EDITORS' INTRODUCTION

Jan De Leeuw
Willem Heiser
Jacqueline Meulman
Frank Critchley

This volume contains the proceedings of the workshop Multidimensional Data Analysis, which was held in Pembroke College in Cambridge from june 30 to july 2 1985, just before the 4th European Meeting of the Psychometric Society and the Classification Societies.

The topic we selected was, for various reasons, Multidimensional Data Analysis. In the first place these techniques form, of course, one of the cornerstones of psychometrics. In the second place we, the organizers, were all active in the field of multidimensional scaling and multivariate analysis, and we knew the relevant literature and the people active in the field quite well. In the third place we wanted to use the opportunity to organize, within the meeting, another one of the recently popular encounters of French data analysis and Anglo-American statistics. And in the fourth place we felt that in the recent upsurge of structural covariance modelling, log-linear modelling, and simple graphical data analysis, the classical multidimensional techniques have sometimes been forgotten, or at least neglected.

We decided on the following format. Four main topics were selected, which seemed to cover the field, and which seemed to be quite active at the time. These were

- Algebraic Aspects of Distance-based Methods
- Algebraic Aspects of Scalar Product and Biplot Analyses
- Probabilistic and Statistical Aspects
- Classification and Seriation

For each topic two main speakers were invited, plus two main discussants. An additional 30 experts in the field of multidimensional data analysis were invited, mainly of course from the psychometric community, and many of them decided to contribute short talks, which were organized into 3 sessions. During the composition of this book, we gave the main speakers the opportunity to react to the points made by the discussants

of their papers. The contributions of the main speakers and discussants are all presented in these proceedings, along with abstracts of the contributed papers.

In this introduction we shall briefly summarize the main contributions, and try to make some additional points which we think to be important. Topic 1 was **Algebraic Aspects of Distance-based Methods**, with main speakers John Gower and Bernard Fichet, and with discussants Frank Critchley and Willem Heiser. The topic was chosen because of its obvious relevance for multidimensional scaling, a field which is relatively algorithm-rich and theory-poor. Fichet and Gower discuss various similarity and dissimilarity indices in terms of the metric axioms, and in terms of their imbeddability into Euclidean space. Gower also pays attention to the question what happens if we apply metric multidimensional scaling methods to non-Euclidean distances, or even to non-Euclidean non-distances. Both authors review the classical embedding theorems due to Fréchet, Schoenberg, and Young-Householder. Fichet also contributes to the Anglo-French dialogue by relating embedding of distances to principal component analysis and correspondence factor analysis, using the famous duality diagram. Heiser concentrates, in his discussion, on properties of classical scaling and the relationships with correspondence analysis. Critchley has some useful additional results, both on the families of similarity measures and on the metric scaling problem. The papers, with discussion, give a very complete coverage of the results in this area, which is at least partly due to the fact that the area is a well-defined one, which is not too broad.

The second main topic was **Algebraic Aspects of Scalar Product and Biplot Analyses.** The main speakers were John Van de Geer and Ruben Gabriel. This area is much broader, and much more work has been done on the topic in the past. Consequently the main speakers had to make a selection, choosing the material with which they were most familiar. Van de Geer talked about relations between K sets of variables, using his characteristic geometrical approach. Gabriel presented the three-dimensional Biplot, used to diagnose various simple algebraic models generating the data. Jos ten Berge discussed Van de Geer's paper, and contributed a general algorithm to solve the optimization problem involved in the the maximization of the matching criteria. Michael Greenacre, in his discussion of Gabriel's paper, related the biplot to correspondence analysis and classical scaling, another Anglo-French interaction.

The main speakers for the third topic, **Probabilistic and Statistical Aspects**, were Yoshio Takane with Tadashi Shibayama, and Henri Caussinus. Of

course the topic was intended to cover probabilistic and statistical aspects of multidimensional data analysis, but even with this specification the speakers were left with a very wide area. Takane talked about distance based maximum likelihood methods for the analysis of stimulus recognition data (or, more generally) for square frequency tables. His methods are multidimensional extensions of the quasi-symmetry model, which was introduced into statistics by Henri Caussinus more than 20 years ago. Caussinus himself talked about a much more philosophical topic, the role of models in principal component analysis. Again this is important because the role of models are perceived quite differently in French data analysis than in Anglo-American statistics. In his discussion of the Takane and Shibayama paper, Ivo Molenaar makes several interesting remarks about the comparison of models in general, and he discusses a simple nonparametric model as a possible alternative for the complicated parametric models. De Leeuw discusses the Caussinus paper, finding himself largely in agreement with Caussinus' discussion of the role of models. He also proposes a very general factor analysis model, which has the models discussed by Caussinus as special cases.

The fourth and final main topic was **Classification and Seriation.** Because of the Cambridge meeting there were many eminent classifiers and clusterers around, and we invited Edwin Diday and Larry Hubert, with Phipps Arabie, as our main speakers. Geert de Soete discussed the Hubert-Arabie paper, and Phipps Arabie, who confounded our design because Hubert chose him as a co-author and we chose him as a discussant, discussed Diday's paper. The Hubert-Arabie paper was about unidimensional scaling, which is a very special technique because the problem it tries to solve looks like an ordinary continuous optimization problem, but turns out to be basically combinatorial. Unidimensional scaling is a clear example of seriation. De Soete discussed additional computational aspects of this scaling problem. Diday talked about Pyramids, which are elegant French generalizations of the more familiar hierarchical trees. Many aspects of Pyramids were presented, and as Phipps Arabie pointed out in his discussion, they are very promising tools for data analysis, certainly because good software and graphics are available.

We end this introduction by thanking various persons. In the first place the staff of Pembroke College, who were very helpful and unfailingly cheerful throughout all the local arrangements. In the second place the authors of abstracts, main papers, and discussions, who (as a rule) observed our deadlines, or at least sent in their contributions very rapidly after some prompting. In the third place, the Royal Society of London, for providing an enabling grant. And in the fourth place the Repro-dienst

Psychologie and the Secretarial Department of the DSWO, who were a great help in the preparation and production of the various versions of this book. We, the editors, thoroughly enjoyed both the workshop, and putting together this collection of excellent papers.