termed an agglomerative, polythetic, unweighted pair-group method (Sokal and Sneath, 1963), or a monotonic hierarchical method based on average within-group pair-wise distance (McCammon, 1968b, p. 1664). To achieve an algorithm capable of handling 1000 samples, it was obvious that the entire matrix could not be in memory at one time. As the Q-mode distance functions are calculated, they are sorted and only the N smallest (closest resemblance) are retained, along with the indices of the pairs under comparison. After distance functions for all pairs of samples have been calculated, the clustering procedure is started. The pair of samples with the smallest distance function is selected and recorded, the factor measurements for the two samples are averaged forming a new sample retaining the lowest sequential sample designator and deleting the highest. All distance function comparisons in the sorted list that involve either of the samples in the selected pair are recomputed using the averaged factor measurements of that pair. Then the next pair in the list is combined and the above procedure is repeated.

The number of samples entering into each group is recorded, so that a weighting factor of one is used for each original sample in the averaging procedure--that is, the stems in each group are "unweighted". After N/2 samples have been clustered, the distance functions are all recalculated and sorted before continuing the clustering, and again when N - N/5 samples have been clustered this procedure is repeated. This is done as the major clusters come together to uncover any new relationships.

After all the samples have been clustered, the cluster list is resorted for the dendrogram. The largest distance encountered is set to the page width, and the dendrogram is "drawn" on a blanked out section of core memory, and then written onto the output tape to be printed by the line printer. The number of samples entering into each cluster is used to center the stems on the cluster diagram. Both the sequential sample numbers and the original alphanumeric sample designators are printed down the left-hand margin.

## PROGRAM DESCRIPTION

This program performs a number of operations in sequence:

(1) Normalizes data for each variable by finding maximum and minimum values, subtracting the minimum value from each sample measurement for that variable and dividing by the difference between the maximum and minimum value. All measurements then range from 0.0 to 1.0, giving equal weight to each variable.

(2) Computes an R-mode similarity matrix, comparing each variable with all other variables across all samples, using either the product-moment correlation coefficient or the simple distance function. The formula used for the simple distance function for R-mode analysis is

$$D_{1,2} = 1.0 - \left[ \sum_{i=1}^{N} (X_{1i} - X_{2i})^2 / N \right]^{1/2} .$$

As this matrix is symmetrical, only the upper half of the matrix and the principal diagonal are stored in a linear array in order to conserve memory space.

(3) Computes an R-mode principal components analysis (factor analysis with unities in the principal diagonal) on either the correlation coefficient matrix or the distance function matrix, using a modification of Householder's iterative method. The eigenvalues and eigenvectors are extracted sequentially, terminating when the cumulative variance accounted for exceeds the stopping criterion ECD set on the control card, Card 2 (default value=80.0), or when the variance accounted for by the last factor extracted is less than the stopping criterion ECK specified on Card 2 (default value=5.0), or when IFM desired number of factors has been extracted, or when the maximum of 10 factors has been extracted.

(4) Computes factor score measurements for each sample [by Harman's short regression and square root method (Harman, 1960, p. 349-356)] and normalizes these measurements column-wise.

Steps (1) through (4) produce new orthogonal (uncorrelated) variables, and also reduce the number of variables needed for the following Q-mode cluster analysis.

(5) Computes Q-mode similarity matrix, comparing each sample with all other samples across all variables (factor score measurements). The similarity coefficient used is the simple distance function. The formula for distance function for Q-mode analysis is (Parks, 1966, p. 704),

$$D_{1,2} = \left[ \sum_{i=1}^{M} (X_{i1} - X_{i2})^2 / M \right]^{1/2} .$$

For the 1000 sample capability of this program, this matrix, or even the upper half of the matrix, would far exceed the core memory capacity of nearly all computers, and in fact storing this matrix is not necessary. The distances between all possible pairs of samples are calculated, but only the smallest distances are saved in a linear array. This array then is processed, and as samples are paired, one sample name (or number) is eliminated, the measurements for the combined sample are averaged (weighted according to the number of samples entering into the combination from each side) and the new sample takes the designation of the lowest sequential number of the pair, or the sequential designator of the previously combined pair (or pairs) entering into a new combination that contains the largest number of original samples. All distances involving the samples farther down the list are recalculated. One pass through all the sample pairs is sufficient to form all the clusters, but in practice the clustering is performed in three passes.

(6) Finally the clusters are sorted and regrouped in the proper sequence for plotting the dendrogram, and the dendrogram is plotted by the line printer as part of the output, without recourse to an external X-Y plotter.

## DISCUSSION OF THE OPTIONS

One purpose in writing this program was to make the computer classification of multivariate data as objective as possible--on the Data Control Card (Card No. 2) there are default options for every control parameter except one, the number of variables used--but despite all efforts, computer classification remains something of an art, and user's options are therefore provided. These options include:

(1) Selection of either (a) product-moment correlation coefficient or (b) one-complement of normalized distance function for the R-mode principal component analysis (in practice on geological data I have found that the distance function gives more meaningful results).

(2) Stopping criterion for R-mode principal component analysis: the default option provides for extracting principal components until more than 80 percent of the total variance has been accounted for, by components accounting for more than 5 percent of variance and in practice, for the initial classification attempt, this seemingly works well. After the first results are seen, the user may wish to set (a) either a higher or lower percent total variance accounted for stopping criterion, or (b) higher or lower limits to minimum percent of variance accounted for by <u>each</u> component, or (c) a specified number of components (up to a maximum of 10).

(3) The criterion EPSI for accuracy of extraction of components is preset at 0.005, which appears to work satisfactorily for most applications, but this also may be specified by the user--a smaller value would increase the number of iterations required and lengthen the running time of the program.

(4) Provision is made for variable format, to be read in with data (Card No. 3) to allow compatability with existing data decks. Each sample must have an alphanumeric identifier preceding the data: if a preexisting data deck has no sample numbers, it will be necessary to insert another card in front of the data for each sample (which can be left blank if there are no specific sample numbers or identifiers available, as sequential sample numbers are automatically provided on the cluster diagram). An example format (Card No. 3) for this situation might be (2A6/(20F4.0)). It is important to note that a two-word identifier is needed, and if the first word is not six characters long (default value of control parameter I on Card No. 2 is 6), "I" must be set on Card No. 2 to one of the following integer values: 1, 3, 5, or 10. For example, if a data deck has a 4 character sample identifier, "I" can be set to 1 (with a variable format on Card 3 of (A1, A3, 10F6.3)); or "I" may be set to 3, with a variable format like (A3, A1, 10F6.3).

(5) The Q-mode clustering of samples may be performed on "unweighted" normalized distances, or these distances may be "weighted" according to the percent of total variance accounted for by each

principal component used. Not enough comparisons have yet been made on the same sets of data clustered both ways to determine which method is most effective. Other options, such as a logarithm or square-root transformation of the raw data, could be added but have not been included in this program.

## OPERATING INSTRUCTIONS

This cluster analysis program consists of seven object decks which are called and loaded as OVERLAYS as needed. Information is transferred from OVERLAY to OVERLAY through COMMON and four scratch tapes (or simulated tapes on disk).

Input:

Card 1: Title Card
- Cols. 1-80: Any desired alphanumeric information

Card 2: Control Card
- Cols. 1-4: M, number of input variables, right justified (Maximum=200) (I4)

**********

NOTE: This is all the information that is required on the control card. All other parameters have default values, which can be changed for more control over the program.

**********

- Cols. 5-6: I, number of alphanumeric characters in first A-field of sample designation (right justified); (if left blank, default value is 6) (I2)
- Cols. 7-8: IFM, maximum number of factors to be extracted (if left blank, default value is 10) (I2)
- Cols. 9-10: JKL, weighting code for factor measurements . If zero, or left blank, factor measurements are unweighted. If JKL equals or is greater than 1, factor measurements are weighted according to percent of total variance accounted for by each factor.
- Cols. 11-14: EPSI, criterion for accuracy in extracting factors (if left blank, default value is 0.005) (F4.0)
- Cols. 15-24: ECK, the minimum percent variance accounted for by a factor for that factor to be accepted for further calculations, after 60.0 percent variance has been accounted for (if left blank, default value is 5.0) (F6.0)
- Cols. 25-34: ECH, value of distance function on cluster diagram at which sample designators are double-spaced to emphasize separation between major clusters (if left blank, default value is 1/2 largest distance). (F6.0)
- Cols. 35-44: ECD, minimum cumulative variance accepted for total number of factors used in further calculations (if left blank, default value is 80.0) (F6.0)
- Col. 48: ICH, determines similarity function used: If Col. 48= 1, R-mode factor matrix is computed on simple distance function; if Col. 48=2, R-mode factor matrix is computed on correlation coefficient (If ICH left blank, default value = 1, for distance function matrix) (I1)

Card 3: Variable Format Card, Format used to read data cards, enclosed in parentheses, must begin with 2A fields. Example format as used in test problem later described: (2A6, 10F6.0/(12X, 10F6.0))

Card 4: Data Cards (as used in the example problem)
- Cols. 1-12: Alphanumeric sample designator
- Cols. 13-80: Data, according to variable format on Card 3.

NOTE: According to the value of I placed in column 6 of Card 2 (Control Card) and the format placed on Card 3, these column values may be changed.

Card 5: End of Data Card, indicated by 999999 in Columns 1-6, followed by as many blank cards as needed to equal number of cards in data set for each sample. In test problem described later with 12 variables using FORMAT (2A6, 10F6.0/(12X, 10F6.0)), there are 2 cards in the data set for each sample, so Card 5

is to be followed by one blank card. Two or more jobs may be done in sequence, starting with Cards 1 through Card 5.
Last card in data deck must have in Columns -6: FINISH

TEST EXAMPLE

The test example, complete with input data and output cluster diagrams shown in the Appendix following the computer program listing, is a 12 variable 40 sample subset from a study of Recent carbonate sediment constituents on the Bahamas Platform around Andros Island (Purdy, 1960, 1963a, 1963b). A number of different multivariate computer classification techniques have used this set of data as an illustrative example (Ward, 1963; Parks, 1964, 1966, 1969; McCammon, 1968; Bonham-Carter, 1967).

