E. Diday, Y. Escoufier, L. Lebart, J. Pagès, Y. Schektman and R. Tomassone, editors, *Data Analysis and Informatics, IV*, Proceedings of the 4th International Symposium on Data Analysis and Informatics, Versailles (France), October 9-11, 1985. Amsterdam: North-Holland, 1986, pp. xiii + 764.

This book contains the proceedings of the Fourth International Symposium on Data Analysis and Informatics, held in Versailles (France) in 1985. The purpose of the Symposium was to present and discuss recent advances in both the theory of data analysis and its practice (computing aspects). Data analysis consists essentially of the multivariate methods of descriptive statistical analysis; it is defined more specifically by the Editors, in the Foreword, as the set of methods aimed at extracting "useful information that can be easily apprehended by the user."

The book contains 64 papers, among which are 15 invited contributions. I will not attempt to review them all, but will limit this review to pointing out those that I found to be of greatest interest to the readership of this Journal. The book is bilingual, with 36 papers in French and 28 in English. Most French papers contain an abstract in English, but there is no reciprocity for the papers in English.

Chapter 1 — Clustering (12 papers). M. Roux presents a new algorithm for computing a network among points, by gradually modifying the distances in such a way as to satisfy the 4-point metric condition. B.G. Mirkin reviews three methods for qualitative-data factor analysis that had previously been developed by his collaborators and himself, and were originally published in Russian. He extends their use to the problem of additive clustering. New algorithms for clustering large data sets are presented by M. Reinert, and by I.C. Lerman and P. Peter. C.J.F. ter Braak proposes to relate hierarchical classifications to external information, by computing a simple equivalent to discriminant functions at each node of the tree. He argues that discriminating variables are more useful when they are attached to specific nodes than when they are computed for all clusters simultaneously. Finally, R.R. Sokal reviews methods of spatial autocorrelation analysis, used to infer causal processes from observed patterns of spatially distributed variables. The review covers

Reviewer's Address: Pierre Legendre, Département de sciences biologiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec, Canada H3C 3J7.

single-variable spatial correlograms, the Mantel test, the comparison of correlograms by clustering, and introduces for the first time the idea of multivariate correlograms based on Mantel statistics.

Chapter 2 — Factor analysis (8 papers). H. Caussinus argues that probabilistic models can turn out to be useful even in data analysis; this is the case for instance in principal components analysis and in derived methods. G. Bove proposes a procedure that improves the representation of those modalities that are badly represented in correspondence analysis plots. B. Escofier and J. Pagès present an extension of multiple factorial analysis (MFA) to qualitative variables; MFA was introduced in 1983 by these same authors as a method for comparing groups of variables. In a paper that presents partial correspondence analysis in the framework of canonical correlation analysis, H. Yanai proposes to apply redundancy coefficients to correspondence analysis. P.G.M. van der Heijden shows that correspondence analysis is a suitable method for the analysis of transition matrices, especially when the number of categories is large. Other papers deal with specialized aspects of correspondence analysis and with nonlinear factor analysis.

Chapter 3 — Discrimination (6 papers). Fuzzy sets are highlighted in the first two papers. J. Aguilar-Martin and N. Piera I Carrete propose to adapt a form of fuzzy clustering, with "connective" operators, to self-teaching classification algorithms. J.L. Mallet et al. present a nonlinear form of discriminant analysis based on fuzzy characteristic functions.

P. Baufays and J.-P. Rasson propose to use, in discriminant analysis, an assigning rule whose computation is performed by a geometric analysis algorithm; the new rule is described and compared to more classical methods. P. Meunier et al. describe a method to solve the "classification" problem in discriminant analysis (building identification schemes or keys); they propose a new criterion well adapted to the treatment of ordinal dependent variables.

V.V. Alexandrov et al. describe a computer package for statistical analysis, multidimensional analysis and pattern recognition, developed in the USSR. This package presents an "expert system" interface to the user. I would like to challenge the authors' claim that users of such a system can get the most out of their data without knowledge of ANY data processing methods. Experience in teaching data analysis shows, on the contrary, that the more the users in specialized fields of application understand the methods they use, the more they become able to formulate and test informative hypotheses; no expert system can replace scientists at the task of generating pertinent hypotheses.

