CABIOS
Vol.5. no.4. 1989
Pages 287–292

Statistical analysis and graphical display of multivariate data on the Macintosh

J. Thioulouse

Abstract

Two Macintosh programs written for multivariate data analysis and multivariate data graphical display are presented. MacMul includes principal component analysis (PCA), correspondence analysis (CA) and multiple correspondence analysis (MCA), with a complete, original and unified set of numerical aids to interpretation. GraphMu is designed for drawing collections of elementary graphics (curves, maps, graphical models) thus allowing comparisons between variables, individuals, and principal axes planes of multivariate methods. Both programs are self-documented applications and make full use of the user-oriented graphical interface of the Macintosh to simplify the process of analysing data sets. An example is described to show the results obtained on a small ecological data set.

Introduction

Multivariate data analysis methods in biological sciences have gained more and more popularity over the last 20 years. The great improvements in computing facilities have allowed an increasing number of research laboratories to make use of statistical software. However, until the development of powerful micro-computers, these statistical packages were restricted to people having access to mainframe computers, and they had only very poor graphical outputs. Nowadays many statistical packages have been written for or adapted to the IBM PC or compatible micro-computers (e.g. BMDP, SAS-PC, SPSS/PC+, SYSTAT, PCSM, etc.). Several statistical packages are also available on the Apple Macintosh, such as SYSTAT, PCSM, STATWORK or STATVIEW. The friendly, user-oriented graphical interface of the Macintosh makes it easier for biologists to use intricate statistical methods. Nevertheless, few packages really deal with complete multivariate data analysis methods (STATVIEW for principal component analysis (PCA), PCSM for PCA, correspondence analysis (CA) and multiple linear regressions, and SYSTAT for discriminant analysis and canonical correlation analysis). Moreover, in spite of the great deal of graphical display software available on the Macintosh, there is a lack of *multivariate data* graphical display programs. Two main features are required in this area: (i) labelled scatter diagrams for the display of principal axes planes and (ii) collection and superimposition of elementary graphics (including principal axes planes) for the comparison of variables, sets of individuals or principal axes planes.

Here we present two Macintosh programs. The first one (MacMul) is intended to make it easier to use three multivariate methods: PCA for quantitative variables, CA for counts tables, and multiple corespondence analysis (MCA) for qualitative variables. These methods are described in detail in the book by Lebart *et al.* (1984). CA has been particularly used in the scope of molecular biology (see for example Grantham *et al.*, 1980; Fichant and Gautier, 1987; Quentin, 1988, 1989). The second one (GraphMu) offers several facilities for multivariate data graphical display and for graphical output of multivariate data analysis methods. Until now these two programs have a French interface language but we are working on an English version.

A small data set was used to show the potential of both programs. The complete analysis was described by Thioulouse and Chessel (1987) and Doledec and Chessel (1987). Further biological discussion may be found in Bournaud *et al.* (1983). The data set consists of nine physico-chemical variables measured four times (February, June, August, November) in five sampling sites (A, B, C, D, E) distributed along a small French stream (Table I presents the data set, each variable being standardized at each date).

MacMul

The use of the three multivariate methods is built on the same model. For each method, three main steps are executed: preparation, computation and interpretation. During each step disk files are created that contain, for example, factor scores to provide direct interfacing with GraphMu. The Macintosh user interface has been employed to simplify the use of the program; scrolling menus, dialog boxes and mouse clicks allow any user to carry out a complete analysis quickly and easily. Keystrokes have been reduced to a minimum and are almost reserved to numerical answers.