This 40 sample 12 variable problem ran on the CDC 6400 computer in 5.9 seconds central processor time, utilizing 4 factors accounting for 83.4 percent of the total variance. The full 216 sample Bahama sediment problem (Purdy, 1960) ran in 89 seconds CP plus 315 seconds PP time.

This example was run with the default control values--only the number of variables (12) was entered on Card No. 2. It is readily apparent that there are six distinct clusters present in the data.

## REFERENCES


Bonham-Carter, G.F., 1965, A numerical method of classification using qualitative and semi-quantitative data, as applied to the facies analysis of limestones: Canadian Petroleum Geology Bull., v. 13, no. 4, p. 482-502.

Bonham-Carter, G.F., 1967, FORTRAN IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers: Kansas Geol. Survey Computer Contr. 17, 28 p.

Harman, H.H., 1960, Modern factor analysis: University of Chicago Press, Chicago, 469 p.

Imbrie, J., and Purdy, E.G., 1962, Classification of modern Bahamian carbonate sediments: Am. Assoc. Petroleum Geologists Mem. 1, p. 253-272.

McCammon, R.B., 1966, Principal component analysis and its application in large scale correlation studies: Jour. Geology, v. 74, no. 5, pt. 2, p. 721-733.

McCammon, R.B., 1968a, Multiple component analysis and its application in classification of environments: Am. Assoc. Petroleum Geologists Bull., v. 52, no. 11, pt. 1, p. 2178-2196.

McCammon, R.B., 1968b, The dendrograph: a new tool for correlation: Geol. Soc. America Bull., v. 79, no. 11, p. 1663-1670.

Oxnard, C.E., 1969, Mathematics, shape and function: Am. Scientist, v. 57, no. 1, p. 75-96.

Parks, J.M., 1964, Cluster analysis applied to multivariate geologic problems (abs.): Am. Assoc. Petroleum Geologists Bull., v. 48, no. 4, p. 540.

Parks, J.M., 1966, Cluster analysis applied to multivariate geologic data: Jour. Geology, v. 74, no. 5, pt. 2, p. 703-715.

Parks, J.M., 1968, Classification of mixed mode data by R-mode factor analysis and Q-mode cluster analysis on distance function: Proc. Colloquium in Numerical Taxonomy, Univ. of St. Andrews, Sept. 1968, p. 187-192.

Parks, J.M., 1969, Multivariate facies maps: Kansas Geol. Survey Computer Contr. 40, p. 6-18.

Purdy, E.G., 1960, Recent calcium carbonate facies of the Great Bahama Bank: Unpubl. doctoral dissertation, Columbia Univ., 174 p.

Purdy, E.G., 1963a, Recent calcium carbonate facies of the Great Bahama Bank, 1. Petrography and reaction groups: Jour. Geology, v. 71, no. 3, p. 334-355.

Purdy, E.G., 1963b, Recent calcium carbonate facies of the Great Bahama Bank, 2. Sedimentary facies: Jour. Geology, v. 71, no. 4, p. 472-497.

Rohlf, F.J., 1963, Multivariate methods in taxonomy: Proc. IBM Scientific Computing Sym. on Statistics, Oct. 21-23, 1963, White Plains, New York, p. 3-14.

Sokal, R.R., and Sneath, P.H.A., 1963, Principles of numerical taxonomy: W.H. Freeman and Co., San Francisco, 359 p.

Ward, J.H., 1963, Hierarchical grouping to optimize an objective function: Am. Stat. Assoc. Jour., v. 58, p. 236-244.

