Dimensional Analysis of Rank-Order and Categorical Data

Harvey A. Skinner and Wen-Jenn Sheu
Addiction Research Foundation
Toronto, Ontario

Skinner (1979) has described a generalized principal components model for classification research that assumes interval or quasi-interval data. First, a parsimonious set of typal dimensions is sought through a multiple replication design, and then relatively homogeneous subgroups are identified within this low dimensional space. The purpose of this paper is to describe preliminary transformations whereby the model may be extended to situations where the data are of either categorical or rank-order metric.

In a previous article, Skinner (1979) described a generalized principal components model for classification research. Although this model is essentially a dimensional or ordination approach to classification, there may be a subsequent search for relatively homogeneous clusters within the low dimensional space. The computations are based on a sequential application of the singular-value decomposition (Stewart, 1973) to samples in a replication design. The purpose of the analysis is (1) to identify a small number of typal dimensions *within* each sample under consideration, (2) to determine the replicability of these dimensions *across* samples, and (3) to use the replicated dimensions to classify new or independent samples. An important feature of this approach is that it allows differentiation of the independent contribution of profile elevation, scatter, and shape parameters in defining similarity among individuals. Three examples of the model's use include classifications of alcoholic patients (Skinner, Jackson, & Hoffmann, 1974), general psychopathology (Skinner & Jackson, 1978), and personality (Burger & Cross, 1979).

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 6, No. 1, Winter 1982, pp. 41–45

0146-6216/82/010041-05$1.25

The discussion by Skinner (1979) and the computer program developed by Skinner and Lei (1980) both focus upon interval or quasi-interval scale data, such as a structured personality inventory (e.g., Jackson, 1974). The purpose of this paper is to outline simple transformations whereby the model may be extended to rank-order and categorical data. This modification will allow investigators to use this dimensional approach for classifying individuals in a broader range of situations, where the metric quality of the data may be either categorical, rank-order, or interval scale.

Consider a data matrix $\mathbf{X}_j$ of n_j individuals by k variables. To simplify the presentation, the subscript j used by Skinner (1979) to denote sample j will be deleted, since attention here will focus on a single sample. The singular-value decomposition of $\mathbf{X}$ is given by

$$X = U \Gamma W' , \qquad [1]$$

where $\mathbb{U}$ $(n \times k)$ and $\mathbb{W}$ $(k \times k)$ are matrices whose columns are, respectively, the left-hand and right-hand eigenvectors ($\mathbb{U}'\mathbb{U} = \mathbb{W}'\mathbb{W} = \mathbb{I}$), and Γ is a diagonal matrix of singular values. Then, the principal component weightings among individuals are computed by $\mathbb{A} = \mathbb{U}\Gamma$, whereas $\mathbb{W}$ contains principal component scores among variables. Only the most important principal components are retained, and the reduced-rank solution is generally transformed to simple structure in the person space (Skinner, 1979). The weights in $\mathbb{A}$ may be examined to identify individuals who are salient on a particular dimension, whereupon the corresponding vector in $\mathbb{W}$ gives the characteristic or modal profile of variables for these individuals. This step is repeated *within* each sample, and then a second stage evaluates the robustness of important dimensions *across* samples.

Rank-Order Data

Let $\mathbb{R}$ denote a matrix of n individuals by k ranked variables. These data may represent, for example, the rank-ordering of symptoms manifested by psychiatric patients. For each scalar element r of $\mathbb{R}$, define e as

$$e = k + 1 - 2r \quad . \qquad [2]$$

Once the new matrix $\mathbb{E}$ $(n \times k)$ is formed by Equation 2, the analysis proceeds as described by Skinner (1979), that is,

$$E = U \Gamma W' \quad , \qquad [3]$$

where $\mathbb{U}$, Γ, and $\mathbb{W}$ are defined above. Similarly, the principal component weightings among individuals are given by $\mathbb{A} = \mathbb{U}\Gamma$, and $\mathbb{W}$ yields the characteristic profiles of (rank-order) variables.

