# Book Review

## Brigitte Le Roux and Henry Rouanet (with a foreword by Patrick Suppes): Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis

**Dordrecht: Kluwer, 2004. 475 pages, 155 Euros**

To many sociologists, a book with the title Geometric Data Analysis may at first seem of little relevance to their own work. The subtitle, however – From Correspondence Analysis to Structured Data Analysis – should indicate otherwise. As is well known, multiple correspondence analysis was the preferred statistical tool of the late Pierre Bourdieu, with whom the two authors collaborated from the mid-1990s and onwards. The book gives a state-of-the-art introduction not only to correspondence analysis, but also, through several examples, to dealing with problems met in practical multivariate analysis using methods for analysing latent structures in the data. The book is organised into 10 chapters, covering all the fundamental issues in geometric data analysis.

Chapter 1, Overview of Geometric Data Analysis, gives a short review of the history of correspondence analysis, a general introduction to the chapters to come, and also some very clear messages about the book's epistemological foundation. Citing Ludovic Lebart, the authors make it clear that 'Statistics does not explain anything – but provides potential elements for explanation' (pp. 19–20). The widely used and misleading opposition between so-called exploratory and explanatory techniques is thus strongly refuted. The same goes for the understanding that correspondence analysis and statistical inference exclude each other (pp. 18–19).

Chapter 2, Correspondence Analysis, starts with a presentation of the 'strong alliance between the mathematical underlying structures and their geometrical representations' (p. 22). Most sociologists with little or no background in mathematics will find these sections hard to follow, and may benefit from first reading the sections on how to interpret the results from the analysis, and thereafter return to the mathematical sections. Le Roux and Rouanet's strategy of interpretation is based on the method of contributions of points and deviations, developed by the authors themselves, where the principles of correspondence analysis are combined with variance analysis. Towards the end of the chapter, a gauntlet is thrown down, this time to the dominant modelling tradition in categorical data analysis. Correspondence analysis, the authors make clear, can also be presented as a probabilistic model. If one keeps all the axes, one gets a saturated model. Retaining fewer axes gives an unsaturated model. Where the interpretation of log-linear models and association models is based on cross-product ratios, the interpretation of correspondence analysis is based on correlations, which the authors find easier to interpret. Whether or not researchers trained in the log-linear modelling tradition agree with the latter claim remains an open question.

Chapters 3 and 4 – Euclidian Cloud and Principal Components Analysis – follow a similar structure. First, the mathematical and statistical properties of the methods are introduced. Thereafter, the principles of interpretation are outlined. Unfortunately, the example in Chapter 3 does not seem to be structured around 'real-life' data, as this reviewer would have liked, if only for pedagogical purposes. In Chapter 4, principal components analysis is 'recast as a geometrical method' (p. 129) for the analysis of metrical data. As an advantage over standard principal components analysis, the geometrical approach makes it possible not only to study the relations between the active variables in the analysis, but also the cloud of individuals. Thus, the interpretation is based on a simultaneous exploration of the space of variables and the space of individuals. For a more detailed demonstration of these principles, however, the reader must go to the analysis of the Parkinson data set in subchapter 9.1. Although this is only a minor inconvenience, the contrast to the strategy followed in Chapter 5 – Multiple Correspondence Analysis (MCA) – is striking.

After first going through the mathematical properties of the analysis, the cloud of modalities and the cloud of individuals, the authors here give a demonstration of MCA and its interpretation, using survey data from 1997 about cultural practices in France as

the example. From section 5.3 onwards, readers are guided carefully through all the important steps in the analysis of survey data in a straightforward way, and also shown how to construct their own spaces from survey data. Even though one could ask for a more detailed section on cloud shapes and data structures, to this reviewer, these 35 pages are simply a masterpiece and merit a wide readership in the social sciences.

