Review of I. Borg & J. Lingoes, "Multidimensional similarity structure analysis"

Kroonenberg, P.M.

PSYCHOMETRIKA—VOL. 55, NO. 1, 169–172
MARCH 1990
REVIEWS

REVIEWS

I. Borg and J. Lingoes. *Multidimensional similarity structure analysis*. New York/ Berlin; Springer-Verlag, 1987, xii + 390 pp.

Preliminaries

The title of this book suggests that it covers an entirely new set of techniques, that it deals with the analysis of similarities otherwise known as multidimensional scaling, or that it deals with a variant of smallest space analysis (SSA) under a new name. Apparently, the authors had the second possibility in mind, especially considering the definition of the title: "Multidimensional similarity structure analysis (SSA) comprises a class of models that represent similarity coefficients among a set of objects by distances in a multidimensional space". The authors continue that "[h]istorically, the search for such dimension[al system]s was so prevalent that it almost seemed as if there could not exist any other regularities in the point spaces. Consequently, SSA was termed 'multidimensional scaling' (MDS)", but this term "seemed to us so unfortunate for a general understanding of the purpose of these data analysis techniques that we decided to use the uncommon term 'SSA' instead". To my mind this is highly unfortunate, because it has the (unintended?) effect of implicitly claiming that the whole field of what is commonly understood to be MDS, is a derivative of the authors' (and their close associates') smallest space analysis (SSA).

The book opens with a preface called "Introduction", which contains the above quotations and a diagram how the chapters may be read if one does not want to read them sequentially. A proper introduction into the field and its ramifications is lacking, as well as an overview of what the book is going to be about and for whom it is intended, apart from the statement that "we are dealing with all aspects of SSA: the general data-analytic perspective, the MDS point-of-view, and the distance-formula-as-composition-rule approach". Each chapter is prefaced by a short abstract of its contents and an (extensive) list of keywords, eliminating according to the authors the need for an overview of the contents of the book. I beg to differ.

Contents

Chapter 1 (Construction of SSA representations) starts with the by-now famous reconstruction of a map from a distance table (Germany, this time). This is done by stepwise introducing new points (cities) into a two-dimensional plane. The ideas of (similarity) transformations of a configuration are introduced next, followed by a similar introduction to map reconstruction for ranked similarities. Chapter 2 (Ordinal SSA by iterative optimization) continues this discussion with the detailed and computationally oriented explanation of an iterative rank-image optimization procedure for SSA. Next, in Chapter 3 (Monotone regression), with equally great computational detail, an "alternative" (i.e., Kruskal's) algorithm is presented based on monotone transformations of distances to optimally fit disparaties in a disparaties versus distances (or Shepard) plot. Missing data and the possible treatments of ties is discussed as well.

In Chapter 4 (SSA models, measures of fit, and their optimization) some SSA models are defined, primarily by defining the type of (permissible) transformation which relate the proximities (or better disparaties) to distances, leading to absolute, ratio, linear, metric, nonmetric, and nominal SSA. Various badness-of-fit measures are defined, such as Kruskal's stress for the monotone regression approach and Guttman's alienation coefficient for the rank image approach. On the basis of these error measures, the basic iterative procedure for solving an SSA model is defined by the "2-phase algorithm": Step 1—given coordinates, compute distances, and find the optimal transformation of the disparities; Step 2—given the target distances, find an optimal set of coordinates. Via differentiation of the loss function, a gradient algorithm is developed to solve the minimization problem.

These four chapters form the root from which (sets of) chapters are branching. They concern primarily the technical problem of how to construct an optimal SSA representation, and may be skipped, according to the authors, by readers with more background knowledge, or by readers with relatively applied interests.

Chapter 5 (Three applications of SSA) give three examples of SSA—Ekman's color data, Rothkopf's Morse code data, and Engen et al.'s data on the Schlosberg series. The examples are treated in considerable detail and several practical issues are addressed, such assessing stress coefficients, choosing dimensionalities, relationships with substantive theories, and fitting solutions to external scales. In the same vein, chapter 6 (SSA and facet theory) presents Guttman's facet theory, primarily via several extensive analyses of studies of wellbeing, introducing the ideas of a simplex and radex structure.

Chapter 7 (Degenerate solutions in SSA) and chapter 8 (Computer simulation studies on SSA) discuss what can go wrong with SSA analyses and how error and subsampling of dimensions influence solutions. Such knowledge allows evaluation of an obtained solution against simulation results.

In chapters 8, 9, and 10 primarily the situation where only proximities between sets but not within sets are available, that is, unfolding, and a few specialized topics are discussed. In chapter 12 and 13, confirmatory SSA is treated, and Minkowski distances are the focus of attention in chapter 14, while—finally—in chapter 15 "true" multidimensional scaling comes on the scene, where MDS is defined as the generalization of unidimensional scaling, that is, a procedure for constructing a multidimensional scale system for a given set of data. In particular, a city-block metric example and a rectangle-comparison example are used to illustrate the idea of physical and psychological judgemental dimensions.

The final series of chapters move on to more advanced and complex topics. In particular, scalar products (chapter 16), matrix algebra (chapter 17), relationships between data, proximities, distances, and the additive constant problem (chapter 18) are treated. Most of the subjects so developed are necessary to understand the inner working of the methods to follow. The treatment of these topics is done in the by-now familiar elementary and computational manner, which should make it fairly easy to follow for most readers.

