Review of Michael J. Greenacre, Theory and Application of Correspondence Analysis

Gilbert Saporta

Review of : Michael J. Greenacre, Theory and Application of Correspondence Analysis, London: Academic Press, 1984

Dedicated to Jean-Paul Benzécri, a leader of the French school of data analysis, this book presents very comprehensively a method of data reduction and graphical display which is used extensively in France but not so much elsewhere.

The publication of this book is thus an important event for it provides a large audience access to many works available until now only in French.

Note that this book is not the only one devoted to this topic in English since, by a lucky coincidence, a translation of a book by Lebart and Morineau also was published recently.

After a short introduction, Chapter 2 is devoted to a presentation of the geometrical concepts used in the book. The mathematical treatment is simple and the stress is put upon weighted Euclidean spaces and the chi-square distance.

Chapter 3 is of the same style and simply presents the output of a correspondence analysis of a contingency table.

Chapter 4 deals with the mathematics of correspondence analysis: it is of special interest both historically and methodologically. Connections with various techniques (canonical correlations, dual scaling) are shown, and the reader will be convinced that correspondence analysis is one of the main statistical methods. The fundamental formulae were discovered very early (in the thirties) but its use as a powerful tool of data analysis stems from the graphical potentialities emphasized by Benzécri. Section 4.6 provides useful comments on various mathematical properties.

Chapter 5 presents multiple correspondence analysis, an extension of correspondence analysis to multivariate nominal data using only bivariate marginals; though well documented this Chapter suffers from some gaps. The pioneering work of Guttman (1941) quoted in Chapter 4 should have been presented here as well as a list of optimal properties of the components. The work of the Dutch group in homogeneity analysis is also ignored. A useful reference for the interested reader is Tenenhaus and Young (1985) but the most important gap concerns the use of supplementary variables, which is one of the main features of the French style for analyzing questionnaires (see Lebart, Morineau and Warwick 1984).

Chapters 6 and 7 deal with nonstandard uses of correspondence analysis (preference or rating data, regression, discrimination, clustering) and demonstrate its relevance in various contexts, though the superiority claimed for correspondence analysis over classical techniques might be discussed at greater length.

Chapter 8, "Special Topics," is a compendium of less-known properties and applications. The most original points are about stability and statistical inference by using jackknife and bootstrap techniques. It can be argued whether the best way of examining the variability of graphical displays is the use of convex hulls for the replicates, since the hulls are not at all robust.

One of the charms of this book is in the great variety of worked examples presented. Chapter 9 gives in addition eleven case-studies of correspondence analysis in various fields. However, the reader might get the erroneous impression that two-dimensional displays are always sufficient. Advice for choosing the right number of dimensions is also missing.

An extensive list of references (nearly 300) ends the book, but the choice of the French ones is surprisingly biased: they all belong to the Benzécri team (83 references from the Cahiers de l'Analyse des Données) though numerous books and papers (some in English) have been written by authors from other schools. In particular, the approach based on the duality diagram due to Cailliez and Pagès (1976) is ignored.

In summary, this book is a welcome one and I recommend it to anyone willing to learn what is behind correspondence analysis and how to use it.

References

CAILLIEZ F., and PAGES, J.P. (1976), Introduction à l'Analyse des Données, Paris: Société de Mathématiques Appliquées et de Sciences Humaines.

GUTTMAN, L. (1941), "The Quantification of a Class of Attributes: A Theory and Method of Scale Construction," in The Prediction of Personal Adjustment, Ed. The Committee on Social Adjustment, New York: Social Science Research Council, 319-348.

LEBART, L., MORINEAU, A., and WARWICK, K. (1984), Multivariate Descriptive Statistical Analysis, New York: Wiley.

TENENHAUS, M., and YOUNG, F.W. (1985), "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis, and Other Methods for Quantifying Categorical Multivariate Data," Psychometrika, 50, 91-119.

Gilbert Saporta