Rerum Naturalium Fragmenta

No. 358

Glasgow
1983

.

Rerum Naturalium Fragmenta
Tamas Jasko editor
34 Lomondside Ave, Clarkston, Glasgow G76 7UJ, Scotland

CHART: Plotting of Palynology Charts

by T. Jasko

1. Purpose

This program produces plots of species occurrence data for a series of samples. Two different versions are available for palynological and micro-palaeontological data. Also, the palynology plots can be drawn with or without the palynofacies columns.

2. Input data

Sample data are read from a file in the format produced by CHIN (or MPIN). Species names/codes are read from the same file as for CHIN/MPIN. For details, see CHIN description.

3. Procedure

To generate palynology charts, type in TSM>

```
CHART filename
```

where filename = the file containing the sample data

To produce micro-palaeontology charts, type in TSM>

```
MPLOT filename
```

If no filename is specified data are read from (default) files CH-1 or MP-1.

4. Plot Parameters

Enter these as prompted by the program. In each case, pressing Enter (CR=carriage return) alone will indicate that the default value is to be used.

4.1 Author

Enter name to be displayed, up to 10 characters. Default=blank.

4.2 Depth unit

Enter M or FT as wanted (2 characters). Default=FT.

4.3 Depth Interval selection

The specification has two purposes: to select data from the input file i.e. samples outside the range specified will be ignored; and to control the range of depth displayed.

4.31 Top depth

Enter number with or without decimal point. Default=0.0

4.32 Bottom depth

Enter number as for top. Default = top depth + 1000.

4.4 Number of fossils plotted

Enter an integer number. Default = 100 (minimum). This parameter, if entered, will allow plotting more than 100 species (up to 500) by adding extra columns to the plot. Each additional column will increase the 'Standard' plot dimensions by 3.3 mm in the X (horizontal) direction.

4.5 Depth scale extension

Enter integer number of 15 mm subdivisions wanted. Default = 10 (minimum). If entered this parameter allows stretching of the plot in the Y (vertical) direction. Useful, if the plot would be otherwise crowded by too many samples.

4.6 Plot destination

Enter 1 to plot at user terminal (must be a 'graphic' terminal e.g. Westward or Tektronix). Enter 2 for main plotter (default). Not yet available: film plotting (=3), colour terminal (=6).

If the plot is drawn on the terminal, it will be scaled in both directions to fit the screen and the effect of defining the number of fossil species and depth scaling will affect only the relative size of heading space etc. Because of limited resolution the matrix is displayed without the fine (dotted) net.

4.61 Absolute sizing

If destination is the main plotter, absolute size of the plot can be changed by entering the maximum horizontal size as wanted. The plot will be produced with the proportions kept as defined by previously entered parameters. Default: Standard size.

5. Output

Besides the plot, a record of all fossil codes and their display columns is output to a temporary file. These records are cumulated in a permanent file and can be used to produce cross-reference listings for a set of plots.

6. I/O Channels and files used

LO	PRINT messages/prompts to user terminal
6	WRITE error messages (from graphic routines)
1	input data file
PL	input last plot number (file PLAST)
PP	Output incremented plot number (file SPAR)
BO	PUNCH cross-reference record
SL	input species name/code list (file BUGS for palynology, file MICRO for micro-palaeontology)
UT	ACCEPT Parameters from terminal

7. Associated program modules

Catalogued modules used: GH.ART, MP.ART, EDIT

Fortran source:
- specific routines (In Fortran 77): CH:xxxxx and MP:xxxxx
- graphic routines (DI-3000 system): Jxxxxx
- general routines:
    OPTION = logical function to get program options

8. Program options

Options 1 & 2 are used by TSM macros to define plot version.
Option 7 : Print sample data for checking.
Options 11 & 12 are used to debug graphics.

[Software Update no.22: 4 May 1983]

DISCRIM: Discriminant Functions

by T. Jasko

DISCRIM computes the linear discriminant function that separates two groups. Group A is an n1 by m data matrix where n1 is the number of observations and m is the number of variables.

Group B is an n2 by m data matrix. Here n1 and n2 are not necessarily equal. As a third group, a number of 'unknown' samples can be entered for classification.

If the data are entered in Ratiplot G format, all numeric variables are used in computing the discriminant function; while variables designated as alphabetic (using the / Separator) will be ignored.

Access is through the CRISTAN menu (option DISCRIM) or directly from TSM> by typing

```
DISCRIM
```

Program options (parameters) are input interactively (I/O unit 5).