The final two chapters, chapter 19 (Procrustes Procedures), and 20 (Individual differences models) discuss models and methods for designs which lead to more than one similarity matrix. The main emphasis in these chapters is the authors' work in this area leading up to their Procrustean INdividual DIfferences Scaling (PINDIS), and ending with a short discussion of dimensional-weighting models, such as INDSCAL. In passing, differentiation of matrix traces and differentiation under side constraints, and indices for measuring configurational similarities are taken up as well.

Some General Comments

The book does not have a very strong *leitmotiv*, or at least it is not presented. The chapter structure diagram on page xiv shows a rather flat profile, suggesting that the book is primarily an integrated collection of papers, which together cover the intended field of description. This impression is reinforced by limited internal cross-referencing, and the (as such useful) abstracts and keyword lists with which the chapters open.

The approach to explanation in the book is generally computational rather than conceptual, which does not always make for easy reading. Because of this, the writing is fairly demanding, notwithstanding the relatively elementary level of explanation. The authors assume that the reader needs no more than a high school background in mathematics, at least to get started. This may be true but a considerable aptitude and appreciation for mathematics will certainly be helpful, and probably necessary for the later chapters. After reading the book one should have a fairly good working knowledge of the linear algebra necessary to read most papers in the field. All through the text new mathematical techniques are introduced and explained, both in examples and in separate sections. Their scattered locations seem to make the book less fit as a textbook. On the other hand, the wealth of topics and the large number of realistic examples treated in considerable detail and with a proper emphasis on the role of substantive theory in model building, make the book a worthwhile reference book and a source of ideas for ways to handle similarity data.

Notwithstanding its computational slant, very little is said about computer programs, which is not necessarily a bad thing. Basic principles and understanding should come before computing, but some guidance about the availability of programs for the analyses presented would have been useful, even if such information tends to become outdated very quickly. Incidentally, the speed of developments in this area, may be illustrated by a quote on page 73. ''Drawing a 3-D configuration is always difficult. It is often worthwhile to construct a physical model using knitting needles or straws on a styrofoam board''. In most departments such models will by now reside in dusty cellars or be play objects for somebody's children.

The most serious shortcoming of the book is its *introversion*. Nowhere is a discussion given of the place of SSA in the whole field of multivariate or multidimensional analysis. For instance, cluster analysis often used in conjunction and as support of multidimensional analyses is never mentioned. As a term, it does not appear in the subject index, nor does, for instance, principal component analysis, principal coordinate analysis, and three-mode scaling; factor analysis is mentioned once. In addition, no mention is made anywhere of how similarities arise, that is, design and data collection aspects are outside the frame of reference of the authors. Clearly, there is no need to treat everything in one book, but at least some guidance towards the literature would be helpful. Another aspect of its introversion is the choice of topics and references. For instance, Ramsay's work (1982) on maximum likelihood MDS is essentially ignored, as is Heiser's work (1981) on unfolding and constrained multidimensional scaling, Carroll's work (Carroll, Pruzansky & Kruskal, 1980) on INDSCAL with linear constraints, and Young's work (1984) on generalized Euclidean scaling. No mention is made of the books by Coxon (1982), Davison (1983), or the survey papers by Shepard (1972), and Carroll and Arabie (1980). Even though I cannot always judge this properly, I also have the impression the authors largely focus on their own (and their close associates') work when discussing various details in other chapters. Similar remarks were made and additional neglected references were mentioned by DeSoete's (1982) review of the German precursor of this book (Borg, 1981).

As a general conclusion, one could say that the book presents by and large the contributions of the authors' (and their close associates') in an integrated and competent fashion, that there is more to similarities and their structure than what the authors present, that what they present they do in a very computational, rather than a conceptual style, that the level of explanation is detailed and understandable, and that their examples are detailed, well-done, and often related to substantive theory. In other words, the book is a useful and valuable reference, rather than a real textbook, but it should not be taken as representative of the entire field of multidimensional scaling and its relatives.

UNIVERSITY OF LEIDEN *Pieter M. Kroonenberg*

References

Borg, I. (1981). *Anwendungsorientierte multidimensionale Skalierung* [Applied multidimensional scaling]. Berlin: Springer.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 31, pp. 607–649). Palo Alto, CA: Annual Reviews.

Carroll, J. D., Pruzansky, S., & Kruskal, J. B. (1980). CANDELINC: A general approach to multidimensional analysis of multi-way arrays with linear constraints on the parameters. *Psychometrika*, *45*, 3–24.

Coxon, A. P. M. (1982). *The user's guide to multidimensional scaling*. Exeter, UK: Heinemann.

Davison, M. L. (1983). *Multidimensional scaling*. New York: Wiley.

De Soete, G. (1982). Review of I. Borg, *"Anwendungsorientierte multidimensionale Skalierung"* [Applied multidimensional scaling]. *Psychologica Belgica*, *22*, 184–185.

Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Leiden: University of Leiden, Department of Data Theory.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling [With discussion]. *Journal of the Royal Statistical Society, Series A*, *145*, 285–312.

Shepard, R. N. (1972). A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences. Vol. 1 Theory* (pp. 21–47). New York: Seminar Press.

Young, F. W. (1984). The general Euclidean model. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 440–469). New York: Praeger.