Description of software

The first step (preparation) includes the creation of a descriptive file containing information on the data table (number of rows and columns, mean and variance of each variable) and the computation of the table on which the second step will work (original data or averaged or standardized variables for PCA, conditional probabilities for CA, and the so-called 'complete

Laboratoire de Biométrie, URA CNRS 243, Université Claude Bernard, Lyon 1, 43 Boulevard du 11 Novembre 1918, F-69622 Villeurbanne Cedex, France

© IRL Press

Table I. Example data set

	June	August	November	February	
Temperature	−1.37	−1.17	−1.60	+0.50	
Flow	−1.62	−1.27	−1.88	−1.54	
pH	+0.20	+0.94	+0.56	−1.88	
Conductivity	+0.00	−1.53	−0.70	−0.98	
BOD	−0.96	−1.09	−0.60	−1.09	Site A
pO_2	−0.64	−0.77	−0.74	−0.98	
NH_4	−0.74	−1.14	−0.65	−1.18	
NO_3	−0.53	−0.86	−0.57	−1.97	
H_3PO_4	−1.05	−1.64	−1.14	−1.09	
Temperature	−0.39	−1.17	+1.07	+0.50	
Flow	−0.43	−0.78	−0.08	−0.49	
pH	−1.83	−1.40	−1.31	+0.68	
Conductivity	+1.19	+0.94	+1.87	+0.93	
BOD	+1.94	+1.42	+1.98	+1.27	Site B
pO_2	+1.99	+1.95	+1.97	+1.27	
NH_4	+1.98	+1.33	+1.96	−1.04	
NO_3	−1.62	−1.49	−1.62	+0.41	
H_3PO_4	+1.62	+0.61	+1.71	0.41	
Temperature	−0.39	+0.50	−0.27	+0.50	
Flow	−0.02	−0.23	+0.67	+0.00	
pH	0.20	−0.73	−0.84	+0.68	
Conductivity	−0.48	+1.17	+0.12	+1.47	
BOD	−0.42	+0.62	−0.19	+1.03	Site C
pO_2	−0.50	−0.28	−0.23	+1.13	
NH_4	−0.46	+0.84	−0.10	+1.31	
NO_3	0.40	+0.49	+1.14	+0.58	
H_3PO_4	−1.07	+1.15	+0.48	+1.74	
Temperature	0.59	+1.33	+1.07	+0.50	
Flow	0.82	+0.86	+0.95	+1.43	
pH	1.22	−0.07	+0.09	+0.68	
Conductivity	−0.48	+0.04	−0.35	−0.71	
BOD	−0.31	+0.21	−0.57	−0.14	Site D
pO_2	−0.50	−0.21	−0.41	−0.45	
NH_4	−0.41	+0.08	−0.57	−0.36	
NO_3	0.40	+0.69	+0.84	+0.24	
H_3PO_4	0.14	+0.48	−0.45	−0.42	
Temperature	+1.57	+0.50	−0.27	−2.00	
Flow	+1.24	+1.41	+0.34	+0.59	
pH	+0.20	+1.27	+1.50	−0.17	
Conductivity	−0.95	−0.63	−0.94	−0.71	
BOD	−0.25	−1.17	−0.60	−1.06	Site E
pO_2	−0.36	−0.70	−0.59	−0.98	
NH_4	−0.38	−1.11	−0.64	−0.81	
NO_3	+1.34	+1.15	+0.21	+0.75	
H_3PO_4	+0.35	−0.61	−0.60	−0.65	

The five groups of rows correspond to the five sampling sites and the four columns correspond to the four sampling dates. Each variable is standardized at each date (average of five sites equals zero, variance equals one). Sites A, B, C, D and E are arranged upstream to downstream.
(i) Water temperature, (ii) flow, (iii) pH, (iv) conductivity, (v) BOD, (vi) pO_2, (vii) NH_4^+, (viii) NO_3^-, (ix) PO_4^{---}.

disjunctive table' for MCA). The user is asked for a 'generic filename' that is used to build the names of all other files created for the current analysis by concatenation of extension strings.

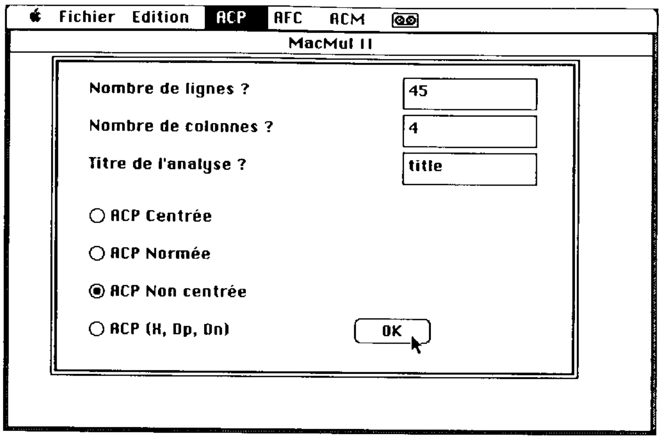

Fig. 1. Main dialog box of the preparation step for a PCA with MacMul. The user is asked for the number of rows and of columns of the data table and also for a 'generic filename', which is used to build the names of all other files created in the current analysis (by concatenation of extension strings). Four 'radio buttons' permit to choose between the various types of PCA (centered, standardized, non-centered or generalized).