Data are read from files (I/O unit IN).

After performing the required function the program returns to the main CRISTAN menu.

Input:

- two data files containing the two populations
- data file of 'unknown' samples (optional)

Output:

- values of r0, r1, r2
- linear discriminant function
- statistics for testing the significance of the function
- Mahalanobis distance
- F statistic (see table for critical values in Penguin-Honeywell*)
- discriminant scores for each group

*Critical Values of F for n1, and n2 Degrees of Freedom and 5% (a = 0.05) or 1% (a = 0.01) Level of Significance can be found in Table 22 of *The Penguin-Honeywell Book of Tables*. F. W. Kellaway (ed.). Honeywell Controls Ltd. (E.D.P. Division). 1968

(Software Update no.31: 12 January 1983)

FACTOR: Principal Component and Factor Analysis

by

T. Jasko

This set of statistical routines calculates principal components and performs factor analysis.

Access is through the CRISTAN menu (option FACTOR) or directly from TSM> by typing

```
FACTOR
```

To select a particular function, type its name. Function names can be abbreviated to 3 characters e.g. FAC for FACTOR ANALYSIS.

Program options (parameters) are Input Interactively (I/O unit 5). Data are read from files (I/O unit IN).

After performing the required function the program returns to the main CRISTAN menu.

PCA - Principal Component Analysis

The program accepts an n by m data matrix where n is the number of observations and m is the number of variables. If the first option is 1, an m by m matrix of covariances between columns will be computed. If this option is 2, an n by matrix of covariances between rows will be computed. If the option is 0, the program calls exit, as the program loops back and restarts after completion of an analysis.

If the second option is 1, a standardized covariance (correlation) matrix is created. If this option is 2, a raw covariance matrix is created.

Input:

- transp, similm (2 numbers separated by comma):
    - transp:
        - 0 = end of job
        - 1 = do not transpose input data matrix
        - 2 = transpose input data matrix
    - similm:
        - 1 = calculate correlation matrix
        - 2 = calculate covariance matrix
- Input data file

Procedure:

1. If correlation matrix is to be calculated, standardize the input data matrix and print the standardized data matrix
2. Transpose input data matrix (if required)
3. Calculate and print similarity matrix
4. Calculate eigenvalues and eigenvectors
5. Move eigenvalues to first column
6. Calculate sum of eigenvalues
7. Calculate percent contribution of each eigenvalue

Output:

- eigenvalues and percent contribution
- eigenvectors (note that eigenvectors are stored column-wise)
- component scores

FACTOR ANALYSIS

The program accepts an n by n data matrix where n is the number of observations and m is the number of variables. If the first option is 1, an m by m similarity matrix between columns will be computed. If this option is 2, an n by m similarity matrix between rows will be computed. If the Option is 0, the program calls exit, as the program loops back and restarts after completion of an analysis.

If the second option is 1, a standardized covariance (correlation) matrix is created. If this option is 2, a cosine theta similarity matrix is created. The third option specifies the number of factors to be retained.

To perform R-mode analysis, select 1 both on option one and on option two.

To perform Q-mode factor analysis, select 2 on both option one and on option two.

Input:

- transp, similm, nufact (3 numbers separated by comma):
    - transp:
        - 0 = end of job
        - 1 = do not transpose data matrix
        - 2 = transpose data matrix
    - similm:
        - 1 = calculate correlation matrix
        - 2 = calculate cosine theta matrix
    - nufact: number of factors desired

- input data file

Procedure:

- If correlation matrix is to be calculated standardise input data matrix and print standardised data
- Transpose data matrix (if required)
- Calculate similarity matrix. Save correlation matrix
- Calculate eigenvalues and eigenvectors. Move eigenvalues to first column
- Calculate sum of eigenvalues. Calculate percent contribution of each eigenvalue
- Calculate factor loadings
- Calculate reproduced correlation matrix and residual correlation matrix
- Calculate factor scores. Rotate factor matrix
- Calculate varimax factor scores

Output:

- similarity matrix
- eigenvalues and percent contribution
- eigenvectors (Note: eigenvectors are stored column-wise)
- factor loadings
- reproduced correlation matrix and residual correlation matrix
- factor scores
- rotated factor matrix
- varimax factor scores

(Software Update no.33: 30 November 1983)

TREND : Trend Surface Analysis

by T. Jasko

This set of statistical routines calculates grids and contours and performs trend surface analysis and double Fourier transformation.

