BOOK REVIEW

Elements of Dual Scaling: An Introduction to Practical Data Analysis

by Shizuhiko Nishisato
Hillsdale NJ: Erlbaum, 1994, 381 pp., $79.95

Elements of Dual Scaling: An Introduction to Practical Data Analysis discusses dual scaling, a collection of techniques for the analysis of discrete multivariate data. Dual scaling also is known as quantification theory in Japan (Hayashi, 1952), correspondence analysis in France (Greenacre, 1984, 1993), and homogeneity analysis in the Netherlands (Gifi, 1990).

This is the second book on the topic by the same author. Whereas the first book, *Analysis of Categorical Data: Dual Scaling and Its Applications* (Nishisato, 1980), primarily emphasized the mathematical exposition of dual scaling, this new book, *Elements of Dual Scaling: An Introduction to Practical Data Analysis*, reviews applications and developments in dual scaling that have taken place in the past 15 years. The book is lucidly written, and provides an excellent text for a one-semester course on dual scaling at an advanced undergraduate or graduate level. Indeed, a copy of the book arrived when I was teaching dual scaling in my scaling course last year, and I could not help but immediately borrow several examples of applications from this book.

One aspect of the book that I particularly like is the well-balanced treatment of dual scaling. It is not too mathematical (even for nonquantitative students) nor is it too cookbookish. Part I provides a very informal introduction to dual scaling and its use, and gradually explains its basic mechanics. Parts II and III provide a more detailed account of applications to two representative types of data–incidence data (Part II) and dominance data (Part III). The incidence data discussed include contingency tables, multiple-choice data, and sorting data; the dominance data discussed include pair-comparison, rank-order, and successive-categories (rating) data. Part IV discusses topics of special interest that represent newer developments in dual scaling such as forced classification, graphical display, outliers and missing data, analysis of multiway data, and other miscellaneous topics.

Nishisato's books always are thought-provoking. That was the impression I had about his first book (Takane, 1982) and I have exactly the same impression about the new one.

Part I

Part I is superbly written, except for a minor misquotation in section 2.5. The example on p. 19 of a possible violation of triangular inequality in similarity judgments originally actually comes from Tversky (1977).

Part II

Part II is also nicely written. The inclusion of a chapter on sorting data is particularly nice. My only

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 18, No. 4, December 1994, pp. 379–382

0146-6216/94/040379-04$1.45

quibble is the title of this part, "Incidence Data." Although all data types discussed are of this type, this unnecessarily limits the scope of dual scaling. The version of dual scaling applicable to incidence data should be equally applicable to any kind of proximity data (Coombs, 1964), of which incidence data are but a special case.

Part III

Extensions of dual scaling to dominance data (Coombs, 1964) are primarily due to Nishisato. These extensions are all intuitively appealing. But what criterion do they optimize? Is it in any sense comparable to Guttman's (1941) internal consistency criterion? These questions keep hovering in my mind. It seems, however, that various extensions of dual scaling in Parts III and IV—as well as the original version described in Part II—can all be viewed as a reduced-rank approximation technique with or without external constraints and special metrics used to evaluate the goodness of approximations. Takane & Shibayama (1991) call this general technique constrained principal component analysis (CPCA).

Part IV

Forced classification. The problem to be dealt with seems to be that of weighting a particular item more heavily than others. In the framework of CPCA mentioned above, this amounts to supplying a special metric matrix. There is a minor error on pp. 254–255. The $\mathbf{A}^*$ should be

$$\mathbf{A}^* = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 3 & 0 & -3 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad (1)$$

and $\mathbf{F}_1\mathbf{A}^*$ on the previous line should be $\mathbf{FA}^*$.

The problem of outliers. Low frequency categories, when they exist, tend to dominate dual scaling solutions. Nishisato proposes robust techniques against outliers. But what really is causing the problem? Dual scaling allows for individual differences. The marginal variance in selecting a category should be (at times substantially) larger than that deduced from a binomial or multinomial distribution. This phenomenon is known as "over dispersion" in the statistical literature (e.g., McCullagh & Nelder, 1989). This implies that metric matrices should be used other than those normally used in dual scaling.

Graphical display. How to display row and column configurations simultaneously has been an issue of fierce debate over the past several years. Assuming that the interpretable relations in dual scaling are interpoint distances, Nishisato rejects the idea of a joint display. Is it, however, always the distance that is interpretable, or is the scalar product a more natural candidate? If so, a joint display still seems feasible.

Analysis of multiway data. Nishisato proposes three alternative approaches. According to the framework of CPCA, however, only the projection operator method makes sense. Nishisato, however, dismisses the projection method for computational economy. I disagree.