For the following steps, the user will be asked only for this generic filename. The 'non-centered PCA' option was used in the example (Figure 1).

The second step (computation) consists of the computation of the matrix that is to be diagonalized, the eigenvalues and eigenvectors, and the factor scores. Files containing these factor scores are stored on disk at the end of this step. Tables II and III show the eigenvalues and the factor scores (for columns and rows) obtained for our small data set.

The third step may include any of the following aid to interpretation:

– for PCA, CA and MCA: inertia analysis for rows or columns (absolute and relative contributions) and projection of additional rows or columns (see Lebart *et al.*, 1984, p.46; Greenacre, 1984, p.66),

– for PCA and CA: data reconstitution (re-computation of initial data table using one or more factors, Lebart *et al.*, 1984, p.8),

– for PCA: PCA of row percentages tables (Doledec, 1986, p.244),

– for CA: double discrimination (CA considered as a double discriminant analysis between rows and between columns (see Greenacre, 1984; D.Chessel and J.Thioulouse, in preparation),

– for MCA: variables/factor scores correlation ratios (Pialot *et al.*, 1984; Tenenhaus and Young, 1985, p.101).

The system of file names conventions avoids the possibility of mixing different methods or using an illegal interpretation aid (like correlation ratios for PCA). Inertia analyses of columns and rows for the above example give the results presented in Tables IV and V. As pointed out by Lebart *et al.* (1984), inertia analyses are designed to help interpreting the results of multivariate methods; they provide two kinds of coefficients which have symmetric meanings:

Table II. Eigenvalues (on the first line) and factor scores of the columns of Table I (first three axes)

	F1	F2	F3
Eigenvalues	2.813	0.7521	0.2541
Factor scores			
June	0.8500	−0.4174	−0.2008
August	0.9060	0.0087	0.4208
November	0.9242	−0.1452	−0.1285
February	0.6448	0.7461	−0.1423

– absolute contributions indicate the part of variance of a principal component due to each item (row or column)

– relative contributions indicate the part of variance of one item (row or column) explained by each principal component.

Algorithms, constraints and additional features

All computation algorithms for multivariate methods are well known. The main difficulty is the diagonalization of a symmetric semi-definite positive matrix. We chose subroutines TRIDI and VPROP of Lebart *et al.* (1977, 1984) from the algorithm of Wilkinson and Reinsch (1971). This procedure provides both good numerical stability and high computation speed. The first step consists of the reduction of the matrix to tridiagonal form (subroutine TRIDI) and the second consists of the extraction of eigenvalues and eigenvectors (subroutine VPROP).

Two kinds of constraints are to be taken into account. (i) The size of the data table. The maximum number of variables allowed is about 200, according to the number of rows of the table. The file containing the initial data table (ASCII file of type TEXT) cannot have more than 1024 characters per line (this is due to a limitation of the compiler), however, this does not limit the number of variables since one logical line of the data table may extend on several consecutive lines of the file. (ii) The computation time. We compared the time used in the computation step for the PCA of a 50 rows × 20 columns table on a Macintosh SE and a Macintosh II, with MacMul and with STATVIEW. For MacMul the computation times were 20 s on a Mac SE and 4 s on a Mac II. For STATVIEW the computation times were 80 s on a Mac SE and 16 s on a Mac II. The computation step for the CA of a 200 rows × 105 column table takes about 2 min 30 s with MacMul on a Mac II.