APPENDIX

```
      OVERLAY(JPCL,0,0)
      PROGRAM CLUST6(INPUT,TAPE 3=INPUT,OUTPUT,TAPE4=OUTPUT,TAPE6,TAPE7,
     1TAPE8,TAPE5)
C     CLUSTER PROGRAM - LIMIT 200 VARIABLES, 1000 SAMPLES
C     PERFORMS R-MODE PRINCIPAL COMPONENTS ANALYSIS ON VARIABLES
C         LIMIT 10 FACTORS OR PRINCIPAL COMPONENTS, EITHER ON
C         CORRELATION COEFFICIENT MATRIX OR DISTANCE FUNCTION MATRIX
C
C     CARD 1 --- TITLE CARD    80 COLUMNS ALPHANUMERIC
C     CARD 2 --- COLUMNS 1-4 = NUMBER OF VARIABLES (RIGHT JUSTIFIED)
C                COLS 5-6 = I, NUMBER OF ALPHANUMERIC CHARACTERS IN
C                   FIRST A-FIELD OF SAMPLE DESIGNATION (RIGHT JUSTIFIED)
C                   IF  LEFT BLANK, DEFAULT VALUE IS  6
C
C                COLS 7-8 = IFM, MAXIMUM NUMBER OF FACTORS DESIRED
C                   (RIGHT JUSTIFIED))  IF LEFT BLANK, DEFAULT VALUE
C                   IS 10 TO ALLOW OTHER CRITERIA TO CONTROL NUMBER OF
C                   FACTORS (ECD OR ECK)
C                COLS 9-10 = JKL, WEIGHTING CODE FOR FACTOR
C                   MEASUREMENTS   IF ZERO OR LEFT BLANK, FACTOR
C                   MEASUREMENTS ARE UNWEIGHTED
C                   IF JKL = 1, FACTOR MEASUREMENTS ARE WEIGHTED
C                        ACCORDING TO PERCENT VARIANCE ACCOUNTED FOR
C
C                COLS 11-14= EPSI, CRITERION FOR ACCURACY OF SELECTING
C                   FACTORS)  IF LEFT BLANK, DEFAULT VALUE IS 0.005
C
C                COLS 15-24 = ECK, THE SMALLEST PERCENT VARIANCE
C                   ACCEPTED FOR FURTHER CALCULATIONS
C                   IF LEFT BLANK, DEFAULT VALUE IS 5.0
C                COLS 25-34 = ECH, VALUE OF DISTANCE FUNCTION ON CLUSTER
C                   DIAGRAM AT WHICH CLUSTERS ARE DOUBLE SPACED
C                   IF LEFT BLANK, DEFAULT VALUE IS 1/2 LARGEST DISTANCE
C                COLS 35-44 = ECD, MINIMUM TOTAL PERCENT VARIANCE
C                   ACCEPTED FOR TOTAL NUMBER OF FACTORS TO BE USED
C                   IF LEFT BLANK, DEFAULT VALUE IS 80.0
C                COLS 45-48 = ICH, (RIGHT JUSTIFIED)
C                   IF COL 48 = 1, FACTOR MATRIX ON VARIABLES IS COMPUTED
C                   ON SIMPLE DISTANCE FUNCTION MATRIX OF VARIABLES
C
C                   IF COL 48 = 2, FACTOR MATRIX OF VARIABLES IS COMPUTED
C                   ON CORRELATION COEFFICIENT MATRIX
C                   IF COL 48 LEFT BLANK, DEFAULT VALUE = 1 (FOR
C                   DISTANCE FUNCTION MATRIX)
C
C     CARD 3 --- VARIABLE FORMAT CARD, MUST BEGIN WITH 2 A-FIELDS
C                   FIRST A-FIELD MUST BE A6
C                   EXAMPLE VARIABLE FORMAT = (2A6,10F6.0,/(12X,10F6.0))
C
C     CARD 4 --- DATA CARDS, USUALLY
C                   COLS 1-12 ALPHANUMERIC SAMPLE DESIGNATION
C                   COLS 13-80 DATA (ACCORDING TO VARIABLE FORMAT CARD 3)
C
C     CARD 5 --- END OF DATA IS INDICATED BY 999999 IN COLS 1-6
C         CARD TYPE 5 MUST BE FOLLOWED BY AS MANY BLANK CARDS AS NEEDED
C         TO EQUAL NUMBER OF CARDS IN DATA SET FOR EACH SAMPLE
C
```

```
C         TWO OR MORE JOBS MAY BE DONE IN SEQUENCE, STARTING WITH CARD 1
C         THROUGH CARD 5
C
C     LAST CARD IN DATA DECK MUST HAVE IN COLS 1-6 FINISH
C
C     FACTOR MEASUREMENTS ARE COMPUTED FOR EACH SAMPLE, WHICH ARE THEN
C         NORMALIZED, AND USED AS INPUT INTO CLUSTERING ROUTINE BASED ON SIMPLE
C         DISTANCE FUNCTION
C
C     CLUSTERING OF SAMPLES IS PERFORMED ON SIMPLE DISTANCE FUNCTION
C          DATA ARE NORMALIZED ( NOT STANDARDIZED) BY VARIABLES
C
C     CLUSTER DIAGRAM IS PRINTED OUT BY COMPUTER PRINTER, WITH
C         SEQUENTIAL SAMPLE NUMBER IN FIRST COLUMN, THEN ACTUAL
C         ALPHANUMERIC SAMPLE DESIGNATOR IN SECOND COLUMN
C
C
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECD,EPSI,ECK,IFM
     1,JKL,BEND,IMT(8),I
      COMMON/BLB/IQ(1000),IS(1000),RCOEF(1000),IC(1000,2),IY(1000),IWYE(
     11000)
      COMMON/BLC/IX(1000)
      COMMON/BLD/PC(10)
      JPCL=4LJPCL
      FINI=(5HFINIS)
    3 CONTINUE
      REWIND 5
      REWIND 6
      REWIND 7
      REWIND 8
      READ(3,201) (CHAR(I),I=1,16)
  201 FORMAT(16A5)
      IF(CHAR(1).EQ.FINI) GO TO 4
      READ(3,202) M,I,IFM,JKL,EPSI,ECK,ECH,ECD,ICH
  202 FORMAT(I4,3I2,F4.0,3F10.0,I4)
      IF(ICH.LE.0) ICH=1
      IF(ICH.EQ.1) CALL OVERLAY(JPCL,5,0)
      IF(ICH.EQ.2) CALL OVERLAY(JPCL,6,0)
      CALL OVERLAY(JPCL,1,0)
      CALL OVERLAY(JPCL,2,0)
      CALL OVERLAY(JPCL,3,0)
      CALL OVERLAY(JPCL,4,0)
      GO TO 3
    4 CONTINUE
      END
```

```
      OVERLAY(JPCL,1,0)
      PROGRAM FACMES
C     PROGRAM FACMES
C    TO CALCULATE FACTOR MEASUREMENTS FROM RAW DATA MATRIX
C       INPUT, 200 VARIABLES BY 1000 SAMPLES
C       NORMALIZES VARIABLES
C       CALCULATES CORRELATION COEFFICIENT MATRIX BY VARIABLES
C       EXTRACTS PRINCIPAL FACTORS LARGER THAN 5 PERCENT OF VARIANCE
C       TAKES SQUARE ROOT OF FACTOR MATRIX
C       CALCULATES BETA COEFFICIENTS
C       ESTIMATES FACTOR MEASUREMENTS
C       READY FOR INPUT TO CLUSTER BY DISTANCE
C
C
C     FUNCTION ISYMM TO FIND LINEAR SUBSCRIPT IN HALF OF A
C     SYMMETRIC MATRIX WITH ITS PRINCIPAL DIAGONAL STORED IN ROW-WISE
C     ORDER, GIVEN ROW(I) AND COLUMN(J) AND SIZE OF MATRIX(M)
C
      DIMENSION RTT(200),VEE(200),R(20100),X(200),TT(200),T(200),F(200,1
     10),VV(10),H(10),AR(200) ,SUMASQ(10),HSQ(200),       TN(200),IMT(8)
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECC,EPSI,ECK,IFM
     1,JKL
      COMMON/BLD/PC(10)
      EMM=M
      NNN=(M*M+M)/2
      PCO=0.0
      DO 296 J=1,10
      SUMASQ(J)=0.0
      VV(J)=0.0
  296 PC(J)=0.0
      DO 417 J=1,M
      HSQ(J)=0.0
      DO 417 K=1,10
  417 F(J,K)=0.0
      WRITE(4,300)
  300 FORMAT(25X,38HPRINCIPAL FACTOR ANALYSIS OF VARIABLES//)
      READ(8) (R(I),I=1,NNN)
      REWIND 8
      PEPSI=EPSI
  412 DO 401 MP=1,10
      MP=MP
      DO 324 J=1,M
  324 VEE(J)=0.0
      SUV=0.0
      LLL=0
      NQ=0
      LL=1
  400 CONTINUE
      DO 313 J=1,M
      TT(J)=0.0
  313 T(J)=0.0
      DO 6001 K=1,M
      DO 6003 J=1,M
 6003 AR(J)=0.0
      MK=M-K+1
      L=1
      DO 6005 I=K,M
```

```
      DO 6002 J=1,M
      MJ=ISYMM(J,K,M)
      MI=ISYMM(J,I,M)
      AR(L)=AR(L)+R(MI)*R(MJ)
 6002 CONTINUE
      L=L+1
 6005 CONTINUE
      WRITE(7)  (AR(L),L=1,MK)
 6001 CONTINUE
      REWIND 7
      K=M
      L=1
      DO 6006 I=1,M
      READ(7)  (R(J),J=L,K)
      L=K+1
 6006 K=K+M-I
      REWIND 7
      DO 101 J=1,NNN
      IF(R(J).NE.0.0)  GO TO 102
  101 CONTINUE
      READ(5)(R(J),J=1,NNN)
      REWIND 5
      NQ=1
  102 CONTINUE
      DO 314 J=1,M
      DO 314 I=1,M
      IA=ISYMM(I,J,M)
  314 T(J)=T(J)+R(IA)
      TM=0.0
      DO 315 J=1,M
      IF(ABS(TM).LT.ABS(T(J)))  TM=T(J)
  315 CONTINUE
      DO 316 J=1,M
      TN(J)=T(J)/TM
      IF(T(J).EQ.0.0)  TN(J)=0.0
  316 CONTINUE
   43 DO 317 J=1,M
  317 TT(J)=0.0
      DO 318 J=1,M
      DO 318 I=1,M
      IA=ISYMM(I,J,M)
  318 TT(J)=TT(J)+R(IA)*TN(I)
      TM=0.0
      DO 319 J=1,M
      IF(ABS(TM).LT.ABS(TT(J)))  TM=TT(J)
  319 CONTINUE
      DO 320 J=1,M
      RTT(J)=TT(J)/TM
      IF(TT(J).EQ.0.0)  RTT(J)=0.0
  320 CONTINUE
  414 DO 321 J=1,M
      TEST1=RTT(J)+EPSI
      TEST2=RTT(J)-EPSI
      IF(TN(J).LT.TEST1.AND.TN(J).GT.TEST2)  GO TO 321
      LLL=LLL+1
      IF(LL.EQ.0)  GO TO 413
      LL=LL+1
      IF(LL-4 )413,413,411
```

```
411 EPSI=EPSI+EPSI/5.0
    LL=0
    GO TO 414
413 CONTINUE