Access is through the CRISTAN menu (option TREND) or directly from TSM> by typing

```
TREND
```

To select a particular function, type its name. Function names can be abbreviated to 3 characters e.g. GRI for GRIDDING.

Program options (parameters) are input interactively (I/O unit 5). Data are read from files (I/O unit IN). After performing the required function the program returns to the main CRISTAN menu.

GRIDDING

This computes a regular grid of values from irregularly spaced map data. The matrix of grid values is printed out and also plotted as a line printer contour map by subroutine PLOT.

Parameters are first read that control the size of the finished map.

Next the data are read in as an n by 3 matrix, where n is the number of observations. The first column contains x1 (east-west or across the map) co-ordinate, the second column contains x2 (north~south or down the map), and the third column contains the dependent variable. The map as produced by plot will have 9 equally spaced contour bands scaled between the maximum and minimum values of y.

Input:

1. control parameters: width, left, right, bottom, top
    - width = width of map in inches
    - left = xl value of left edge of map
    - right = xl value of right edge of map
    - bottom = x2 value of bottom edge of map
    - top = x2 value of top edge of map
2. data file

Procedure:

For each grid point calculate the squared distance between current grid point and all data points. Find the 6 nearest data points and calculate sums. Calculate value at grid point and store in matrix.

Output:

- grid values
- map

TREND SURFACE ANALYSIS

The program first reads a control card specifying desired degree and various map parameters.

Input data matrix is n by 3, where n is the number of observations. The first column contains xl (east-west or across the map) coordinate, the second column contains x2 (north-south or down the map), and the third column contains the dependent variable.

Input:

Parameters: order, width, refc, intv, left, right, bottom, top
- order = order of trend surface
- width = width of map in inches
- refc = value of reference ($) contour
- intv = contour interval
- left = xl value at left edge of map
- right = xl value at right edge of map
- bottom = x2 value at bottom edge of map
- top = x2 value at top edge of map

Data file

Output:

- trend surface coefficients
- error measures
- trend map

DBL Double Fourier Transformation

This program calculates the double Fourier transform of a data matrix, which has nr rows and mc columns representing measurements made at the nodes of an nr by mc grid.

Input:

1. - input data matrix

2. - After the data are read in, 4 map control parameters are requested:

   (1) nh, the number of harmonics to be used in map (maximum 100)
   (2) width of map in inches (up to 10 inches)
   (3) value of reference ($) contour
   (4) contour interval

3. - Next follow nh pairs of numbers, which give the subscripts of harmonics to be used in map.

Output:

printout of coefficients and power spectrum
calculate and print map one line at a time.

(Software Update no.34: 7 December 1983)

CRISTAN Statistical Analysis Package

by T. Jasko

The statistical methods used in this program system are described in detail by J.C. Davis (*Statistics and Data Analysis In Geology)*

A set of Davis's routines were adapted for interactive use on the SEL by L. Holman; at subsequent stages routines for multivariate analysis were added and the system was enhanced by graphic output routines and additional data formats.

To access the system, sign on to the terminal by typing your username & key and in TSM> type

```
CRISTAN
```

The CRISTAN menu is displayed:

- RATIPLOT for Ratio definition & Plots
    - X/Y scatter plots
    - histograms
    - triangle diagrams
- ANOVA for Mean, Variance and Regression
- DISCRIM for Discriminant Functions
- CLUSTER for Cluster Analysis/Dendrogram
- FACTOR for Principal Component and Factor analysis
- MERGE for Merging Files and Selecting Data
- TIMESER for Time Series analysis
- TREND for Trend Surface Analysis

Select a menu item by typing its name (or X to exit).

Programs producing graphic output can be best run on graphic terminals but the graphic output can be diverted to the plotter. Therefore, all programs can be run on any terminal.
[Software Update no.35: 1 December 1983]

RENATA: Transfer of Files from one Username to Another

by T. Jasko

Using the Editor, create a job stream as follows:

```
$JOB TRANSFER
$RENATA
username1
userkey1
username2
userkey2
file1  newname1
file2  newname2
file3 newname3
        .... and so on
$EOJ
```

Then run it. file1, file2, file3,... will be transferred from username1 to username2 (and renamed to the new name) unless any of the following errors is found:

- invalid username or key specified
- no such file found under username1
- there is such a file already under username2
- file is busy

Errors are indicated by a warning message - file left as is.

If used on line, the end of the job can be indicated by entering X.

(Software Update no.37: 12 December 1983)