Computers & Geosciences Vol. 16, No. 8, pp. 1235–1240, 1990
Printed in Great Britain.

0098-3004/90 $3.00 + 0.00

SHORT NOTE

MACMUL AND GRAPHMU: TWO MACINTOSH PROGRAMS FOR THE DISPLAY AND ANALYSIS OF MULTIVARIATE DATA

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

(*Received* 23 *January* 1990; *accepted* 3 *April* 1990)

INTRODUCTION

Multivariate analysis methods have been in use for several years in geology, and the work of Davis (1973) has helped to make the basic methods (and in particular principal component analysis) well-known. Thanks to today's high-performance microcomputers, these methods are now more accessible and much easier to use. In this field, the Macintosh interface can simplify the use of software for non-specialists in computer science.

Two programs are presented in this article. The first, MacMul, makes it possible to perform computations for the three basic multivariate data analysis methods (Lebart, Morineau, and Warwick, 1984): principal component analysis (PCA) for quantitative variables, correspondence analysis (CA) for contingency tables, and multiple correspondence analysis (MCA) for qualitative variables. Use of the three methods has been standardized completely, both for input files and program outputs. The second program, GraphMu, allows the user to draw the graphics that usually are associated with these methods (planes of principal axes) as well as different types of graphics which are adapted specially to the simultaneous study of several variables.

An example is taken from the work of Davis (1973) to show some of the possibilities of the two programs. Thioulouse (1989) gives another example, using a set of hydrobiological data.

COMPUTATIONS

The example consists of a table with 50 rows, representing 50 soil samples, and seven columns, corresponding to seven size classes (table 7.21 from Davis, 1973, p. 494–495). The 50 samples belong to five different groups: beach and foreshore sands (I), silty channel sands (II), silty channel margin sands (III), organic bottom silts (IV), and organic muds from lees of islands (V). The table gives, for each sample, the distribution of the particles into the seven diameter categories (Table 1). We used MacMul to carry out the PCA of this table.

The first step consists in inputting the table, using one of the classical Macintosh programs. It can be carried out with the help of a spreadsheet (Excel, StatView, . . .) or a wordprocessing program (MacWrite, Word, Edit, . . .). All these programs allow the user to save the table into a text file. The second step involves the transformation of this text file into a binary file, so that it can be used by MacMul. The "Transfo. TEXT→BIN" option of the "File" menu of MacMul or GraphMu makes this possible (Fig. 1).

The computations as such are carried out by the "Preparation" and "Computation" options of the "PCA" menu. After having supplied a certain amount of information (numbers of rows, numbers of columns, title of the analysis), and selected the type of analysis (centered, standardized, . . .), the program displays the list of eigenvalues and asks how many factor scores are to be calculated (Fig. 2). The files containing the factor scores of the rows and columns then are obtained. All the output files have a name that is created automatically from a generic filename (the title of the analysis, supplied by the user), and to which are added four-letter extensions. The first two letters of the extension identify the method used: "cp" for PCA (ACP in French), "fc" for CA (AFC in French), and "mc" for MCA. The following two letters identify the contents of the file: "li" for the row coordinates ("lignes" in French), "co" for the column coordinates, "vp" for eigenvalues ("valeur propres" in French). These files are binary and can be used directly by GraphMu to draw the graphs. It also is possible to transform them into text files with the "Transfo. BIN→TEXT" option of the "File" menu of MacMul or GraphMu, so that they can be used with other programs.

Table 1. Data from Davis (1973, p. 494–495), table 7.21: 50 grain-size analyses (Barataria Bay, Louisiana). Samples are classified into five types (I–V), and grain sizes into seven categories