    IF(LLL.GT.8) GO TO 40
    IF(NQ.EQ.1) GO TO 40
    GO TO 41
 40 DO 42 K=1,M
    IN(K)=RTT(K)
 42 CONTINUE
    GO TO 43
 41 GO TO 400
321 CONTINUE
    EPSI=PEPSI
    READ(5) (R(J),J=1,NNN)
    REWIND 5
    DO 326 J=1,M
    DO 326 I=1,M
    IA=ISYMM(I,J,M)
326 VEE(J)=VEE(J)+RTT(I)*R(IA)
    DO 327 I=1,M
327 SUV=SUV+RTT(I)*VEE(I)
    SUV=ABS(SUV)
    SUV=SQRT(SUV)
    WRITE(4,301) MP
301 FORMAT(1H ,5X,6HFACTOR,I5)
    DO 328 J=1,M
    F(J,MP)=VEE(J)/SUV
328 IF(VEE(J).EQ.0.0.AND.SUV.EQ.0.0) F(J,MP)=0.0
    IA=0
    DO 329 J=1,M
    DO 329 K=J,M
    IA=IA+1
329 R(IA)=R(IA)-F(J,MP)*F(K,MP)
    WRITE (5) (R(J),J=1,NNN)
    REWIND 5
    DO 330 J=1,M
330 SUMASQ(MP)=SUMASQ(MP)+F(J,MP)*F(J,MP)
    WRITE(4,302) LLL
302 FORMAT(15X,27HNUMBER OF ITERATIONS EQUALS,I7)
    WRITE(4,303) SUMASQ(MP)
303 FORMAT(15X,25HVARIANCE EXTRACTED EQUALS,F9.3//)
    PC(MP)=SUMASQ(MP)*100./EMM
    PCO=PCO+PC(MP)
    IF(MP.LT.IFM) GO TO 401
    IF(MP.EQ.IFM) GO TO 415
    IF(PCO.GT.ECD) GO TO 415
    IF(PCO.GT.60.0.AND.PC(MP).LT.ECK)GO TO 331
401 CONTINUE
    GO TO 415
331 MP=MP-1
415 DO 332 J=1,M
    DO 332 K=1,MP
332 HSQ(J)=HSQ(J)+F(J,K)*F(J,K)
    WRITE(4,304)
304 FORMAT(1H1,5X,30HPRINCIPAL AXIS FACTOR LOADINGS//)
    WRITE(4,305)
305 FORMAT(12X,3HF 1,7X,3HF 2,7X,3HF 3,7X,3HF 4,7X,3HF 5,7X,3HF 6,7X,3
```

```
   1HF 7,7X,3HF 8,7X,3HF 9,6X,4HF 10,8X,3HHSQ//)
    DO 333 K=1,M
    DO 334 J=1,MP
334 VV(J)=F(K,J)
333 WRITE(4,306) K,(VV(I),I=1,10),HSQ(K)
306 FORMAT(2X,I4,11F10.3)
    WRITE(4,307) (SUMASQ(I),I=1,MP)
307 FORMAT(1H0,5X,10F10.3)
    WRITE(4,307) (PC(J),J=1,MP)
    READ(8) (R(J),J=1,NNN)
    REWIND 8
    DO 336 J=2,M
    LJ=J-1
    II=ISYMM(J,J,M)
    DO 337 L=1,LJ
    IL=ISYMM(L,J,M)
337 R(II)=R(II)-R(IL)*R(IL)
    R(II)=SQRT(R(II))
    KJ=J+1
    DO 336 K=KJ,M
    LJ=J-1
    IJ=ISYMM(J,K,M)
    DO 338 L=1,LJ
    IK=ISYMM(L,K,M)
    IL=ISYMM(L,J,M)
338 R(IJ)=R(IJ)-R(IK)*R(IL)
336 R(IJ)=R(IJ)/R(II)
    DO 342 K=1,MP
    DO 342 J=1,M
    DO 343 L=1,J
    IF(L.EQ.J) GO TO 416
    IL=ISYMM(L,J,M)
343 F(J,K)=F(J,K)-F(L,K)*R(IL)
416 IJ=ISYMM(J,J,M)
342 F(J,K)=F(J,K)/R(IJ)
    WRITE(4,309)
    WRITE(4,305)
309 FORMAT(1H1,5X,28HSQUARE ROOT OF FACTOR MATRIX//)
    DO 344 K=1,M
    DO 345 J=1,MP
345 VV(J)=F(K,J)
344 WRITE(4,306) K,(VV(I),I=1,MP)
    DO 346 K=1,MP
    DO 347 J=1,M
    JJ=M+1-J
    JK=ISYMM(JJ,JJ,M)
    DO 348 I=1,M
    II=M+1-I
    IF(II.EQ.JJ) GO TO 347
    JI=ISYMM(JJ,II,M)
348 F(JJ,K)=F(JJ,K)-R(JI)*F(II,K)
347 F(JJ,K)=F(JJ,K)/R(JK)
346 CONTINUE
    WRITE(4,110)
110 FORMAT(1H1,10X,17HBETA COEFFICIENTS//)
    WRITE(4,311)
311 FORMAT(8X,3HF 1,8X,3HF 2,8X,3HF 3,8X,3HF 4,8X,3HF 5,8X,3HF 6,8X,3H
   1F 7,8X,3HF 8,8X,3HF 9,8X,4HF 10//)
```

```
      DO 349 K=1,M
  349 WRITE(4,350) K,(F(K,J),J=1,MP)
  350 FORMAT(1X,I4,10F11.5)
      DO 561 I=1,N
  561 READ(6) Q1,Q2
      DO 352  J=1,N
      READ(6)  (X(I),I=1,M)
      DO 353 K=1,MP
      RTT(K)=0.0
      DO 353 I=1,M
      RTT(K)=RTT(K)+F(I,K)*X(I)
  353 CONTINUE
      WRITE(8)  (RTT(K),K=1,MP)
  352 CONTINUE
      END FILE 8
      REWIND 8
      REWIND 6
      RETURN
      END
```

```
      FUNCTION ISYMM(I,J,M)
      IF(I.GT.J) GO TO 111
  110 ISYMM=(I-1)*M-((I-1)*I)/2+J
      RETURN
  111 ISYMM=(J-1)*M-((J-1)*J)/2+I
      RETURN
      END
```

```
      OVERLAY(JPCL,2,0)
      PROGRAM CLUST5
      DIMENSION TT(10),RT(10),X(10,1000),Y(1000),L(1000,2)
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECD,EPSI,ECK,IFM
     1,JKL
      COMMON/BLB/IQ(1000),IS(1000),RCOEF(1000),IC(1000,2),IY(1000),IWYE(
     11000)
      COMMON/BLC/IX(1000)
      COMMON/BLD/PC(10)
      INTEGER TMP
      INTEGER P
      NN=N-1
      M=MP
      DO 41 J=1,M
      TT(J)=999999.
   41 RT(J)=-99999.
      DO 33 J=1,N
      READ(8) (X(I,J),I=1,M)
      DO 32 I=1,MP
      IF(X(I,J).LT.TT(I)) TT(I)=X(I,J)
   32 IF(X(I,J).GT.RT(I)) RT(I)=X(I,J)
   33 CONTINUE
      REWIND 8
      IF(JKL.GE.1) GO TO 538
      WRITE(4,410)
  410 FORMAT(1H1,42HNORMALIZED FACTOR MEASUREMENTS, UNWEIGHTED//)
      DO 541 K=1,MP
  541 RT(K)=RT(K)-TT(K)
      GO TO 539
  538 CONTINUE
      WRITE(4,537)
  537 FORMAT(1H1, 40HNORMALIZED FACTOR MEASUREMENTS, WEIGHTED//)
      DO 46 K=1,MP
   46 RT(K)=(RT(K)-TT(K))*100./PC(K)
  539 CONTINUE
      DO 47 J=1,MP
      DO 47 I=1,N
   47 X(J,I)=(X(J,I)-TT(J))/RT(J)
      DO 412 J=1,N
  412 WRITE(4,411) J,(X(I,J),I=1,MP)
  411 FORMAT(1X,I4,10F11.5)
      EM=M
      DO 191 J=1,N
      IC(J,1)=1
      IC(J,2)=1
      IQ(J)=0
      IS(J)=0
  191 RCOEF(J)=0.0
      IP=1
      NI=N/2
   19 CONTINUE
      JI=1
      DO 11 I=1,N
      Y(I)=1.0
      DO 11 J=1,2
      L(I,J)=0
   11 CONTINUE
      DO 215 J=1,NN
      IF(X(1,J).LT.0.0) GO TO 215
      IJ=J+1
```

```
      DO 216 K=IJ,N
      IF(X(1,K).LT.0.0) GO TO 216
      R=0.0
      DO 122 I=1,M
  122 R=R+(X(I,K)-X(I,J))*(X(I,K)-X(I,J))
      R=R/EM
      DO 12 I=1,JI
      IF(R.GT.Y(I)) GO TO 12
      II=I
      GO TO 13
   12 CONTINUE
      II=JI+1
   13 TEMP=Y(II)
      JEMP=L(II,1)
      KEMP=L(II,2)
      IF(II.EQ.1) GO TO 16
      IB=II
      DO 14 I=1,IB
      IF(K.EQ.L(I,2).OR.J.EQ.L(I,2))GO TO 216
   14 CONTINUE
   16 Y(II)=R
      L(II,1)=J
      L(II,2)=K
      JI=JI+1
      IB=II+1
      IF(IB.GT.JI) GO TO 216
      DO 15 I=IB,JI
      IF(K.EQ.KEMP) GO TO 17
   21 TTEMP=Y(I)
      JTEMP=L(I,1)
      KTEMP=L(I,2)
      Y(I)=TEMP
      L(I,1)=JEMP
      L(I,2)=KEMP
      TEMP=TTEMP
      JEMP=JTEMP
      KEMP=KTEMP
   15 CONTINUE
      GO TO 216
   17 JI=JI-1
  216 CONTINUE
  215 CONTINUE
      II=1
      DO 211 P=IP,NI
      KP=P
      LP=P-1
      R=0.0
      RCOEF(P)=Y(II)
      IF(LP.EQ.0) GO TO 210
      IF(NI.LT.(N-N/5).AND.RCOEF(P).GT.(RCOEF(LP)+0.002)) GO TO 212
  210 CONTINUE
      IQ(P)=L(II,1)
      IS(P)=L(II,2)
      J=P-1
  147 IF(J.EQ.0) GO TO 59
      DO 703 I=1,J
      K=P-I
      IF(IQ(K).EQ.IQ(P)) GO TO 707
  703 CONTINUE
      GO TO 704
```

```
707 IC(P,1)=IC(K,1)+IC(K,2)
704 J=P-1
    DO 706 I=1,J
    K=P-I
    IF(IQ(K).EQ.IS(P)) GO TO 705
706 CONTINUE
    GO TO 59
705 IC(P,2)=IC(K,1)+IC(K,2)
 59 IEMPJ=0
    IF(IC(P,1)-IC(P,2)) 261,62,62
261 IEMPJ=IQ(P)
    IQ(P)=IS(P)
    IS(P)=IEMPJ
    IEMPJ=IC(P,1)
    IC(P,1)=IC(P,2)
    IC(P,2)=IEMPJ
    IEMPJ=L(P,1)
    L(P,1)=L(P,2)
    L(P,2)=IEMPJ
    IJ=II+1
    IF(IJ.GT.JI) GO TO 62
    DO 262 I=IJ,JI
    IF(IS(P).EQ.L(I,1)) L(I,1)=IQ(P)
    IF(IS(P).EQ.L(I,2)) L(I,2)=IQ(P)
262 CONTINUE
 62 AK=IC(P,1)
    BK=IC(P,2)
    JA=IQ(P)
    JB=IS(P)
    DO 63 J=1,M
    X(J,JA)=(X(J,JA)*AK+X(J,JB)*BK)/(AK+BK)
 63 X(J,JB)=-9.0
    K=II+1
    IF(K.GT.NI) GO TO 67
    DO 64 J=K,NI
    JB=J
    IF(JA.EQ.L(J,2)) GO TO 66
    IF(JA.EQ.L(J,1))GO TO 65
 64 CONTINUE
 65 JC=L(JB,2)
    GO TO 67
 66 JC=L(JB,1)
 67 DO 222 I=1,M
    R=R+(X(I,JA)-X(I,JC))*(X(I,JA)-X(I,JC))
222 CONTINUE
    R=R/EM
    II=II+1
    Y(JB)=R
    IF(KP.EQ.NN) GO TO 20
211 CONTINUE
    KP=KP+1
212 IP=KP
    IF(IP.LE.(2*N/3)) GO TO 9
    NI=NN
    GO TO 19
  9 NI=N-N/5
    GO TO 19
 20 WRITE(4,10)  (CHAR(J),J=1,16)
 10 FORMAT(1H1,16A5)
    RETURN
    END
```

```
      OVERLAY(JPCL,3,0)
      PROGRAM CLPLIT
C     SUBROUTINE CLPLIT(I)
      DIMENSION BD(11),B(24)
      DIMENSION KEY(500),LE(500),IL(500)
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECC,EPSI,ECK
      COMMON/BLB/IQ(1000),IS(1000),RCOEF(1000),IC(1000,2),IY(1000),IWYE(
     11000)
      COMMON/BLC/IX(1000)
