Analysis of Categorical Data: Dual Scaling and Its Applications. By Shizuhiko Nishisato. Mathematical Expositions No. 24, University of Toronto Press, Toronto, Ontario, 1980. xiii + 276 pp. C$25.00. ISBN 0-8020-5489-7, ISSN 0076-5333.

Review by Yoshio Takane, McGill University.

In social and behavioural sciences categorical data are almost ubiquitous; in opinion polls you may be asked to express your opinion by indicating to what extent you agree or disagree to an attitude statement (e.g., strongly agree, not so strongly agree, neutral, etc.); in certain types of personality inventories you are asked to mark, from a list of adjectives, those which adequately describe your own behavioural disposition; in testing situations you either pass or fail certain test items. Indeed, examples of categorical data abound in those scientific disciplines, and this is precisely why a method to analyze them is so important (and often crucial!) in those areas. Dual scaling, to which the present book is addressed, represents one versatile method for the analysis of categorical data. This method is applicable to a wide variety of categorical data. Yet the basic idea underlying the method is quite simple. So simple that it is hard to believe that a comprehensive English text on the method has not been produced previously. In this regard Nishisato's monograph is a very timely endeavour.

The book is divided into eight chapters, of which the first is an introductory chapter describing the kinds of data (and their numerical representations) to be analyzed by dual scaling. The remaining chapters can be grouped into two major parts, one part focusing on the theoretical and mathematical foundations of the method and the other on applications of the basic formula to a variety of situations. Dual scaling assigns numbers to categories according to an "internal consistency" criterion. This leads to an analytical method similar to the principal components analysis analogue of discrete categorical data. Chapters 2 and 3 discuss the precise definition of the criterion, the derivation of the basic formula as well as other mathematical properties of the solution (including algebraic *duality* from which the name of the method derives). The basic formula of dual scaling is applied, modified and extended in various ways in subsequent chapters according to specific forms of

* See review by R.G. Stanton in Vol. 8, No. 2, 1980, p. 259.

categorical data. Chapter 4 deals with contingency tables and response frequency data, Chapter 5 with response pattern tables, Chapter 6 with paired-comparison and rank-order data, and Chapter 7 with multidimensional tables. The exposition is very clear everywhere except perhaps in the portion discussing Greenacre's account of dual scaling at the end of Chapter 3. On the whole the book is highly readable.

There are several ways to use the book: e.g., (1) as a source book for application oriented readers, (2) as a textbook in an applied multivariate analysis course, or (3) as a side reader in a psychological scaling course. (The mathematical prerequisites for such a course can be kept minimal. The coverage of the book is perhaps too narrow to be used as the sole textbook, but several books can be used concurrently. In view of the fact that no comprehensive up-to-date accounts of scaling methods presently exist, that is what one has to do anyway in teaching a course on scaling. Of course, the book is quite stimulating from the viewpoint of a mature scaling methodologist.)

There are a few points which invite my specific comments. (1) The author develops methods of dual scaling which specifically apply to paired-comparison and rank-order data. This formulation itself is quite sensible. However, it gives the impression that the same kind of interpretation as in the analysis of proximity data can be afforded in the analysis of dominance data as well. Despite the equivalence of the analytical method there seems to be an important distinction between these two cases, which roughly parallels Tucker's distinction between intra- and inter-individual multidimensionality. (2) The treatment of multidimensional tables is very incomplete. The author's main strategy here is to spread out a multidimensional array into a two-way layout and apply additive models analogous to those in a multi-factor ANOVA design. However, since the model of dual scaling for two-way tables is a bilinear form, the multilinear decomposition seems more natural in this context. There are some significant developments in the multilinear decomposition of quantitative data recently, e.g., CANDECOMP/PARAFAC, three-mode factor analysis, etc. (3) A direct application of Bartlett's chi-squared criterion to dual scaling seems too naïve. I am particularly annoyed with the fact that the same statistic is used when some order restrictions are imposed on quantified vectors. In a similar situation I have the impression that this criterion tends to give too few significant eigenvalues in comparison with the results, for example, from the likelihood ratio criterion. Some empirical as well as theoretical investigations are urgently needed in this regard. (4) I feel that the internal consistency criterion used in dual scaling is somewhat arbitrary from an empirical scientist's point of view. In order to empirically validate the criterion one needs to construct a model which more closely simulates the process by which the subjects elicit categorical responses, like the item response model in test theory. I have recently constructed such a model for sorting data, and have confirmed that it gives a very similar result to that of dual scaling. Similar models could be constructed for more general categorical response situations.

All in all, however, the book is highly commendable. It is provocative, full of potential research ideas, guiding our way to a lot of interesting questions. I enjoyed reading it very much.