		1-2	2-3	3-4	4-5	5-6	6-7	7-8
I	A	0.6	70.2	29.2	0.0	0.0	0.0	0.0
	B	1.0	69.9	29.1	0.0	0.0	0.0	0.0
	C	0.8	73.7	25.5	0.0	0.0	0.0	0.0
	D	0.9	75.3	23.8	0.0	0.0	0.0	0.0
	E	0.6	62.5	36.9	0.0	0.0	0.0	0.0
	F	1.1	68.8	30.1	0.0	0.0	0.0	0.0
	G	0.8	10.2	79.2	9.8	0.0	0.0	0.0
	H	1.0	16.3	73.8	8.9	0.0	0.0	0.0
	I	1.8	35.7	61.9	0.6	0.0	0.0	0.0
II	A	9.5	15.8	59.0	8.4	0.9	0.9	1.4
	B	2.4	14.5	53.9	12.2	5.5	1.6	2.5
	C	2.2	38.8	42.2	7.9	1.4	1.8	1.0
	D	1.7	30.4	44.5	11.2	3.0	1.9	2.9
	E	0.0	40.0	32.5	3.8	4.5	6.5	2.7
	F	0.0	37.0	45.4	7.3	3.8	3.3	3.8
	G	0.3	15.6	54.1	21.3	4.1	2.6	2.0
	H	0.3	24.4	56.0	15.1	4.2	0.0	0.0
	I	10.5	29.2	37.3	15.1	4.2	3.7	0.0
	J	0.3	13.3	63.5	14.2	4.0	3.4	1.3
	K	1.2	26.9	54.7	11.0	3.9	2.3	0.0
III	A	0.4	3.9	45.2	24.7	3.7	8.1	3.0
	B	0.0	13.8	39.3	15.4	9.1	4.5	6.4
	C	0.4	4.0	38.2	28.5	6.0	4.3	4.7
	D	1.9	11.5	49.5	22.4	5.7	4.5	2.0
	E	0.4	5.1	31.8	30.3	5.4	7.8	3.0
	F	0.5	5.9	32.2	32.7	4.9	5.4	2.7
	G	1.1	4.9	31.1	41.9	13.9	7.8	3.7
	H	7.9	8.5	21.0	19.9	8.9	5.9	6.3
	I	0.9	13.6	43.9	20.1	7.2	4.8	9.5
	J	2.9	15.5	37.0	30.3	5.1	1.9	2.2
	K	2.1	16.7	39.6	17.7	8.3	8.3	7.3
	L	0.3	20.6	55.4	16.6	6.2	6.1	5.5
IV	A	1.2	1.6	15.3	38.4	13.0	9.5	5.6
	B	2.3	7.9	23.9	25.5	9.2	7.9	7.7
	C	1.0	3.1	15.2	32.0	14.3	10.0	7.2
	D	0.0	11.5	28.4	19.1	7.3	7.8	4.8
	E	0.8	7.0	31.6	21.1	10.2	9.0	6.3
	F	0.5	2.1	14.0	37.2	19.9	11.4	6.1
	G	0.0	3.4	19.7	25.4	15.7	10.2	9.9
	H	1.4	1.9	14.4	40.2	8.5	8.4	7.1
	I	0.4	3.5	18.8	29.5	11.2	10.4	7.5
	J	0.8	6.3	18.2	28.0	9.1	9.7	9.9
V	A	1.0	2.3	6.6	16.2	12.0	11.4	13.3
	B	3.2	3.9	10.5	24.1	14.2	15.4	13.5
	C	2.1	2.1	10.7	23.6	15.1	14.0	11.8
	D	4.4	8.1	8.9	19.9	12.0	11.4	10.8
	E	0.6	3.6	4.2	17.8	12.4	10.8	9.9
	F	0.5	4.1	9.8	27.9	13.5	13.5	7.4
	G	0.7	2.3	5.2	23.2	19.4	14.1	10.1
	H	3.4	1.6	4.4	18.0	14.7	15.3	15.1

GRAPHICS

The graphical display usually used after a PCA-type multivariate analysis is the plane of principal axes. The plane of the first two factors of the PCA of Table 1 is given in figure 7.25 in Davis (1973, p. 499). It was drawn using GraphMu, and added the text and the envelope curves separating the five types of samples to the drawing with MacDraw (Fig. 3). To draw this figure, GraphMu needs the file of the factor scores of the PCA, and a text file containing the characters to be drawn on the plane. The latter file is optional: it may be created with any text editor, but in its absence it is the number of the element which is entered on the principal plane.

One of the main advantages of GraphMu is that it makes it possible to draw automatically **collections** and **superimpositions** of graphics. The different graphics of a collection may correspond on the one hand to different columns of the input table, and on the other hand to different groups of rows. For example, it is possible here to separate the five sample groups, and to draw automatically the five graphs of Figure 4. On this figure, the samples are represented by a letter (A, B, C, ...), so that they can be located individually (and in particular samples G, H, and I of group 1). There are two ways of selecting the rows and columns associated with an elementary graphic: either manually (in a dialog box where you can type for example "20 to 31" to select rows 20–31), or by using a selection file which gives, for each row, the number of the graphic to which it belongs.