      DATA BOX/4H    /
      DATA (BD(J),J=1,11)/4H 0.0,4H 0.1,4H 0.2,4H 0.3,4H 0.4,4H 0.5,4H 0
     1.6,4H 0.7,4H 0.8,4H 0.9,4H 1.0/
      WRITE(4,650) N
  650 FORMAT(5X,I6,2X,7HSAMPLES//)
      WRITE(4,106)
  106 FORMAT(//30X,23HCLUSTER TYPE 6 ANALYSIS/)
      WRITE(4,108)
  108 FORMAT(1X,59H   LEFT SAMPLE   RIGHT SAMPLE   DISTANCE  NO. IN
     1NO. IN)
      WRITE(4,109)
  109 FORMAT(1X,59H     NUMBER         NUMBER        FUNCTION   LEFT
     1 RIGHT/)
      DO 381 J=1,NN
      RCOEF(J)=SQRT (RCOEF(J))
  381 WRITE(4,110) IQ(J),IS(J),RCOEF(J),IC(J,1),IC(J,2)
  110 FORMAT(1X,I10,I15,F12.7,2I8)
      RCOEFO =0.0
      J=NN
      DO 5 K=1,10
      IF(RCOEFO-RCOEF(J))6,5,5
    6 RCOEFO=RCOEF(J)
      IJ=J
    5 J=J-1
      DO 4 L=IJ,NN
    4 RCOEF(L)=RCOEFO
      COP=1.0
      DO 7 K=1,10
      IF(RCOEFO-COP)7,7,8
    7 COP=COP-0.1
    8 RCOEFO=COP+0.1
      IF(ECH)911,911,912
  911 ECH=RCOEFO/2.0
  912 COP=24./RCOEFO
      IECK=ECH*COP
      DO 9 J=1,NN
      IX(J)=RCOEF(J)*COP
      IF(IX(J).LT.1) IX(J)=1
    9 CONTINUE
      NNN=NN-1
      WRITE(4,206)
      WRITE(4,207)
      DO 41 J=1,24
   41 B(J)=BOX
      I=24.0/(RCOEFO*10.0)+.2
      L=1
      DO 42 J=1,11
      B(L)=BD(J)
```

```
      L=L+I
      IF(L-24)42,42,43
   42 CONTINUE
   43 CONTINUE
      WRITE(4,204) (B(J),J=1,24)
      DO 15 J=1,NNN
      NNNN=NNN-J
      JAY=IX(J)
      DO 15 I=1,NNNN
      IF(IQ(J+I).NE.IQ(J)) GO TO 13
      IF(IX(J+I).LT.JAY) IX(J+I)=JAY
      JAY=IX(J+I)
   13 IF(IS(J+I).NE.IQ(J)) GO TO 15
      IF(IX(J+I).LT.JAY) IX(J+I)=JAY
      JAY=IX(J+I)
   15 CONTINUE
  206 FORMAT(1H1,15X,15HCLUSTER DIAGRAM/)
  207 FORMAT(10X,20HVALUE OF COEFFICIENT/)
  204 FORMAT(30X,24A4//)
      DO 3 J=1,N
    3 IWYE(J)=0
      DO 1 J=1,1000
    1 IY(J)=0
      DO 2 J=1,500
      KEY(J)=0
      IL(J)=0
    2 LE(J)=0
      I=1
      IP=1
      IY(IP)=IQ(I)
      IP=IP+1
      IF(IX(I)-IECK) 901,901,902
  902 IP=IP+1
  901 IY(IP)=IS(I)
      IWYE(I)=IP
      IP=IP+1
      KEY(I)=IQ(I)
      IQ(I)=-IQ(I)
      IS(I)=-IS(I)
      LP=1
      K=2
      IL(I)=N
   21 CONTINUE
      DO 20 J=K,NN
      IF(IQ(J)-KEY(I))24,23,24
   23 LE(I)=IC(J,2)
  128 IF(IX(J)-IECK)26,26,25
   25 IP=IP+1
   26 IY(IP)=IS(J)
      IWYE(J)=IP
      LP=LP+1
      IP=IP+1
      IL(I)=IL(I)-LE(I)
      IF(IC(J,2)-1)20,27,32
   27 IQ(J)=-IQ(J)
      IS(J)=-IS(J)
   31 IF(IL(I)-1)29,201,201
  201 IF(J-NN)20,21,45
```

```
29 IF(I-1)21,21,30
30 I=I-1
   GO TO 31
32 I=I+1
   IL(I)=IC(J,2)-1
   KEY(I)=IS(J)
   IQ(J)=-IQ(J)
   IS(J)=-IS(J)
   GO TO 21
24 IF(IS(J)-KEY(I))20,33,20
33 CONTINUE
28 IF(IX(J)-IECK)35,35,34
34 IP=IP+1
35 I=1
   KEY(I)=IQ(J)
   IY(IP)=IQ(J)
   IWYE(J)=IP
   LP=LP+1
   IP=IP+1
   IQ(J)=-IQ(J)
   IS(J)=-IS(J)
   IL(I)=IC(J,1)-1
   GO TO 21
20 CONTINUE
   IF(LP.LT.NN) GO TO 21
45 CONTINUE
   RETURN
   END
```

```
      OVERLAY(JPCL,4,0)
      PROGRAM CLPLAT
      DIMENSION A(24,1000)
      DIMENSION C(1001,2)
      REAL IQ,IS,IWYE
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECD,EPSI,ECK
      COMMON/BLB/IQ(1000),IS(1000),RCOEF(1000),IC(1000,2),IY(1000),IWYE(
     11000)
      COMMON/BLC/IX(1000)
      EQUIVALENCE (IQ,C)
      DATA BOX,BKD,BSD,BLD/4H    ,4H----,4HI   ,4HI---/
      IM=N+N/4
      IF(IM.GT.1000) IM=1000
      DO 340 I=1,24
      DO 340 J=1,IM
  340 A(I,J)=BOX
      K=0
      JI=0
      JK=0
      JL=0
      LL=0
C     JK= ROW FOR UPPER LINE
C     K = COLUMN START FOR UPPER LINE
C     JI = COLUMN END FOR UPPER LINE
C     JL = ROW FOR LOWER LINE
C      LL = CLOUMN START FOR LOWER LINE
C     JI = CLOUMN END FOR LOWER LINE
      DO 1000 I=1,NN
      LK=IC(I,1)
      JM= IC(I,2)
      CALL CPLOT1(I,LK,JM,K,JI,JK,JL,LL)
      IF(JI-1)449,455,449
  449 IF(K-JI)411,445,445
  411 IF(A(K,JK)-BSD)476,477,476
  477 A(K,JK)=BLD
      K=K+1
  476 IF(K-JI)446,446,445
  445 JI=K
      IX(I)=JI
  446 IF(LL-JI)412,447,447
  412 IF(A(LL,JL)-BSD)478,479,478
  479 A(LL,JL)=BLD
      LL=LL+1
  478 IF(LL-JI)448,448,447
  447 JI=LL
      IX(I)=JI
  448 DO 376 L=K,JI
  376 A(L,JK)=BKD
      DO 375 L=LL,JI
  375 A(L,JL)=BKD
  455 IF(JL-JK)485,486,486
  485 MM=JL
      JL=JK
      JK=MM
  486 CONTINUE
      DO 374 L=JK,JL
  374 A(JI,L)=BSD
```

```
1000 CONTINUE
     DO 561 I=1,N
 561 READ(6) C(I,1),C(I,2)
     REWIND 6
     M=IP
     DO 365 J=1,M
     IF(IY(J)-1)366,367,367
 366 WRITE(4,115) (A(K,J),K=1,24)
     GO TO 365
 367 IK=IY(J)
     WRITE(4,116) IY(J),(C(IK,I),I=1,2),(A(K,J),K=1,24)
 365 CONTINUE
 116 FORMAT(3X,I4,3X,2A6,1X,24A4)
 115 FORMAT(23X,24A4)
     RETURN
     END
```

```
      SUBROUTINE CPLOT1(KK,LK,JJ,K,JI,JK,JL,LL)
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECD,EPSI,ECK
      COMMON/BLB/IQ(1000),IS(1000),RCOEF(1000),IC(1000,2),IY(1000),IWYE(
     11000)
      COMMON/BLC/IX(1000)
      IF(LK-1)45,46,47
   46 IF(JJ-1)45,60,45
   47 IF(JJ-1)45,62,63
   60 K=1
      JI=IX(KK)
      LL=1
      JL=IWYE(KK)
      JK=IWYE(KK)-1
      IF(IX(KK)-IECK)72,72,71
   71 JK=JK-1
   72 GO TO 70
   62 L=KK-1
      LL=1
      DO 85 IJ=1,N
      IF(IQ(KK)-IQ(L))85,86,85
   86 JK=IWYE(L)
      K=IX(L)
      IF(IX(L)-IX(KK))87,88,88
   88 K=IX(KK)
   87 JI=IX(KK)
   89 JL=IWYE(KK)
      IWYE(KK)=(IWYE(KK)+IWYE(L))/2
      GO TO 70
   85 L=L- 1
   63 L=KK-1
      DO 95 IJ=1,N
      IF(IQ(KK)-IQ(L))95,96,95
   96 JK=IWYE(L)
      K=IX(L)
      IF(IX(L)-IX(KK))97,98,98
   98 K=IX(KK)
   97 JI=IX(KK)
   99 LK=KK-1
      DO 101 II=1,N
      IF(IQ(LK)-IS(KK))101,102,101
  102 JL=IWYE(LK)
      LL=IX(LK)
      IF(IX(LK)-IX(KK))103,105,103
  105 LL=IX(KK)
  103 IWYE(KK)=(IWYE(L)+IWYE(LK))/2
      GO TO 70
  101 LK=LK-1
   95 L=L-1
   70 CONTINUE
   45 CONTINUE
      RETURN
      END
```

```
      OVERLAY(JPCL,5,0)
      PROGRAM DISSYM
      DIMENSION RTT(200),VEE(200),R(20100),X(200),TT(200),T(200)
      DIMENSION C(1001,2)
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECD,EPSI,ECK,IFM
     1,JKL,BEND,IMT(8),I
      IF(JKL.LE.0) JKL=0
      IF(I.LE.0) GO TO 46
      IF(I.EQ.1) GO TO 47
      IF(I.EQ.3) GO TO 48
      IF(I.EQ.10) GO TO 49
      IF(I.EQ.5) GO TO 2038
   47 BEND=(1H9)
      GO TO 50
   48 BEND=(3H999)
      GO TO 50
 2038 BEND=(5H99999)
      GO TO 50
   49 BEND=(10H9999999999)
      GO TO 50
   46 BEND=(6H999999)
   50 CONTINUE
      READ(3,103) (IMT(I),I=1,8)
  103 FORMAT(8A10)
      IF(IFM.LE.0) IFM=0
      IF(EPSI.LE.0.0)EPSI=0.005
      IF(ECK.LE.0.0)ECK=5.0
      IF(ECD.LE.0.0) ECD=80.0
      DO 203 J=1,M
      TT(J)=999999.
  203 T(J)=-999999.
      DO 291 J=1,1001
      READ(3,IMT) C(J,1),C(J,2),(X(I),I=1,M)
      IF(C(J,1).EQ.BEND) GO TO 299
      DO 292 K=1,M
      IF(X(K).LT.TT(K))TT(K)=X(K)
      IF(X(K).GT.T(K))T(K)=X(K)
  292 CONTINUE
  291 WRITE(8) (X(I),I=1,M)
  299 N=J-1
      END FILE 8
      REWIND 8
      DO 562 I=1,N
  562 WRITE(6) C(I,1),C(I,2)
      DO 293 K=1,M
  293 T(K)=T(K)-TT(K)
      DO 294 J=1,N
      READ(8) (X(I),I=1,M)
      DO 295 K=1,M
      X(K)=(X(K)-TT(K))/T(K)
  295 CONTINUE
  294 WRITE(6) (X(I),I=1,M)
      END FILE 6
      REWIND 6
      REWIND 8
      WRITE(4,10)(CHAR(I),I=1,16)
   10 FORMAT(1H1,16A5)
```

```
      DO 393 K=1,M
      RTT(K)=0.0
  393 VEE(K)=0.0
      NNN=(M*M+M)/2
      DO 310 J=1,NNN
  310 R(J)=0.0
      EMM=N
      DO 561 I=1,N
  561 READ(6) Q1,Q2
      DO 394 J=1,N
      READ(6) (X(I),I=1,M)
      IJ=1
      DO 297 I=1,M