MacMul was developed with the Microsoft FORTRAN compiler. The parts of code concerning intensive computations are duplicated: one set is compiled with MC68000 instructions and software floating point operations and the other with MC68020/68030 microprocessor instructions and MC68881/68882 floating point instructions. MacMul first tests if it is running on a Mac II (or a Mac IIx, IIcx, or SE/30). In this case, the parts of code generated with MC68020/68030 and MC68881/68882 instructions are executed. This feature ensures both high computation speed and compatibility with Mac Plus, Mac SE and Mac II.

Table III. Factor scores of the rows of Table I (first three axes)

F1	F2	F3
−2.0160	1.3460	−0.1639
−3.1350	−0.2433	0.4993
0.1950	−1.7980	1.0930
−1.5890	−0.7413	−0.8220
−1.8250	−0.3862	−0.0668
−1.5250	−0.4190	0.0775
−1.8030	−0.5617	−0.1580
−1.8040	−1.3530	0.1946
−2.4650	−0.2581	−0.3523
−0.0478	0.4270	−1.2350
−0.8717	−0.2090	−0.3211
−2.1440	1.6710	−0.2974
2.8640	−0.4228	−0.7156
3.3290	−0.1583	−0.4509
3.6360	−0.1753	−0.0259
3.2020	−0.3730	−0.4719
−2.3610	1.3890	−0.3009
2.2500	−0.7071	−0.6879
0.1159	0.6681	0.5004
0.2348	−0.1048	−0.3548
−0.4945	0.6220	−0.6667
1.0200	1.4870	0.7223
0.4133	1.1260	0.4426
−0.0970	1.2490	−0.2948
0.6691	1.3740	0.5402
1.3190	0.1206	−0.2047
1.0120	1.9430	0.7727
1.7990	−0.0195	0.4611
1.9530	0.6852	−0.2546
0.8914	−0.0179	−0.7593
−0.6874	−0.3208	0.5142
−0.4116	0.1263	0.4836
−0.7657	−0.0800	0.2554
−0.6170	−0.0162	0.4770
1.1310	−0.1197	0.1347
−0.0792	−0.3486	0.5781
0.1481	−2.4260	0.4252
1.8040	−0.1320	0.4297
1.5490	−0.4808	0.6458
−1.6130	−0.0026	0.2927
−1.4970	−0.7030	−0.4248
−1.2620	−0.5781	−0.0138
−1.4560	−0.4180	−0.3833
1.7040	−0.0233	0.1609
−0.7326	−0.6334	−0.3121

An on-line help facility provides information on all menus, dialog boxes and input/output files. The text displayed by this option is stored in a file named 'MacMul.hlp'.

For the three methods (PCA, CA and MCA), MacMul automatically accepts tables having more columns than rows and performs computations in the lowest dimension space. For CA this only saves transposing the table when the number of columns is greater than the number of rows but for PCA and

Table IV. Inertia analysis of the columns of Table I (output listing of program MacMul)

ACP: ANALYSE D'INERTIE DES COLONNES.
TITRE DE L'ANALYSE:
title 45 LIGNES 4 COLONNES 4 FACTEURS
CONTRIBUTIONS ABSOLUES DES COLONNES.
LES FACTEURS SONT EN COLONNES:

1	2568.	2317.	1587.
2	2918.	1.	6967.
3	3036.	280.	650.
4	1478.	7402.	796.

CONTRIBUTIONS RELATIVES DES COLONNES
LES FACTEURS SONT EN COLONNES, SUIVIS PAR LES
CONTRIBUTIONS RESIDUELLES, PUIS LES POIDS,
ET ENFIN LES CONTRIBUTIONS A LA TRACE:

1	7220.	1741.	403.	636.	10000.	2502.
2	8208.	1.	1770.	21.	10000.	2500.
3	8537.	211.	165.	1087.	10000.	2502.
4	4163.	5575.	203.	58.	10000.	2496.

