122

PSYCHOMETRIKA—VOL. 48, NO. 1
MARCH, 1983
REVIEW

REVIEW

F. Cailliez and J.-P. Pages, *Introduction à l'Analyse des Données.* Paris: Societé de Mathématiques Appliquées et de Sciences Humaines, 9 rue Duban, 75016 Paris, 1976, pp XXII + 616, 120 francs.

This remarkable book treats multivariate linear analysis in a manner that is with both distinctive and profoundly promising for future work in this field. With an approach that is strictly algebraic and geometric, it avoids almost all discussion of probabilistic notions, introduces a formalism that transcends matrix algebra, and offers a coherent treatment of topics not often found within the same volume. Finally, it achieves these things while remaining entirely accessible to nonmathematicians and including many excellent practical examples.

The key notion is *duality*: the principle that associated with a vector space E is the dual vector space E^* consisting of the possible linear mappings from E into the real numbers. If $\mathbf{x}$ is a vector in E and $\mathbf{y}$ is a linear mapping in E^* then the real number resulting from applying $\mathbf{y}$ to $\mathbf{x}$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$. In data analytic situations there will also be associated with E the inner product or bilinear form m: $\mathbf{x}_1, \mathbf{x}_2 \to m(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}$. In matrix terms $m(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^t \mathbf{M} \mathbf{x}_2$, and the symmetric positive definite matrix $\mathbf{M}$ defines the *metric* for E. It can then be said the structure (E, m) is a Hilbert space. In order to represent any p-dimensional vector $\mathbf{x} \in E$ in terms of real-valued coordinates, one can equip E with a set of $\mathbf{M}$-orthogonal basis vectors $\mathbf{e}_j, j = 1, \ldots, p$. The exposition of these basic ideas as well as essential statistical and matrix concepts in the first six chapters is painstaking with much repetition and helpful summaries at the ends of sections and chapters.

Associated with the usual p by n data matrix $\mathbf{X}$ are two vector spaces, each equipped with its own dual space, metric, and basis: E the space of dimension p containing n vectors corresponding to individuals or cases, and F of dimension n containing p vectors corresponding to properties or variables. Let $\mathbf{x}_i$ in E be the vector corresponding to the i^{th} individual, and let $\mathbf{e}_j^*$ be a basis vector in the dual space E^*. Then the ji^{th} entry in $\mathbf{X}$ is given by the value $\langle \mathbf{e}_j^*, \mathbf{x}_i \rangle = x_{ij}$, and associated with each property j we have therefore a linear mapping which is the j^{th} basis vector in E^*. Similarly, we also have $\langle \mathbf{f}_i^*, \mathbf{x}_j \rangle = x_{ij}$, where $\mathbf{f}_i^*$ is the i^{th} basis vector for F^* and $\mathbf{x}_j$ in F is the vector corresponding to the j^{th} property. Thus, to each individual i is associated a linear mapping $\mathbf{f}_i^*$ which is the i^{th} basis vector in F^*. We have, then a natural one-to-one correspondence for properties between elements of E^* and elements of F, and a complementary one-to-one correspondence for individuals between elements of F^* and elements of E. The data matrix $\mathbf{X}$ is also the matrix corresponding to the linear mapping from F^* to E while the matrix $\mathbf{X}$ corresponds to the mapping from E^* to F. All this is summarized in the *duality diagram*:

$$\begin{array}{ccc} & \mathbf{X} & \\ E & \longleftarrow & F^* \\ \mathbf{M} \downarrow & & \uparrow \mathbf{N} \\ E^* & \longrightarrow & F \\ & \mathbf{X}^t & \end{array}$$

Note that the diagram indicates by $\mathbf{M}$ and $\mathbf{N}$ the mappings from E to E^* and from F to

Table 1. Multivariate Statistics Texts in the Gifi Analysis

Author	Date	Plotting Label
Anderson	1958	ANDR
Caillez & Pages	1976	C&P
Cooley & Lohnes	1962	C&P1
Cooley & Lohnes	1971	C&P2
Dagnelie	1975	DAGN
Dempster	1969	DEMP
Gifi	1977	GIRI
Gnanadesikan	1977	GNAN
Green & Carroll	1976	G&C
Harris	1975	HARR
Kendall	1957	KEN1
Kendall	1975	KEN2
Kshirsagar	1978	KSHI
Morrison	1967	MOR1
Morrison	1976	MOR2
Roy	1957	ROY
Tatsuoka	1971	TATS
Thorndike	1978	THOR
Van de Geer	1967	VDG1
Van de Geer	1971	VDG2