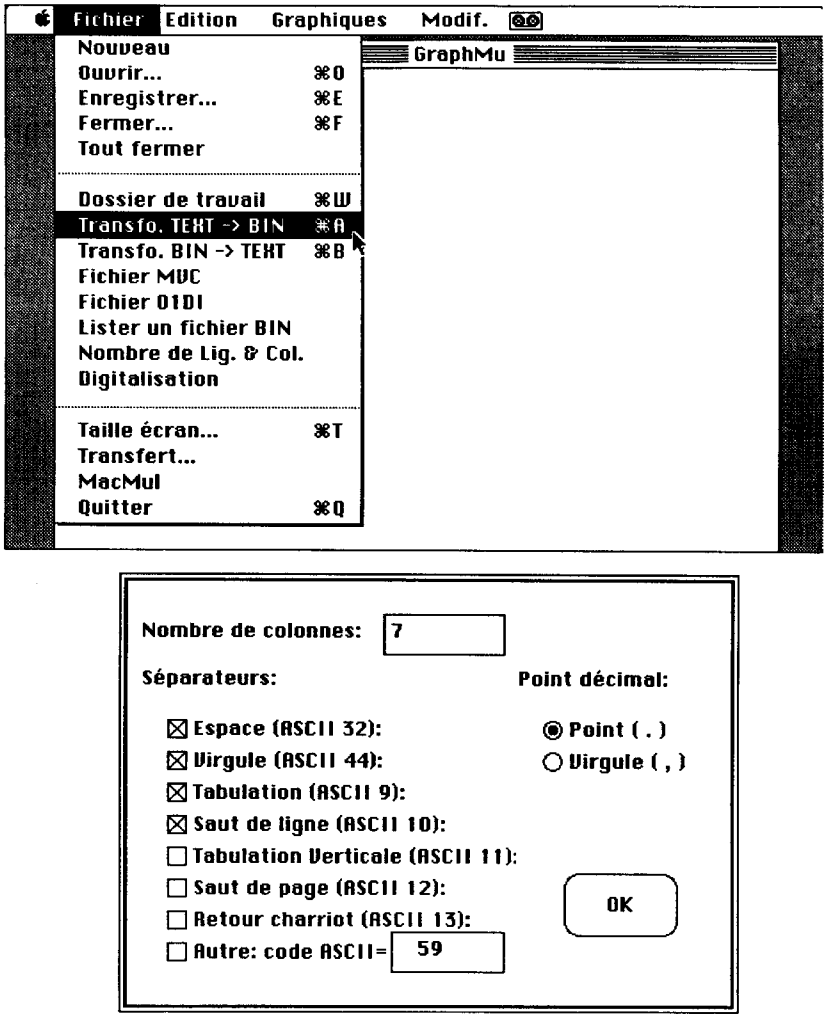

Figure 1. "Fichier" menu of GraphMu, showing "Transfo. TEXT→BIN" option, used to transform text file into binary file, and corresponding dialog box: user is asked for number of columns of data file, and list of possible separators is presented; one also can select between comma (French software) or period for decimal mark.