Table V. Inertia analysis of the rows of Table I (output listing of program MacMul)

ACP: ANALYSE D'INERTIE DES LIGNES.
TITRE DE L'ANALYSE:
title 45 LIGNES 4 COLONNES 4 FACTEURS
CONTRIBUTIONS ABSOLUES DES LIGNES.
LES FACTEURS SONT EN COLONNES:

1	321.	535.	23.
2	775.	17.	218.
3	3.	955.	1044.
4	199.	162.	591.
5	263.	44.	4.
6	184.	52.	5.
7	257.	93.	22.
8	257.	541.	33.
9	480.	20.	109.
10	0.	54.	1334.
11	60.	13.	90.
12	363.	825.	77.
13	648.	53.	448.
14	876.	7.	178.
15	1044.	9.	1.
16	810.	41.	195.
17	440.	570.	79.
18	400.	148.	414.
19	1.	132.	219.
20	4.	3.	110.
21	19.	114.	389.
22	82.	654.	456.
23	13.	375.	171.
24	1.	461.	76.
25	35.	558.	255.
26	137.	4.	37.
27	81.	1116.	522.
28	256.	0.	186.
29	301.	139.	57.
30	63.	0.	504.
31	37.	30.	231.
32	13.	5.	205.
33	46.	2.	57.
34	30.	0.	199.
35	101.	4.	16.
36	0.	36.	292.
37	2.	1739.	158.
38	257.	5.	161.
39	189.	68.	365.
40	205.	0.	75.
41	177.	146.	158.
42	126.	99.	0.
43	167.	52.	128.
44	229.	0.	23.
45	42.	119.	85.

CONTRIBUTIONS RELATIVES DES LIGNES.
LES FACTEURS SONT EN COLONNES, SUIVIS PAR LES
CONTRIBUTIONS RESIDUELLES, PUIS LES POIDS,
ET ENFIN LES CONTRIBUTIONS A LA TRACE:

1	6709.	2990.	44.	256.	222.	336.
2	9689.	58.	246.	7.	222.	564.
3	80.	6775.	2502.	643.	222.	265.
4	6659.	1450.	1782.	109.	222.	211.
5	9105.	408.	12.	475.	222.	203.
6	9260.	699.	24.	16.	222.	139.
7	8873.	862.	68.	198.	222.	203.
8	6230.	3503.	72.	194.	222.	290.
9	9677.	106.	198.	19.	222.	349.
10	8.	625.	5232.	4135.	222.	162.
11	7307.	420.	991.	1281.	222.	58.
12	6140.	3730.	118.	13.	222.	416.
13	9220.	201.	576.	3.	222.	494.
14	9798.	22.	180.	0.	222.	629.
15	9970.	23.	1.	6.	222.	737.
16	9659.	131.	210.	0.	222.	590.
17	7298.	2525.	119.	58.	222.	424.
18	8317.	821.	777.	85.	222.	338.
19	185.	6157.	3454.	204.	222.	40.
20	1098.	219.	2507.	6176.	222.	28.
21	1404.	2223.	2553.	3820.	222.	97.
22	2756.	5861.	1382.	0.	222.	210.
23	1030.	7653.	1181.	135.	222.	92.
24	57.	9401.	524.	18.	222.	92.
25	1694.	7138.	1104.	64.	222.	147.
26	8538.	71.	206.	1184.	222.	113.
27	1790.	6595.	1043.	572.	222.	318.
28	9218.	1.	605.	176.	222.	195.
29	8752.	1077.	149.	22.	222.	242.
30	4046.	2.	2936.	3016.	222.	109.
31	5504.	1199.	3080.	218.	222.	48.
32	3495.	329.	4825.	1351.	222.	27.
33	8821.	96.	982.	101.	222.	37.
34	6053.	4.	3618.	325.	222.	35.
35	9134.	102.	130.	634.	222.	78.
36	100.	1932.	5315.	2653.	222.	35.
37	32.	8672.	266.	1029.	222.	377.
38	8159.	44.	463.	1334.	222.	222.
39	6099.	588.	1061.	2252.	222.	218.
40	9678.	0.	319.	3.	222.	149.
41	7686.	1695.	619.	0.	222.	162.
42	8265.	1734.	1.	0.	222.	107.
43	8683.	715.	602.	0.	222.	136.
44	7798.	1.	69.	2131.	222.	207.
45	4203.	3141.	763.	1893.	222.	71.