F^* induced by $\langle\ \rangle$. In practice, **N** will be the diagonal matrix of reciprocals of standard deviations if one is analyzing standardized data while $\mathbf{M} = \mathbf{I}$. The reader of this review who finds all of this imcomprehensible is assured that the coverage of these ideas in Chapter VII is very slow and well illustrated.

Complementing the mapping m: $E \to E^*/v$: $E^* \to E$ defined by metric is the mapping v: $E^* \to E$ defined by the data in which one takes a tour of the diagram via first F and then F^*. That is, one takes the composition $v = \mathbf{X} \circ n \circ \mathbf{X}^t$ corresponding with the order p positive semidefinite matrix $\mathbf{V} = \mathbf{XNX}^t$. There is, finally, the mapping from E into itself (not necessarily one-to-one) with matrix $\mathbf{XNX}^t\mathbf{M}$. Complementing these mappings are the F^*-to-F and F-to-F mappings with matrices $\mathbf{W} := \mathbf{X}^t\mathbf{MX}$ and $\mathbf{X}^t\mathbf{MXN}$, respectively.

Within the framework of duality theory, principal component analysis and linear regression become the problem of identifying within E the eigenvectors $\mathbf{u}_j$, $j = 1, \ldots, p$, satisfying $\mathbf{XNX}^t\mathbf{Mu}_j = \lambda_j \mathbf{u}_j$ having minimum residual variance (termed the moment of inertia in the French statistical tradition). These topics along with examples are treated in

Table 2. Subject Categories in the Gifi Analysis

Subject Category	Plotting Label
Mathematics other than statistics	MATH
Correlation and regression	CORR
Factor analysis and principal components analysis	FACT
Canonical correlation	CANO
Discriminant analysis & cluster analysis	DISC
Statistical theory	STAT
General linear model & MANOVA	MANO

Chapters VIII to X. Chapter XI extends the duality diagram to accommodate two sets of properties or variables as follows:

$$\begin{array}{ccccccc} & & \mathbf{X} & & \mathbf{Y} & & \\ & E_1 & \longleftarrow & F^* & \longrightarrow & E_2 & \\ \mathbf{I} & \downarrow\uparrow \mathbf{V}_{11} & & N \uparrow & & \mathbf{V}_{22} \uparrow\downarrow & \mathbf{I} \\ & E_1 & \longrightarrow & F & \longleftarrow & E_2^* & \\ & & \mathbf{X}^t & & \mathbf{Y}^t & & \end{array}$$

It is a natural consequence of this layout that canonical correlation analysis is a matter of locating the eigenvalues and right eigenvectors of $\mathbf{X}'\mathbf{V}_{11}^{-1}\ \mathbf{XNY}'\mathbf{V}_{22}^{-1}\ \mathbf{YN}$. Chapter XII applies these principles to multiple discriminant analysis. Chapter XIII considers the corresponding problem when both sets of variables are quantitative, referred to in France as