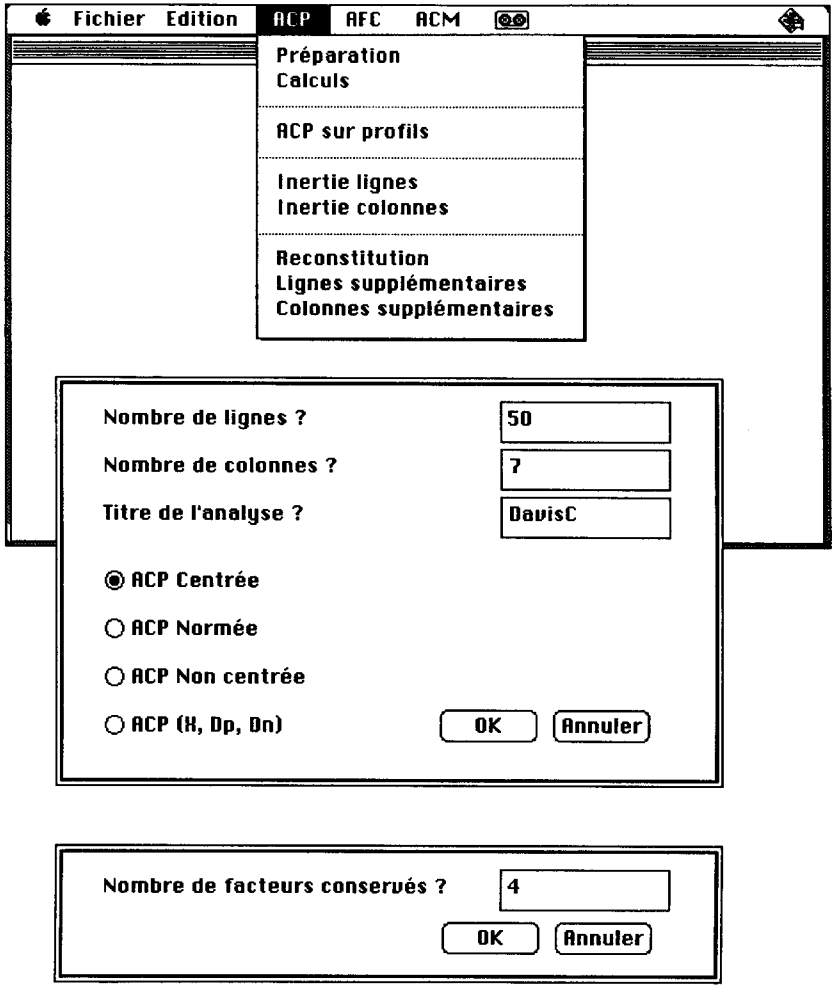

Figure 2. "ACP" menu of MacMul, showing successive steps of analysis: preparation, computation, and interpretation aids (inertia analyses, additional rows and columns, data reconstitution). Below is first dialog box, where user is asked for number of rows and columns of data table, for title (used to build output file names), and for type of PCA (centered, standardized, noncentered, or general). When computations are terminated, list of eigenvalues is presented on screen, and number of factors which are to be written into output files (number of factor scores) may be selected.

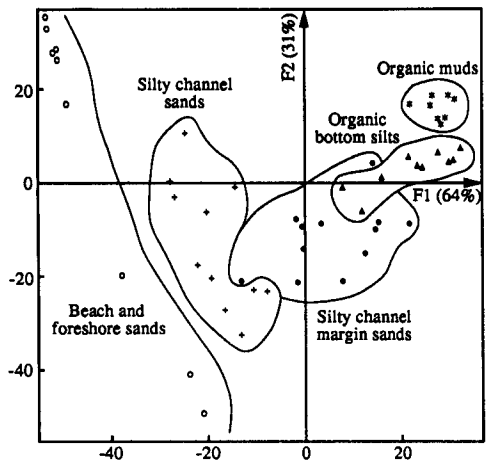

Figure 3. Plane of first two axes of PCA. First axis represents 64% of total inertia, and second one 31%. MacDraw was used to add hand drawings (envelope curves of the five groups), and text (name of each group and percentages of inertia of each axis) to output file of GraphMu.

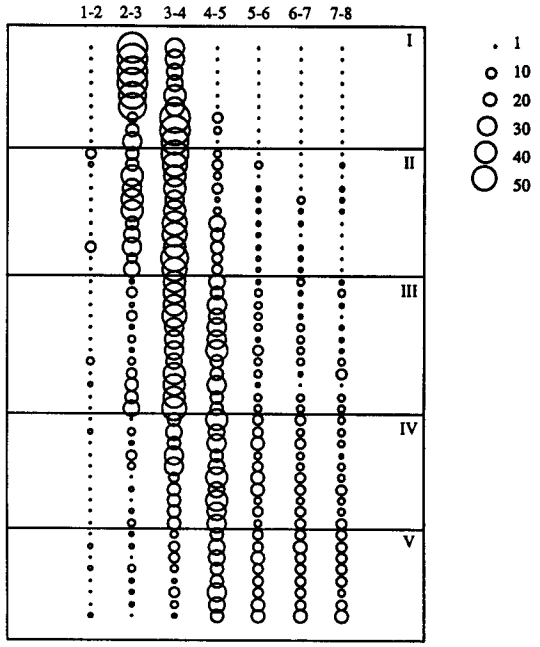

Figure 5. Diameter of circles is proportional to percentage of particles of each size class in 50 samples. Exact relationship between circle sizes (Φ) and original data (x) is given by: $\Phi = 0.1x^{0.5}$.

Other types of grahic displays can be produced with the help of GraphMu. For example, in Figure 5, circles have been drawn with diameters which are proportional to the values of the data. This type of display brings out the overall structure of the table and, in particular, the rise in the percentage of small-size particles between groups 1 and 5. However, it does not bring out the particular feature of the last three samples (G, H, I) of group 1. It also is possible to draw graphics with squares, and mixed graphics (circles for negative values, squares for positive ones). The relationship between the diameters of the circles (or the sides of the squares) and the data depends on two parameters, a multiplying factor and a power factor, which can be selected by the user. The relation can be written:

$$\Phi = mx^p$$

where

Φ = diameter of the circle (or side of the square)
m = multiplier factor
p = power factor
x = data.