      R(IJ)=0.0
      IJ=IJ+1
      L=I+1
      IF(L.GT.M) GO TO 394
      DO 297 K=L,M
      R(IJ)=R(IJ)+(X(I)-X(K))*(X(I)-X(K))
  297 IJ=IJ+1
  394 CONTINUE
      REWIND 6
      IJ=1
      DO 312 J=1,M
      R(IJ)=0.0
      IJ=IJ+1
      KJ=J+1
      IF(KJ.GT.M) GO TO 312
      DO 312 K=KJ,M
      R(IJ)=SQRT(R(IJ)/EMM)
  312 IJ=IJ+1
      DO 6012 J=1,NNN
      R(J)=1.0-R(J)
 6012 CONTINUE
      WRITE(8) (R(J),J=1,NNN)
      END FILE 8
      REWIND 8
      WRITE(5) (R(J),J=1,NNN)
      END FILE 5
      REWIND 5
      WRITE(4,300)
  300 FORMAT(1H0,5X,64HR-MODE PRINCIPAL COMPONENTS ANALYSIS ON DISTANCE
     1FUNCTION MATRIX//)
      RETURN
      END
```

```
      OVERLAY(JPCL,6,0)
      PROGRAM CORSYM
      DIMENSION RTT(200),VEE(200),R(20100),X(200),TT(200),T(200)
      DIMENSION EM(200)
      COMMON/BLOCKA/M,N,MP,NN,IECK,JPCL,IP,CHAR(16),ECH,ECD,EPSI,ECK,IFM
     1,JKL,BEND,IMT(8),I
      DIMENSION C(1001,2)
      IF(JKL.LE.0) JKL=0
      IF(I.LE.0) GO TO 46
      IF(I.EQ.1) GO TO 47
      IF(I.EQ.3) GO TO 48
      IF(I.EQ.10) GO TO 49
      IF(I.EQ.5) GO TO 2038
   47 BEND=(1H9)
      GO TO 50
   48 BEND=(3H999)
      GO TO 50
 2038 BEND=(5H99999)
      GO TO 50
   49 BEND=(10H9999999999)
      GO TO 50
   46 BEND=(6H999999)
   50 CONTINUE
      READ(3,103) (IMT(I),I=1,8)
  103 FORMAT(8A10)
      IF(IFM.LE.0) IFM=0
      IF(EPSI.LE.0.0)EPSI=0.005
      IF(ECK.LE.0.0)ECK=5.0
      IF(ECD.LE.0.0) ECD=80.0
      DO 203 J=1,M
      TT(J)=999999.
  203 T(J)=-999999.
      DO 291 J=1,1001
      READ(3,IMT) C(J,1),C(J,2),(X(I),I=1,M)
      IF(C(J,1).EQ.BEND) GO TO 299
      DO 292 K=1,M
      IF(X(K).LT.TT(K))TT(K)=X(K)
      IF(X(K).GT.T(K))T(K)=X(K)
  292 CONTINUE
  291 WRITE(8) (X(I),I=1,M)
  299 N=J-1
      END FILE 8
      REWIND 8
      DO 562 I=1,N
  562 WRITE(6) C(I,1),C(I,2)
      DO 293 K=1,M
  293 T(K)=T(K)-TT(K)
      DO 294 J=1,N
      READ(8) (X(I),I=1,M)
      DO 295 K=1,M
      X(K)=(X(K)-TT(K))/T(K)
  295 CONTINUE
  294 WRITE(6) (X(I),I=1,M)
      END FILE 6
      REWIND 6
      REWIND 8
      WRITE(4,10)(CHAR(I),I=1,16)
```

```
   10 FORMAT(1H1,16A5)
      DO 200 J=1,M
  200 EM(J)=0.0
      DO 393 K=1,M
      RTT(K)=0.0
  393 VEE(K)=0.0
      NNN=(M*M+M)/2
      DO 310 J=1,NNN
  310 R(J)=0.0
      EMM=M
      DO 561 I=1,N
  561 READ(6) Q1,Q2
      DO 394 J=1,N
      READ(6) (X(I),I=1,M)
      IJ=1
      DO 395 K=1,M
      EM(K)=EM(K)+1.0
      RTT(K)=RTT(K)+X(K)
      VEE(K)=VEE(K)+X(K)*X(K)
  395 CONTINUE
      DO 297 I=1,M
      DO 297 K=I,M
      R(IJ)=R(IJ)+X(I)*X(K)
  297 IJ=IJ+1
  394 CONTINUE
      REWIND 6
      IJ=1
      DO 312 J=1,M
      R(IJ)=1.0
      IJ=IJ+1
      KJ=J+1
      IF(KJ.GT.M) GO TO 312
      DO 312 K=KJ,M
      EMR=EM(J)*R(IJ)-(RTT(K)*RTT(J))
      EMVJ=EM(J)*VEE(J)-(RTT(J)*RTT(J))
      EMVK=EM(K)*VEE(K)-(RTT(K)*RTT(K))
      R(IJ)=EMR/SQRT(EMVJ*EMVK)
  312 IJ=IJ+1
      WRITE(8) (R(J),J=1,NNN)
      END FILE 8
      REWIND 8
      WRITE(5) (R(J),J=1,NNN)
      END FILE 5
      REWIND 5
      WRITE(4,300)
  300 FORMAT(1H0,5X,70HR-MODE PRINCIPAL COMPONENTS ANALYSIS ON CORRELATI
     1ON COEFFICIENT MATRIX//)
      RETURN
      END
C     INSERT REQUIRED END OF FILE CARD HERE
```

```
IMBRIE PURDY 12 VARIABLE BAHAMAS SEDIMENT PROBLEM
 12
(2A6,10F6.0/(12X,10F6.0))
            1   2.0   3.4  10.0   8.8   3.2   2.2   1.0   3.8  14.6   0.4
            1  24.8  23.1
           31   0.0   0.6   0.6   0.2   0.0   0.4   3.6   3.0  12.8  59.0
           31  17.8   7.0
           43   0.2   1.0   0.4   1.4   0.0   1.8   7.4   1.8  25.8  38.8
           43  16.2   6.5
           59   0.0   1.6   0.8   2.8   0.4   3.8   3.4   1.4  44.2  22.0
           59  12.4   2.5
          171   0.0   5.2   4.8   3.0   0.4   9.8  61.0   6.8   0.0   2.6
          171   0.6  31.3
          174   0.0  11.2   5.8   1.2   0.0  14.2  47.6   6.2   0.4   0.6
          174   3.0  40.1
          380   0.0  16.2   4.2   1.8   0.0  10.4  47.0   3.4   0.4   3.0
          380   1.8  51.9
          360   0.0   0.2   2.4   1.8   0.0   4.8  68.8   8.4   0.0   4.2
          360   3.0  58.8
          366   0.0   0.8   3.4   1.2   0.0   3.0  76.8   8.4   0.0   2.8
          366   0.4  35.2
          367   0.0   0.2   2.8   1.2   0.0   2.2  61.8   4.2   0.0  20.0
          367   3.2  31.8
          368   0.0   2.0   6.2   2.0   0.0   4.2  67.4   8.8   0.0   4.2
          368   0.4  51.9
          369   0.0   3.0   4.2   2.6   0.0   5.0  66.8   7.4   0.0   3.2
          369   1.4  63.7
          370   0.0   3.4  10.0   5.4   0.2   5.4  43.0  19.6   0.0   1.6
          370   1.2  90.7
          371   0.0   5.4  15.2   4.2   0.0  10.4  32.0  17.0   0.2   0.4
          371   2.4  87.2
          372   0.0  10.2  12.8   7.2   0.0  11.8  22.4  16.8   0.0   1.2
          372   2.0  85.7
          373   0.0  25.6   1.8   6.4   0.0  12.2  25.8   7.4   0.0   5.2
          373   0.8  89.2
          374   0.0  13.8   0.2   3.6   0.0   2.8  22.0   9.2   0.2  32.8
          374   9.4   5.0
          207   3.4  12.6   1.4   1.8   7.6   8.0   4.2   8.4   4.8   8.4
          207  11.2   2.2
          328   0.0   0.4   0.2   0.6   0.0   1.6  18.6   3.6  35.0  16.0
          328  20.2   7.6
          330   0.0   1.0   0.8   0.2   0.0   1.2  26.2   4.0  10.6  44.6
          330   9.4   5.6
          334   0.0   0.8   0.0   0.0   0.0   1.2  20.0   1.8   4.8  63.8
          334   6.2   4.3
          310   0.2   2.8   0.0   0.6   0.0   1.6  11.6   2.6   3.2  36.4
          310  37.6   8.3
          312   0.0   1.0   0.0   0.4   0.0   2.8  25.6   2.4  12.8  36.8
          312  15.0   5.5
          502   0.0   3.6   0.6   2.8   0.0   9.0   4.4   1.8  14.4  13.6
          502  45.6   4.4
          504   3.6  26.4   1.0   1.8  12.8  12.0   0.2  10.4   2.8   0.0
          504   7.8   2.4
          279   2.0  15.8   2.8   5.4   6.2   5.2   1.8   6.4   1.8   1.4
          279  35.2  18.9
          512   2.8   1.6   0.2   0.2   0.6   4.0   6.4   1.4  38.4  18.0
          512  22.2   2.5
          514   0.0   1.0   2.2   2.6   0.0   1.4   3.2   3.4  49.0   2.2
          514  32.0   4.5
          317   0.0   0.8   0.4   1.8   0.0   1.8   3.0   1.2  48.6   7.8
          317  30.6   4.6
          537   3.8   9.4   1.6   1.2   6.8  10.4   2.2   1.8   3.2   4.4
          537  17.2   2.0
           37   0.2  39.8   2.0   3.2   1.0  10.4   0.8   2.6   3.2   0.0
           37  19.8   4.1
         7-47   2.8  14.2   7.2   2.2  13.8   6.8   0.0   3.8   5.6   0.0
         7-47  27.2   2.5
          409   0.0   5.2   5.5   0.7   0.8   5.1   0.6   1.2  36.7  24.9
          409  15.8   2.9
          411   0.0   0.9   0.8   0.5   0.0   1.6   0.5   1.6  11.4  68.8
          411  11.9   5.9
          418   0.0   0.0   0.0   0.0   0.0   0.3   0.0   0.0   0.3  96.8
          418   2.1   1.4
          417   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0  98.2
          417   1.4   1.3
          414   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0  97.6
          414   2.3   2.3
          413   0.0   0.0   0.0   0.0   0.0   0.0   0.1   0.2   0.3  96.4
          413   2.9   1.5
          438   0.0   0.0   0.0   0.0   0.0   0.1   0.0   0.0   0.5  97.9
          438   1.3   1.3
          250   1.5  22.4   0.6   0.9  13.5  10.6   2.1   3.3  14.2  13.1
          250  11.7   0.6
9999999999

FINISH
```