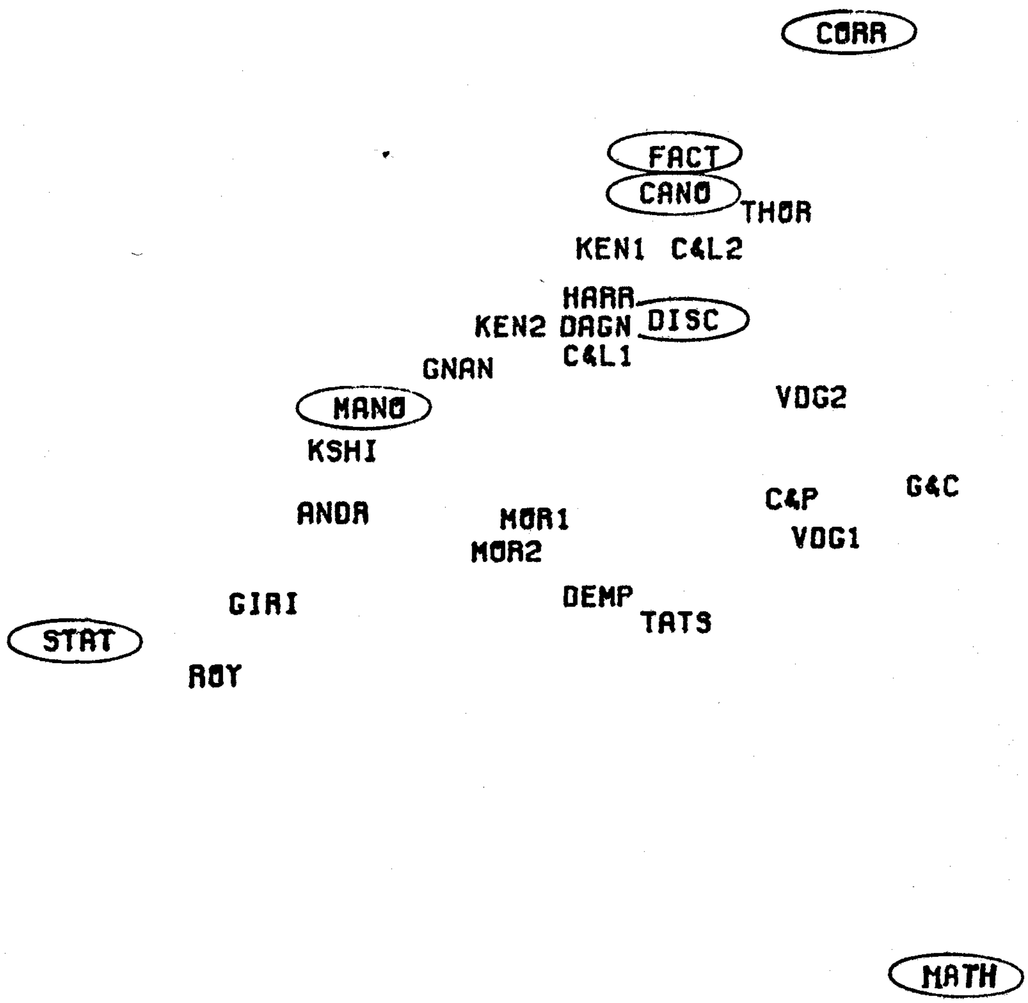

FIGURE 1.
A representation of twenty multivariate statistics texts and seven content areas by correspondence analysis. Full names are given in Tables 1 and 2. Content areas are enclosed in ellipses. The data were the numbers of pages each text devoted to the content areas, and correspondence analysis represents each text as a weighted average of content area positions.

correspondence analysis [Benzécri, 1973] and elsewhere as optimal or dual scaling [Nishisato, 1980]. Chapter XIV discusses other approaches and treats briefly the problem of data in three-dimensional arrays. The final chapter covers classification techniques.

Although the approach is geometrical and algebraic (as opposed to probabilistic), this book is also in effect an application of functional analysis, and provides a reasonable first contact with elementary Hilbert space theory. Along with other introductory works such as that of Dieudonné [1960] or Kreyszig [1978] it could provide an overview of a branch of mathematics that we feel will prove of fundamental importance in data analysis, as it already has in numerical analysis. In effect, there is little in the duality diagram that cannot be transferred to the analysis of other Hilbert spaces such as square-integrable continuous functions [Ramsay, 1982]. Much generalization of this sort is already in progress in France although still in a relatively inaccessible format. Dauxois and Pousse [1976] apply Hilbert analysis much more deeply, Besse [1979] and Saporta [1981] have investigated the principal components analysis of continuous functions, and Jaffrennou [1978] has taken the analysis of three-mode data arrays much further.

Gifi [1981] provides an interesting analysis relating this book to a number of other well known multivariate statistics books. He analyzes data consisting of the number of pages devoted to seven subject categories using correspondence analysis. Table 1 lists the books in the analysis and Table 2 describes the subject categories. Figure 1 plots the books and subject categories on a plane. According to the barycentric principal in correspondence analysis, each book is a weighted mean of subject locations, and thus falls within the triangle defined by MATH, STAT, and CORR. Between STAT and CORR are to be found traditional multivariate statistics books, while between CORR and MATH are to be found texts that might be referred to as data analytic since they use a minimum of statistical theory. Thus the book by Cailliez and Pages is to be found with other data analytic texts such as the two editions of Van de Geer, although there are certainly important differences between it and these others.