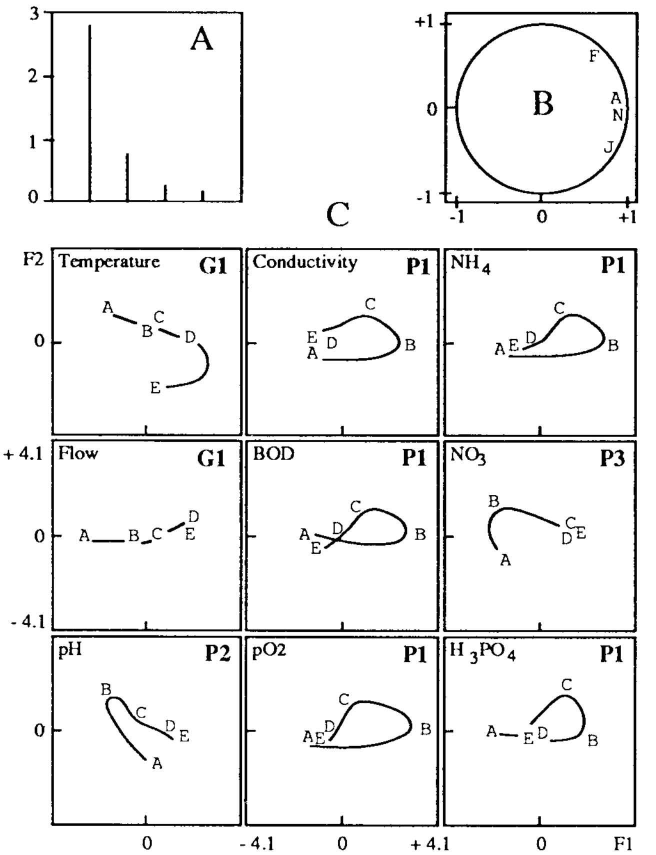

Fig. 2. Example of drawings produced with MacMul and GraphMu and completed with MacDraw. (**A**) Bar chart of first four eigenvalues. (**B**) Correlation circle of the four sampling dates (F, February; J, June; A, August; N, November). (C) First two principal axes plane of the nine variables measured in the five sampling sites (one graphic for each variable); each axis carries the standardized values of a principal component. See text and Thioulouse and Chessel (1987).

MCA this is the only way to obtain the dual analysis (Whittle, 1952; Lebart *et al.*, 1984).

Lastly the presentation of inertia analysis has been unified for rows and columns and for the three methods, thus leading to easier comparisons.

GraphMu

GraphMu is specially designed for graphical display of multivariate data and for graphical output of multivariate statistical methods (Auda, 1983). The main feature of this software is its ability to automatically draw *collections* of graphics. These collections may be ordered either according to the columns of the data table (comparison of sets of variables), or according to its rows (comparisons of sets of individuals) or to both of them (comparison of sets of individuals for several variables). This feature is provided by a selection mechanism among rows and columns. Columns are selected in dialog boxes, and rows are selected either in dialog boxes or through a selection file. Each elementary graphic of a collection may be chosen between three general classes: curves, maps and graphic models. The first class includes five types of curves: dots, lines, bars, histograms and stepped curves. The second class has three types of maps: maps with squares or circles (one circle or one square of size proportional to the data is drawn at each point of the map) or with labels (corresponding to classical principal axes planes of multivariate methods). Currently the third class offers only two types of graphical models: ellipses and Gauss curves. Several dialog boxes allow the user to set up the parameters of a drawing by specifying coordinates file name, selection of sets of variables and individuals, number of horizontal and vertical elementary graphics, minimum and maximum of coordinates, for example.

Files created by MacMul (mainly factor scores) are directly read by GraphMu. A translation option makes it possible to use any ASCII file. Drawings may be saved in disk files of type PICT and altered with any software compatible with this file type (e.g. MacDraw, MacDraft, SuperPaint, CricketDraw, etc).

GraphMu was also written with Microsoft FORTRAN compiler. The maximum size for data files is 256 columns and 10 000 rows. Two help facilities are available: an on-line help, which describes each menu, dialog boxes field and input/output file (text stored in GraphMu.hlp) and a HyperCard stack (GraphStack), which provides a Macintosh-like, user-oriented tutorial.