IMBRIE PURDY 12 VARIABLE BAHAMAS SEDIMENT PROBLEM

R-MODE PRINCIPAL COMPONENTS ANALYSIS ON DISTANCE FUNCTION MATRIX

PRINCIPAL FACTOR ANALYSIS OF VARIABLES

```
FACTOR    1
          NUMBER OF ITERATIONS EQUALS        0
          VARIANCE EXTRACTED EQUALS      7.767

FACTOR    2
          NUMBER OF ITERATIONS EQUALS        5
          VARIANCE EXTRACTED EQUALS       .992

FACTOR    3
          NUMBER OF ITERATIONS EQUALS        5
          VARIANCE EXTRACTED EQUALS       .691

FACTOR    4
          NUMBER OF ITERATIONS EQUALS        5
          VARIANCE EXTRACTED EQUALS       .522
```

PRINCIPAL AXIS FACTOR LOADINGS

| | F 1 | F 2 | F 3 | F 4 | F 5 | F 6 | F 7 | F 8 | F 9 | F 10 | HSQ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .790 | -.346 | -.297 | .134 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .850 |
| 2 | .856 | -.107 | -.198 | .084 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .790 |
| 3 | .868 | .222 | -.001 | -.047 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .805 |
| 4 | .873 | .134 | -.025 | -.122 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .796 |
| 5 | .807 | -.320 | -.302 | .174 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .875 |
| 6 | .808 | .122 | -.210 | -.008 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .711 |
| 7 | .763 | .360 | .161 | .066 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .743 |
| 8 | .876 | .278 | -.010 | .002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .845 |
| 9 | .736 | -.327 | .294 | -.387 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .884 |
| 10 | .646 | -.238 | .540 | .461 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .977 |
| 11 | .786 | -.338 | .143 | -.290 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .836 |
| 12 | .816 | .438 | .060 | -.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | .860 |
| | 7.767 | .992 | .691 | .522 | | | | | | | |
| | 64.723 | 8.269 | 5.761 | 4.353 | | | | | | | |

SQUARE ROOT OF FACTOR MATRIX

| | F 1 | F 2 | F 3 | F 4 | F 5 | F 6 | F 7 | F 8 | F 9 | F 10 | HSQ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .790 | -.346 | -.297 | .134 | | | | | | | |
| 2 | .425 | .187 | .012 | -.012 | | | | | | | |
| 3 | .330 | .518 | .249 | -.175 | | | | | | | |
| 4 | .148 | .139 | .111 | -.244 | | | | | | | |
| 5 | .071 | -.225 | -.150 | .148 | | | | | | | |
| 6 | .107 | .198 | -.157 | -.033 | | | | | | | |
| 7 | .131 | .343 | .315 | .134 | | | | | | | |
| 8 | .072 | .205 | .025 | .042 | | | | | | | |
| 9 | .114 | -.406 | .521 | -.558 | | | | | | | |
| 10 | .074 | -.215 | .634 | .654 | | | | | | | |
| 11 | .056 | -.220 | .120 | -.331 | | | | | | | |
| 12 | .053 | .222 | .044 | -.002 | | | | | | | |

BETA COEFFICIENTS

| | F 1 | F 2 | F 3 | F 4 | F 5 | F 6 | F 7 | F 8 | F 9 | F 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .10176 | -.34840 | -.43015 | .25571 | | | | | | |
| 2 | .11016 | -.10735 | -.28663 | .16110 | | | | | | |
| 3 | .11176 | .22362 | -.00203 | -.08911 | | | | | | |
| 4 | .11246 | .13469 | -.03609 | -.23297 | | | | | | |
| 5 | .10385 | -.32290 | -.43695 | .33255 | | | | | | |
| 6 | .10398 | .12255 | -.30307 | -.01593 | | | | | | |
| 7 | .09830 | .36280 | .23283 | .12727 | | | | | | |
| 8 | .11275 | .28054 | -.01485 | .00379 | | | | | | |
| 9 | .09475 | -.32950 | .42489 | -.74017 | | | | | | |
| 10 | .08317 | -.24001 | .78076 | .88174 | | | | | | |
| 11 | .10115 | -.34112 | .20721 | -.55536 | | | | | | |
| 12 | .10501 | .44109 | .08729 | -.00435 | | | | | | |

NORMALIZED FACTOR MEASUREMENTS, UNWEIGHTED

| | | | | |
|---|---|---|---|---|
| 1 | .75616 | .30843 | .52038 | .29154 |
| 2 | .15399 | .21171 | .93730 | .63570 |
| 3 | .22704 | .20097 | .88552 | .45579 |
| 4 | .36719 | .17553 | .86054 | .24539 |
| 5 | .52340 | .74070 | .58632 | .59276 |
| 6 | .57811 | .72618 | .51119 | .58071 |
| 7 | .52883 | .70276 | .54526 | .60614 |
| 8 | .45920 | .79477 | .70302 | .60171 |
| 9 | .37782 | .75831 | .71040 | .62522 |
| 10 | .29122 | .63468 | .77313 | .67227 |
| 11 | .49918 | .81340 | .69070 | .60679 |
| 12 | .51701 | .82097 | .68112 | .59504 |
| 13 | .86268 | 1.00000 | .63169 | .52074 |
| 14 | .93590 | .99274 | .55788 | .49726 |
| 15 | 1.00000 | .96888 | .50586 | .47375 |
| 16 | .79603 | .79613 | .47208 | .57214 |
| 17 | .34820 | .43672 | .70377 | .66160 |
| 18 | .64220 | .13686 | .24255 | .71115 |
| 19 | .25686 | .22888 | .87804 | .27676 |
| 20 | .16805 | .34166 | .87448 | .65734 |
| 21 | .09319 | .30581 | .91031 | .80253 |
| 22 | .23737 | .20741 | .83131 | .50208 |
| 23 | .18717 | .30916 | .84807 | .57167 |
| 24 | .43001 | .19533 | .72000 | .22698 |
| 25 | .86099 | .09384 | 0.00000 | .80025 |
| 26 | .74023 | .19291 | .37420 | .45138 |
| 27 | .42324 | 0.00000 | .68812 | .34367 |
| 28 | .36708 | .12796 | .88190 | 0.00000 |
| 29 | .31261 | .09107 | .89800 | .05032 |
| 30 | .58224 | .06842 | .20999 | .67053 |
| 31 | .53701 | .30290 | .37901 | .48141 |
| 32 | .80616 | .02521 | .18647 | .60407 |
| 33 | .35567 | .19351 | .81420 | .31541 |
| 34 | .14049 | .21556 | .93745 | .71712 |
| 35 | .00738 | .21813 | .99403 | .98731 |
| 36 | 0.00000 | .21857 | .99984 | 1.00000 |
| 37 | .00602 | .21833 | 1.00000 | .99196 |
| 38 | .00884 | .21602 | .99759 | .98102 |
| 39 | .00275 | .21801 | .99952 | .99555 |
| 40 | .66854 | .04746 | .25410 | .69617 |