In summary *Introduction á l'Analyse des Donńees* offers a treatment of the multivariate linear model which is (a) metric and basis free, (b) offers a unified survey of both quantitative and certain qualitative procedures, (c) incorporates classical multidimensional scaling in a natural way, and (d) invites through its powerful formalism an extension in a number of valuable directions. We hope it will not be long before an English language counterpart appears.

MCGILL UNIVERSITY *J. O. Ramsay*
UNIVERSITY OF LEIDEN *J. de Leeuw*

REFERENCES

Anderson, T. W. *An Introduction to Multivariate Statistical Analysis.* New York: Wiley, 1958.

Benzécri, J. P. *L'Analyse des données II: L'Analyse es Correspondances.* Paris: Duneod, 1973.

Besse, P. Etude descriptive d'un processus. Thése de troisieme cycle, l'Université Paul-Sabatier de Toulouse, France, 1979.

Cailliez, F. & Pagés, J. P. *Introduction à l'Analyse des Données.* Paris: Societé de Mathématiques Appliquées et de Sciences Humaines, 1976.

Cooley, W. W. & Lohnes, P. R. *Multivariate Procedures for the Behavioral Sciences.* New York: Wiley, 1962

Cooley, W. W. & Lohnes, P. R. *Multivariate Data Analysis.* New York: Wiley, 1971.

Dagnelie, P. *Analyse Statistique à Plusieurs Variables.* Gembloux: Presses Agronomiques, 1975.

Dauxois, J. and Pousse, A. Les analyses factorielles en calcul des probabilités et en statistique: Essai d'étude sythétique. Thése d'état, l'Université Paul-Sabatier de Toulouse, France, 1976.

Dempster, A. P. *Elements of Continuous Multivariate Analysis.* Reading, Mass.: Addison-Wesley, 1969.

Dieudonne, J. *Foundations of Modern Analysis.* New York: Academic Press, 1960.

Gifi, N. C. *Multivariate Statistical Inference.* New York: Academic Press, 1977.

Gnanadesikan, R. *Methods for Statistical Data Analysis of Multivariate Observations.* New York: Wiley, 1977.

Green, P. E. & Carroll, J. D. *Mathematical Tools for Applied Multivariate Analysis.* New York: Wiley, 1976.
Harris, R. *A Primer of Multivariate Statistics.* New York: Academic, 1975.
Jaffrennou, P. A. Sur l'analyse des familles finies de variables vectorielles: Bases algebriques et application a la description statistique. Thése de troisieme cycle, l'Université de Saint-Etienne, France, 1978.
Kendall, M. G. *A Course in Multivariate Analysis.* London: Griffin, 1957.
Kendall, M. G. *Multivariate Analysis.* London: Griffin, 1975.
Kreyszig, E. *Introductory Functional Analysis with Applications.* New York: Wiley, 1978.
Kshirsagar, A. M. *Multivariate Analysis.* New York: M. Dekker, 1972.
Morrison, D. F. *Multivariate Statistical Methods.* New York: McGraw-Hill, 1967, 1976.
Nishisato, S. *Analysis of Categorical Data: Dual Scaling and its Applications.* Toronto: University of Toronto Press, 1980.
Ramsey, J. O. When the data are functions. *Psychometrika*, 1982, *47*, 379–396.
Roy, S. N. *Some Aspects of Multivariate Analyses.* New York: Wiley, 1957.
Saporta, G. Methodes exploratoires d'analyse de donnees temporelles. Thése d'Etat, l'Université Pierre et Marie Curie, Paris, France, 1981.
Tatsuoka, M. M. *Multivariate Analysis: Techniques for Educational and Psychological Research.* New York: Wiley, 1971.
Thorndike, R. M. *Correlational Procedures for Research.* New York: Gardner, 1978.
Van de Geer, J. P. *Inleiding in de Multivariate Analyse.* Arnhem: Van Loghem Slaterus, 1967.
Van de Geer, J. P. *Introduction to Multivariate Analysis for the Social Sciences.* San Francisco: Freeman, 1971.