Let $\mathbf{P}_G = \mathbf{G}(\mathbf{G}'\mathbf{G})^-\mathbf{G}'$ and $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^-\mathbf{H}'$ be two orthogonal projection operators, and let $\mathbf{Z}$ denote a data matrix. We would like to obtain the singular value decomposition (SVD) of $\mathbf{J} = \mathbf{P}_G\mathbf{Z}\mathbf{P}_H$. There is an efficient method to obtain the SVD of a matrix of the above form (Takane & Shibayama, 1991). Let $\mathbf{F}_G$ and $\mathbf{F}_H$ be columnwise orthogonal matrices spanning the column spaces of $\mathbf{G}$ and $\mathbf{H}$, respectively. Then, the above $\mathbf{J}$ can be rewritten as $\mathbf{J} = \mathbf{F}_G(\mathbf{F}_G'\mathbf{Z}\mathbf{F}_H)\mathbf{F}_H' = \mathbf{F}_G\mathbf{K}\mathbf{F}_H'$, where $\mathbf{K} = \mathbf{F}_G'\mathbf{Z}\mathbf{F}_H'$. Let the SVD of $\mathbf{K}$ be $\mathbf{K} = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$.

Then the SVD of $\mathbf{J} = \mathbf{UDV}'$ can be obtained by $\mathbf{U} = \mathbf{F}_G\mathbf{U}^*$, $\mathbf{V} = \mathbf{F}_H\mathbf{V}^*$, and $\mathbf{D} = \mathbf{D}^*$. Note that the size of $\mathbf{K}$ is usually much smaller than that of $\mathbf{J}$.

Horseshoe effects. Dual scaling allows for multiple quantifications of categories. When the categories are obtained by discretizing continuous variables, multiple quantifications may be viewed as multiple nonlinear transformations of the discretized continuous variables (Gifi, 1990). The first nonlinear transformation (Solution 1 or Dimension 1) may be monotonic with the discretized variables, but not subsequent transformations. These are called horseshoe effects. A popular view is that they are something undesirable. My opinion is to the contrary. Horseshoe effects are a natural consequence of multiple nonlinear transformations, so they are meaningful to the extent that multiple nonmonotonic transformations of the underlying continuous variables are meaningful. In fact, horseshoe effects occur all the time (even for unordered categorical data) in the sense that all the subsequent dimensions are some nonlinear transformations of the first dimension.

Nonlinear transformations. Nonlinear transformations expand the scope of data analysis. In dual scaling, nonlinear transformations are operationally realized by linear transformations of dummy-coded categorical variables. The nonlinear transformations must, therefore, be identity preserving. If too few observation categories are used, there may be a loss of information. There also is another kind of limitation. In dual scaling, only univariate transformations are feasible. This means that, for example, interactions among variables cannot be captured, unless they are explicitly coded as part of the data. Artificial neural network models, recently popular in pattern recognition and cognitive psychology, have some generality over dual scaling in this respect. They are capable of automatically capturing important interactions among variables. This is done by allowing multivariate nonlinear transformations.

Amount of information that can be explained by each dimension. The following three have to be distinguished: (1) the absolute amount of the sum of squares (SS) explained, given by η_k^2; (2) the total SS to be explained, given by $\mathrm{tr}(C) = \Sigma\eta_k^2$; and (3) the maximum possible total SS. What (1)/(2) and (1)/(3) indicate seems obvious. (1)/(2) can be large if (2) is small (which means there is not much association between the rows and columns of a data matrix initially) or it can be small if (1) is large and (2) is even larger. (1)/(3), however, does not depend on how much association there is to be explained in a particular data matrix (2).

Statistical inference. Assessing the reliability of the results is always important. Nishisato, however, leaves the subject for future investigations. But why not bootstrap? It should be usable without much future investigation. Also, techniques are available for sensitivity analysis based on perturbation theory (Tanaka, 1984). Modelling approaches are also interesting, such as ideal point discriminant analysis (IPDA; Takane, 1987), latent class analysis (LCA; e.g., Hagenaars, 1990), and association models (Goodman, 1986). These methods are equipped with statistical model evaluation capabilities based on a large sample rationale. They have recently been reviewed by van der Heijden, Mooijaart, & Takane (1994).

Quality Issues

The print quality of the book is not optimal. In particular, the first few lines of each page are not aligned properly. It took me awhile to get used to this. However, the print quality has nothing to do with the quality of the content. It is an excellent book, and I would like to thank "Nishi" on behalf of all potential readers of the book for his painstaking effort to produce such an excellent book.

Yoshio Takane
McGill University

References

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis and the usual log-linear model approach in the analysis of contingency tables. *International Statistical Review, 54,* 243–309.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 321–348). New York: Social Science Research Council.

Hagenaars, J. A. (1990). *Categorical longitudinal data.* Newbury Park CA: Sage.

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics, 3,* 69–98.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Takane, Y. (1982). A review of "Analysis of categorical data: Dual scaling and its applications" by Nishisato. *Canadian Journal of Statistics, 10,* 67–68.

Takane, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrika, 52,* 493–513.

Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika, 56,* 97–120.

Tanaka, Y. (1984). Sensitivity analysis in Hayashi's third method of quantification. *Behaviormetrika, 16,* 31–44.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–353.

van der Heijden, P. G. M., Mooijaart, A., & Takane, Y. (1994). Correspondence analysis and contingency table models. In M. J. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 79–111). New York: Academic Press.