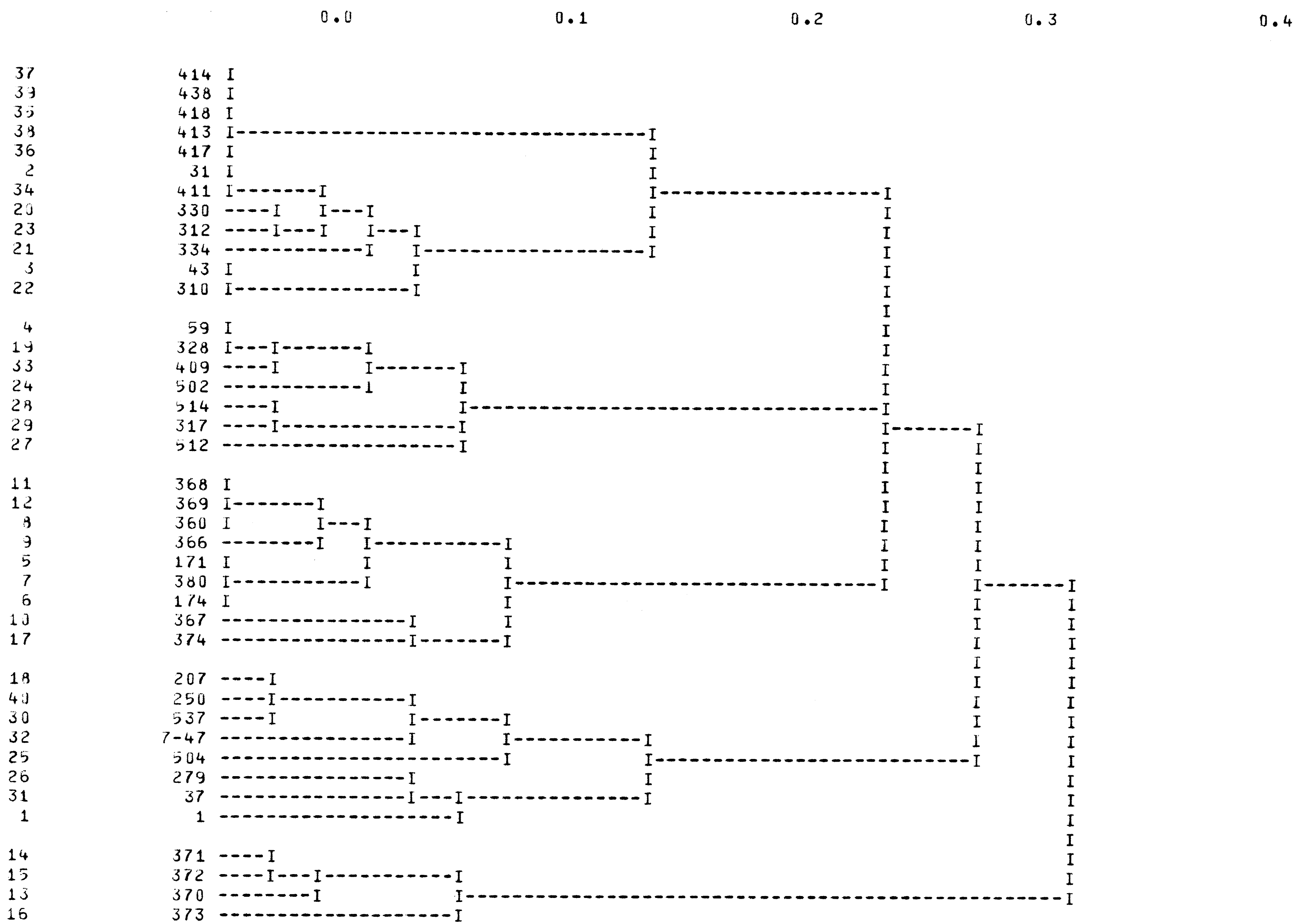
IMBRIE PURDY 12 VARIABLE BAHAMAS SEDIMENT PROBLEM
40 SAMPLES

## CLUSTER TYPE 6 ANALYSIS

| LEFT SAMPLE NUMBER | RIGHT SAMPLE NUMBER | DISTANCE FUNCTION | NO. IN LEFT | NO. IN RIGHT |
|---|---|---|---|---|
| 37 | 39 | .0024445 | 1 | 1 |
| 35 | 38 | .0038323 | 1 | 1 |
| 35 | 37 | .0055372 | 2 | 2 |
| 35 | 36 | .0064430 | 4 | 1 |
| 11 | 12 | .0122984 | 1 | 1 |
| 11 | 8 | .0282201 | 2 | 1 |
| 5 | 7 | .0238663 | 1 | 1 |
| 3 | 22 | .0361593 | 1 | 1 |
| 4 | 19 | .0408324 | 1 | 1 |
| 2 | 34 | .0413095 | 1 | 1 |
| 14 | 15 | .0445466 | 1 | 1 |
| 5 | 6 | .0389097 | 2 | 1 |
| 18 | 40 | .0475457 | 1 | 1 |
| 20 | 23 | .0486310 | 1 | 1 |
| 28 | 29 | .0492440 | 1 | 1 |
| 18 | 30 | .0460419 | 2 | 1 |
| 4 | 33 | .0535876 | 2 | 1 |
| 11 | 9 | .0643489 | 3 | 1 |
| 2 | 20 | .0758914 | 2 | 2 |
| 14 | 13 | .0752575 | 2 | 1 |
| 2 | 21 | .0878980 | 4 | 1 |
| 11 | 5 | .0924154 | 4 | 3 |
| 4 | 24 | .0937014 | 3 | 1 |
| 16 | 17 | .1088090 | 1 | 1 |
| 2 | 3 | .1154333 | 5 | 2 |
| 18 | 32 | .1053786 | 3 | 1 |
| 26 | 31 | .1165346 | 1 | 1 |
| 26 | 1 | .1310782 | 2 | 1 |
| 14 | 16 | .1318150 | 3 | 1 |
| 4 | 28 | .1334705 | 4 | 2 |
| 4 | 27 | .1447120 | 6 | 1 |
| 11 | 10 | .1580452 | 7 | 2 |
| 18 | 25 | .1596563 | 4 | 1 |
| 18 | 26 | .2134851 | 5 | 3 |
| 2 | 35 | .2114926 | 7 | 5 |
| 2 | 4 | .3185507 | 12 | 7 |
| 2 | 11 | .3110394 | 19 | 9 |
| 2 | 18 | .3615795 | 28 | 8 |
| 2 | 14 | .4150381 | 36 | 4 |

CLUSTER DIAGRAM

VALUE OF COEFFICIENT



KANSAS GEOLOGICAL SURVEY COMPUTER PROGRAM
THE UNIVERSITY OF KANSAS, LAWRENCE

PROGRAM ABSTRACT

Title (If subroutine state in title):

FORTRAN IV program for Q-mode cluster analysis on distance function with printed dendrogram

Date: 20 August 1969

Author, organization: James M. Parks, Center for Marine and Environmental Studies and Department of Geological Sciences, Lehigh University, Bethlehem, Pennsylvania 18015

Direct inquiries to: Operations Research Section, Kansas Geological Survey or

Name: James M. Parks Address: as above

Purpose/description: Multivariate Q-mode cluster analysis of samples, OTU's or locations, with computer line printout of resulting two-dimensional hierarchical cluster diagram.

R-mode principal components analysis on normalized distance function (Householder's method); calculates component measurements for each sample (by short regression and square root method of Harman); hierarchical Q-mode cluster analysis on normalized distance function.

Restrictions, range: Maxima of 1000 samples, 200 raw variables, 10 principal components.

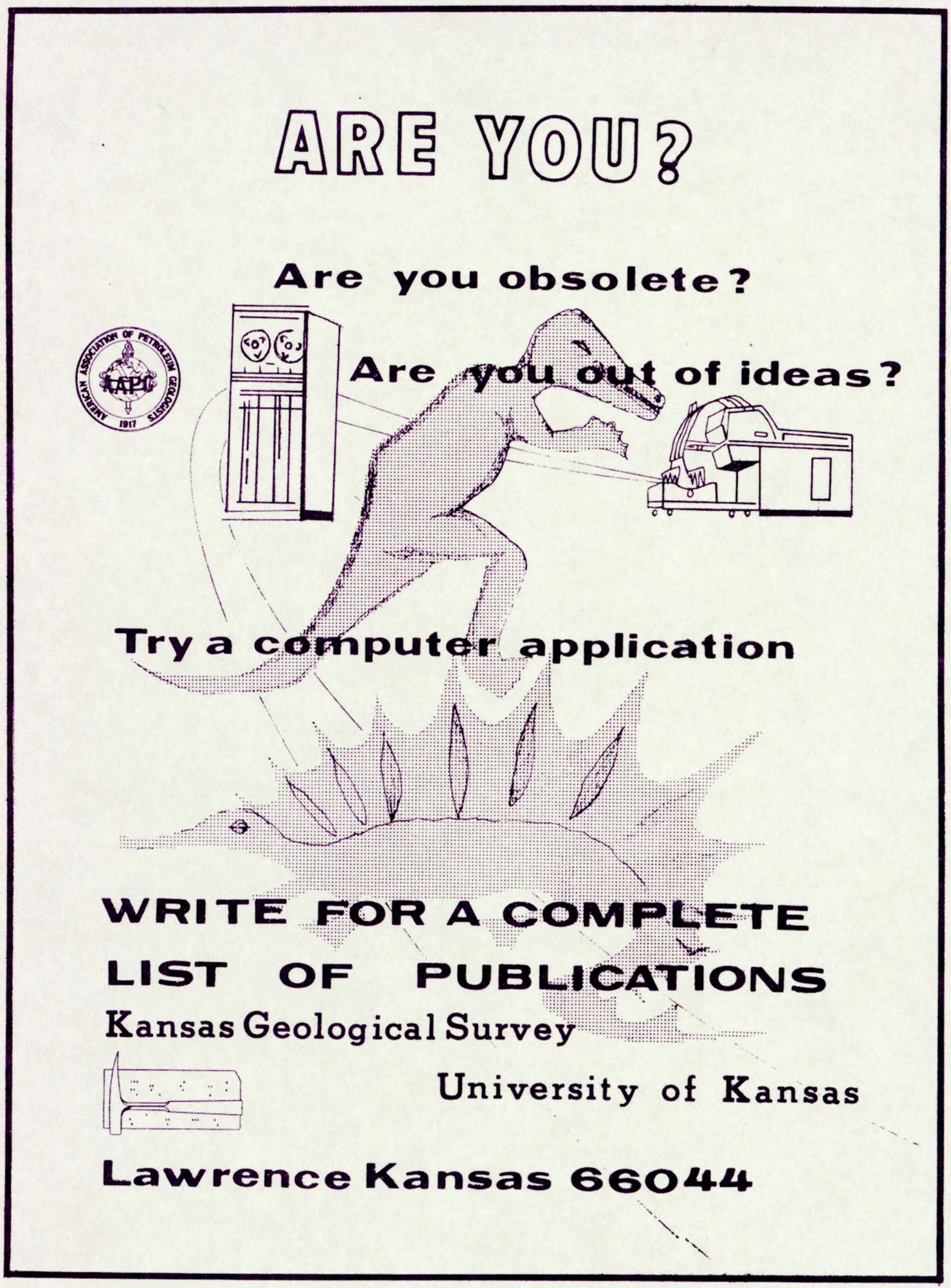
Computer manufacturer: CDC Model: 6400

Programming language: FORTRAN IV

Memory required: 65 K Approximate running time:

Special peripheral equipment required: 4 magnetic tapes and line printer

Remarks (special compilers or operating systems, required word lengths, number of successful runs, other machine versions, additional information useful for operation or modification of program) Example problem (40 samples, 12 raw variables), using 4 principal factors accounting for 83.2 percent of total variance, ran in 5.9 seconds CP and 98.9 seconds PP on CDC 6400. A larger problem (186 samples, 17 raw variables), using 3 principal factors accounting for 81.9 percent of total variance, ran in 77 seconds CP, 265 seconds PP.


ARE YOU?
Are you obsolete?
Are you out of ideas?
Try a computer application
WRITE FOR A COMPLETE
LIST OF PUBLICATIONS
Kansas Geological Survey
University of Kansas
Lawrence Kansas 66044