Chapter 4 — Multidimensional scaling (6 papers). A.W.P. Canton proposes a procedure to construct an index of reliability of decision rules for problems of ranking and selection of the "best" out of a number of groups; such problems arise for instance in plant breeding, in the pharmaceutical

industry, in manufacturing, etc. S. Nishisato summarizes the method of
"forced classification" that he had proposed in 1984, and answers a number
of previously unresolved questions. Forced classification is a correspondence
analysis method in which the cloud of data points can be oriented in any par-
ticular direction the user chooses. The method is here generalized to data of
all precision levels, through a "quantification" procedure. A. Di Ciaccio pro-
poses a new association index among categories, that is easy to compute even
for a large number of categorical variables; he uses the resulting association
matrix for multidimensional scaling, and obtains results that differ from those
of multiple correspondence analysis. J.-B. Kazmierczak proposes a new form
of scaling, called "logarithmic analysis," based on a metric that obeys Yule's
principle of distributional equivalence. This method presents the advantage of
producing equivalent ordinations, whatever the variable is that one uses to
scale all the others. An econometric example is presented that demonstrates
the interest of the new method.

Chapter 5 — Specific data arrays (5 papers). This chapter deals with
special types of data tables, and presents two scaling methods that are of par-
ticular interest. R. Lafosse presents a new stepwise Procrustes analysis pro-
cedure for pairs of data tables, while A. Carlier has adapted to three-way data
tables Le Foll's method of "evolutionary factor analysis" designed to take
into account the contiguity of samples along time, or along some other order-
ing axis.

Chapter 6 — Linear model problems (3 papers). W.J. Heiser discusses
the problem posed by undesirable nonlinearities that can show up even in
nonlinear multivariate analysis methods. Their emergence can be traced back
to two different sources: coding, or the Guttman effect (horseshoe). A. Leroy
and P.J. Rousseeuw propose the method of "least median squares" (LMS)
regression, a robust technique that reduces the influence of outliers on param-
eter estimation.

Chapter 7 — Inference (5 papers). G. De Soete and J.D. Carroll
demonstrate how to fit multidimensional choice models to paired comparison
data. G. Saporta and G. Hatabian draw confidence ellipses around states of
categorical variables in correspondence analysis plots. J. de Leeuw and E.
van der Burg use a permutation (resampling) method to compute the
significance probability of goodness-of-fit statistics obtained in generalized
canonical correlation analysis. In order to visually estimate the effect of
measurement accuracy, J. Benasseni performs principal component analysis
of fuzzy-recoded data. J.-C. Deville presents a classification of the statistical
models used in socioeconomic studies.

Chapter 8 — Data processing. This chapter presents 10 applications of
data analysis to a variety of problems: cross-cultural data, drawing a socio-
economic map, low back pain, heart conditions, drug usage, psycho-

sociological surveys, content analysis of survey data, work accidents, geo-chemistry, and bird community analysis.

Chapter 9 — Informatics (9 papers). Reports on data analysis software for microcomputers are presented by K. Yajima and N. Ohsumi for Japan, and by J.L. Chandon for France. J. Lemaire discusses the importance of certain informatic choices that have to be made by computer scientists who plan to write statistical software. W.A. Gale reviews and criticizes expert systems intended to help naive users to make correct decisions prior to statistical data analysis, while J.R. Barra and M. Becker propose such an expert system to help decide among a variety of ordination methods. I.T. Jolliffe proposes to represent by a dendrogram the results of multiple comparisons among group means. F. Bry and G. Thauront discuss characteristics that statistical data base management systems should have; such a system, integrating data base management and statistical analysis, is presented by G. Jomier et al.

Books of this kind are a useful means of communicating what researchers are currently working on and, for the reader, of getting some introduction to the relevant litterature. It is unfortunate that no subject index is present; an index is always useful in books compiling such a diversity of topics.

Université de Montréal   Pierre Legendre