The program also can be used to draw background maps on which graphics can be superimposed ("Digitalization" option of the "File" menu). GraphMu can draw ellipses and Gaussian curves,

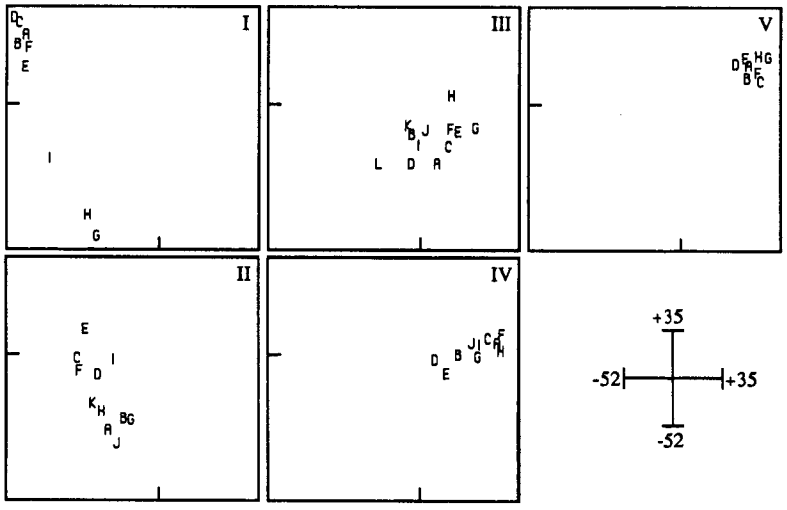

Figure 4. Plane of first two axes of the PCA. Five sample groups are separated, and one graphic is drawn for each group (same scale is used for 5 groups). Within each group, each sample is represented by its letter.

thus making it possible to model graphically the basic parameters (mean, variance, covariance) of several variables measured in several groups of individuals.

Classical graphs (curves, histograms, scatter diagrams, etc.) can be drawn using the options offered by the program of drawing automatically collections and superimpositions. Figure 6 shows the values of Table 1 in the form of a series of histograms. Each graph gives the distribution of the particles making up the 50 samples in seven size classes. This figure makes it easier to understand why the PCA isolates the last three samples in group 1; in these samples, the second diameter class is distinctly under-represented, whereas the third one is over-represented.

GENERAL FEATURES

Several points concerning MacMul and GraphMu need to be emphasized. For MacMul, the maximum number of variables allowed is approximately 200, according to the number of rows; the actual limit is given by:

$$R + 4C + M^2 \leqslant 50{,}000$$

where

R = number of rows in the table

C = number of columns in the table

$M = \min(R,C)$

hence, for $C = 200$ columns, the limit is 9200 rows. As for computation time, Table 2 indicates the values obtained in different situations.

MacMul was developed on a Macintosh using the Microsoft FORTRAN compiler. It tests automatically the type of Macintosh on which it is being run, and uses either processes MC68000 or MC68020/68030, and, if possible, coprocessors MC68881/68882 (compatibility with Mac+, SE, SE/30, II, IIx, IIcx). For all three methods (PCA, CA, and MCA), MacMul automatically accepts tables with more rows than columns, and performs computations in the lowest dimension space. Moreover, interpretation aids (inertia analyses, supplementary rows and columns, data reconstitution) are presented in the same way for the rows, the columns, and for