Figure 2 shows the results obtained with the example. The first two principal axes planes (C) were drawn using GraphMu 'labelled maps' option with a separate graphic for each variable (3×3 graphics). Each elementary graphic contains the five sampling sites (axes limits are ± 4.1 for all graphics): this presentation underlines the existence of four processes relating to the pollution of the stream (P1, P2, P3) and to geographical gradients (G1). Graphics of the variables conductivity, BOD, pO_2, NH_4, H_3PO_4 show that the pollution is maximum at site B (due to a winter resort located between sites A and B) and that restoration then slowly takes place downstream (sites C, D, and E). See Thioulouse and Chessel (1987) or Doledec and Chessel (1987) for a more complete description of the ecological processes involved. Eigenvalues bar chart (A) and correlation circle of the four sampling dates (B) were drawn separately using other GraphMu options, and the final diagram was completed with MacDraw by adding variable names, legends and hand-drawings.

Availability

MacMul and GraphMu are available from the author in exchange for two 800K Macintosh 3.5″ disks.

Acknowledgements

MacMul was developed starting from various BASIC programs which were part of the software library of the Laboratoire de Biométrie (University of Lyon, France) mainly written by D.Chessel on a Data General Eclipse S/140 mini-computer. GraphMu has evolved from the program GRAPHIQUE written by Y.Auda on the same computer. I wish to thank D.Chessel, D.Debouzie and C.Gautier for critical reading of the manuscript.

References

Auda,Y. (1983) *Rôle des Méthodes Graphiques en Analyse des Données: Application au Dépouillement des Enquêtes Écologiques.* Thesis, Université de Lyon.

Bournaud,M., Maucet,D. and Chavanon,G. (1983) Structure des peuplements de macro-invertébrés benthiques en place et an dérive. *Acta Oécol., Oécol Gen.*, **4**, 39–57.

Doledec,S. (1986) *Les Peuplements de Macro-Invertébrés Benthiques du Cours Inférieur de l'Ardèche. Dynamique Spatio-Temporelle.* Ph.D. Thesis, Université de Lyon.

Doledec,S. and Chessel,D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observation complet par projection de variables. *Acta Oécol., Oécol. Gen.*, **8**, 403–426.

Fichant,G. and Gautier,C. (1987) Statistical method for predicting protein coding regions in nucleic acid sequences. *Comput. Applic. Biosci.*, **3**, 287–295.

Grantham,R., Gautier,C. and Gouy,M. (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.*, **8**, 1893–1911.

Greenacre,M.J. (1984) *Theory and Applications of Correspondence Analysis.* Academic Press, London.

Lebart,L., Morineau,A. and Tabard,N. (1977) *Techniques de la Description Statistique: Méthodes et Logiciels pour l'Analyse des Grands Tableaux.* Dunod, Paris.

Lebart,L., Morineau,A. and Warwick,K.M. (1984) *Multivariate Descriptive Statistical Analysis.* Wiley, New York.

Pailot,D., Chessel,D. and Auda,Y. (1984) Description du milieu et analyse factorielle des correspondances multiples. *C.R. Acad. Sci. Paris, Série III*, **298**, 309–314.

Quentin,Y. (1988) The Alu family developed through successive waves of fixation closely connected with primate lineage history. *J. Mol. Evol.*, **27**, 194–202.

Quentin,Y. (1989) Successive waves of fixation of B1 variants in rodent lineage history. *J. Mol. Evol.*, **28**, 299–305.

Tenenhaus,M. and Young,F.W. (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika*, **50**, 91–119.

Thioulouse,J. and Chessel,D. (1987) Les analyses multitableaux en écologie factorielle. I. De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Oécol., Oécol. Gen.*, **8**, 463–480.

Whittle,P. (1952) On principal components and least squares methods of factor analysis. *Skand. Aktaur.*, **35**, 223–239.

Wilkinson,J.H. and Reinsch,C. (1971) *Handbook for Automatic Computation.* Vol. 2. Springer Verlag, Berlin.

Received on January 24, 1989; accepted on July 17, 1989

Circle No. 4 on Reader Enquiry Card