On the whole, Chapter 5 stands out as being state-of-the-art both pedagogically and methodologically. Three 'tools' developed by the authors themselves are also introduced: specific MCA, the combination of MCA and variance analysis, and the use of concentration ellipses in the cloud of individuals. Compared to standard MCA, where rare modalities in active questions, for instance 'Don't know' or 'No response', are defined as illustrative modalities, specific MCA instead defines these as *passive* modalities, i.e. they are ignored in the calculation of the distances between the individuals. This is more than just a detail, not least because it makes the combination of MCA and variance analysis mathematically justifiable. Whereas the coordinates of the overall centre of gravity may not always be located at 0.0000 in a standard MCA with illustrative modalities from active questions, this will always be the case in specific MCA. In this way, the examination of within- and between-class variance can be used in the examination of the cloud of individuals, as the authors do in the chosen example. Finally, concentration ellipses around the mean points in the cloud of individuals reveal intersections and separations between the various sub-groups. Each of these elements is an important contribution to data analysis with MCA. In combination, they constitute three cornerstones in an integrated paradigm baptized Structured Data Analysis.

In Chapter 6, Structured Data Analysis, the epistemological fundamentals of this paradigm are further underlined, stressing the differences between observational and experimental data and approaches, and with a short subchapter on Structured Data Analysis and Regression. Given that the authors have worked extensively on this topic in recent years, one regrets their decision to devote only $1^1/_2$ pages to this topic, especially considering the dominant position regression analysis continues to enjoy in the social sciences.

Whereas Chapter 7 deals with stability problems and how to tackle them, Chapter 8 – Inductive Data Analysis – lays down the final epistemological cornerstone of the framework. Outlining the difference between traditional inference, combinatorial inference and bayesian inference, the authors also introduce confidence ellipsoids as a tool for drawing inferential conclusions about the positions of modalities in the factorial plane.

For an empirical demonstration of these procedures, the reader must once again turn to Chapter 9 – Research Case Studies – which contains the three case studies that are referred to throughout the book: the Parkinson Study, the French Political Space and the Education Program for Gifted Youth (EGPY) study. As Patrick Suppes points out in the foreword, this chapter is almost a book in itself, and can also be read separately from the other chapters. To most social scientists, the study of the French electoral data (pp. 365–94) will probably have the strongest relevance to their own work. All four steps in the overall data analysis philosophy are carefully demonstrated in a pedagogical *tour de force*, and readers with little or no background in statistics could even be recommended to start with this chapter, and thereafter turn to the more mathematically demanding preceding chapters. Finally, in Chapter 10, the mathematical bases of the covered subjects are presented.

All in all, I find this book to be a treasure chest, and social scientists working with survey data will particularly find Chapter 2 and Chapters 5–9 to be highly instructive. Unlike most textbooks in statistics, and to this reviewer's joy, some of the fundamental epistemological questions in the discipline are also raised throughout the book. These topics have had a tendency either to be the subject of separate studies, as for instance in Ian Hacking's or Alain Desrosières' work on the history of statistics, or to be dealt with in rather superficial ways prior to the demonstration of a given statistical technique. In any case, they have too seldom found their way into the curriculum in classes in statistics. This book invites the lecturer to do otherwise.

At the same time, it should also be said that parts of this text are demanding to read. Not being written with the social scientist in mind, non-mathematicians and non-statisticians may be put off by the mathematical complexity at the beginning of most of the chapters. Even so, sociologists and political scientists used to thinking in terms of spatial metaphors will without doubt find this book to be very useful.

Unfortunately, with a price of 155 euros, few students can afford to buy it. Also, the book is probably too complicated to be used as an introductory text to

undergraduate (and perhaps even graduate) students in the social sciences. Next to Geometric Data Analysis therefore, there is still room for a new, shorter, introductory book to GDA on my shelf, written with sociology students in mind, and with a price tag they can afford. In conclusion, a new, introductory volume in the Sage-series Quantitative Applications in the Social Sciences is called for, based on the template used in Chapters 5 and 9, and written by these two authors.

**Johs Hjellbrekke**
*Department of Sociology, University of Bergen, Norway*
*johs.hjellbrekke@sos.uib.no*

DOI:10.1093/esr/jci035, available online at www.esr.oxfordjournals.org