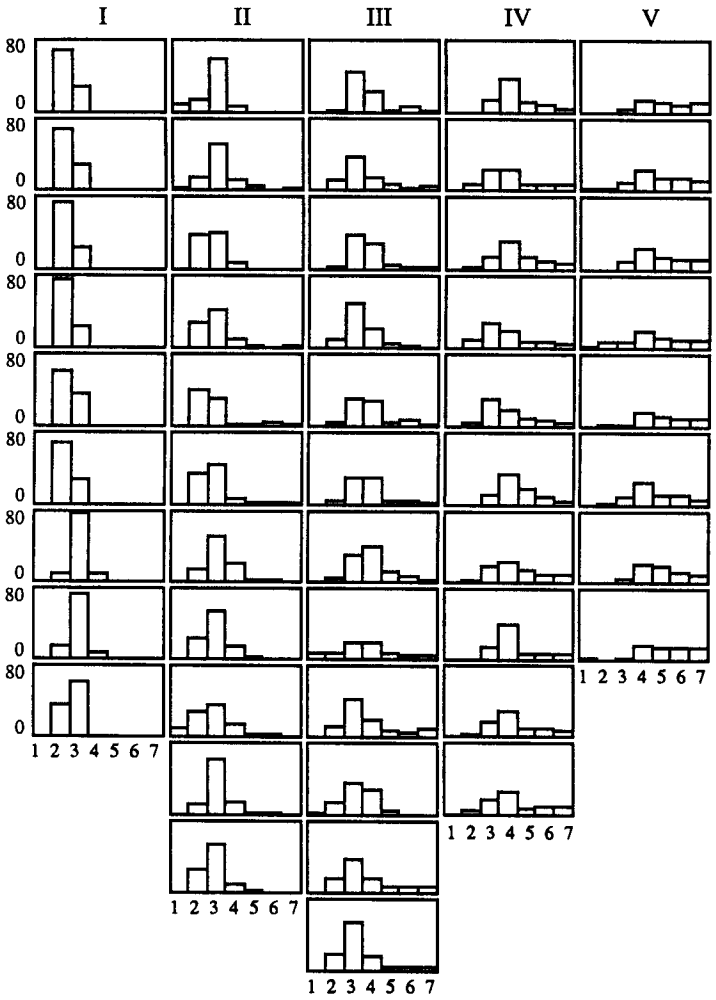

Figure 6. Fifty histograms giving distribution of each sample of each of five groups (I–V) within seven categories of particle size (1–7).

Table 2. Computation times (in seconds) for PCA of 50 rows × 40 columns table, and for CA of 200 rows × 105 columns table, on Macintosh SE, and on Macintosh SE/30. For PCA, results are compared to those obtained with StatView (StatView 512+® on Mac SE StatView II® on Mac SE/30)

	PCA 50x40		CA 200x105
	MacMul	StatView	MacMul
Mac SE	120	292	2160
Mac SE/30	10	8	174

all three methods. An on-line help facility, as well as a Microsoft Word® file, provides information on all menus, dialog boxes, and input/output files.

For GraphMu the maximum number of variables is 256 and the maximum number of points for a graphic is 50,000. Drawings can be stored as PICT files, thus they can be modified using compatible Macintosh software (e.g. MacDraw-like software). The clipboard can be used to copy/paste graphics from one window to another in GraphMu (thus making superimposition easy), as well as between GraphMu and other Macintosh software. A "Modification" menu makes it possible to change the parameters of a graphic rapidly, (number of horizontal and vertical graphics, scales, minima and maxima of abscissae and ordinates, selection of groups of rows and columns, . . .), without having to redefine them all. The program automatically creates windows of size 300 × 450 on a Macintosh Plus or an SE, and of size 400 × 600 on a Macintosh with a larger screen. These values can be modified, either by clicking with the mouse in the bottom right-hand corner of the window, or by using the "Screen Size" option of the "File" menu (in all instances the height:width ratio should be kept to 2:3). Finally, the "MacMul" option of the "File" menu makes it possible to launch MacMul directly from GraphMu (there is a "GraphMu" option in the MacMul "File" menu, which allows the user to launch GraphMu directly from MacMul). A HyperCard stack (GraphStack) provides a Macintosh-like, user-oriented tutorial.

AVAILABILITY

MacMul and GraphMu are available from the author.

Acknowledgments—MacMul was developed from various BASIC programs in the software library of the Laboratoire de Biométrie (University of Lyon-1), written mainly by D. Chessel on a Data General Eclipse S/140 minicomputer. GraphMu was developed from the "Graphique" program, written by Y. Auda (1983) on the same computer.

REFERENCES

Auda, Y., 1983, Rôle des méthodes graphiques en analyse des données: application au dépouillement des enquêtes écologiques: unpubl. these, Univ. de Lyon, 127 p.

Davis, J. C., 1973, Statistics and data analysis in geology: John Wiley & Sons, New York, 550 p.

Lebart, L., Morineau, A., and Warwick, K. M., 1984, Multivariate descriptive statistical analysis: John Wiley & Sons, New York, 231 p.

Thioulouse, J., 1989, Statistical analysis and graphical display of multivariate data on the Macintosh: Computer Applications in the Biosciences, v. 5, no. 4, p. 287–292.