This transformation of rank-order data was proposed by Nishisato (1978) and earlier by de Leeuw (1973). From a different perspective, Woodward and Overall (1976) have demonstrated the value of alternative rank-order transformations of nonlinear data prior to a conventional principal components solution.

Categorical Data

Let $\mathbb{F}$ $(n \times k)$ be the matrix of response-patterns of n individuals on m variables, where $k = k_1 + k_2 + \ldots k_m$ and k_i is the number of options for variable i. For instance, Variable 1 might be marital status (single, married, divorced, separated, widowed), and Variable 2 might be type of residence (apartment, house, condominium, institution). From the dual scaling approach of Nishisato (1980)

$$D_n^{-1} F D^{-1} = U \Gamma W' \qquad [4]$$

where

$\mathbb{D}_n$ is the $(n \times n)$ diagonal matrix of row totals of $\mathbb{F}$,

$\mathbb{D}$ is the $(k \times k)$ diagonal matrix of column total of $\mathbb{F}$, and

$\mathbb{U}$, Γ, and $\mathbb{W}$ are defined above.

Then, $\mathbb{A} = \mathbb{U}\Gamma$ gives the principal component weights for individuals and $\mathbb{W}$ contains the principal component scores for (categorical) data.

With both rank-order and categorical data, a trivial solution of one pair of equally weighted vectors (Guttman, 1946) of $\mathbb{U}$ and $\mathbb{W}$ that do not discriminate among subjects or variables can be eliminated by computing

$$E^* = E - \frac{R \underline{1}\, \underline{1}' R}{\underline{1}' R\, \underline{1}} \qquad [5]$$

$$F^* = F - \frac{F \underline{1}\, \underline{1}' F}{\underline{1}' F\, \underline{1}} \quad . \qquad [6]$$

Then, $\mathbb{E}^*$ is decomposed in Equation 3 instead of $\mathbb{E}$, and $\mathbb{F}^*$ is decomposed in Equation 4 instead of $\mathbb{F}$.

Numerical Examples

In Table 1 a data matrix ($\mathbb{R}$) of eight subjects by five ranked variables has been transformed to give $\mathbb{E}^*$ and then subjected to the singular value decomposition. Since the first two singular values are much larger than the remainder, two principal components ($\mathbb{B}$) were retained and transformed to simple structure (varimax criter-

Table 1
Example Using Rank-Order Data

Subjects	Ranked Variables R				
1	1	2	3	4	5
2	1	3	2	4	5
3	1	2	3	5	4
4	1	2	4	3	5
5	5	1	3	4	2
6	1	5	3	4	2
7	5	1	3	2	4
8	5	1	2	3	4
Singular-Value Decomposition					
	.50	-.10	.09	.04	.00
	.45	-.19	.09	-.61	.00
	.47	-.14	-.39	.15	.00
	.44	-.09	.42	.55	.00
U =	.15	.45	-.66	.28	.00
	.12	-.45	-.37	-.07	.00
	.20	.51	.30	.06	.00
	.25	.51	-.03	-.45	.00
Γ = diag	9.10,	6.18,	2.85,	1.85,	0.00
	.44	-.75	.07	.18	.00
	.51	.64	-.14	.34	.00
B =	.06	.08	-.15	-.88	.00
	-.39	.13	.79	.07	.00
	-.62	-.09	-.57	.28	.00
Varimax Solution					
	4.51	.54		.62	-.61
	4.22	-.11		.33	.75
	4.34	.27	Y =	.04	.09
	4.05	.48		-.41	.02
B =	.61	3.01		-.58	-.25
	1.78	-2.43			
	.95	3.54			
	1.40	3.61			

ion) in the person space. Corresponding principal component scores for variables are listed in matrix $\mathbb{Y}$. The first dimension is defined by Subjects 1, 2, 3, and 4; whereas the second dimension contrasts Subject 6 with Subjects 5, 7, and 8. Corresponding vectors in $\mathbb{Y}$ give the characteristic profile for subjects that are salient on a given dimension in $\mathbb{B}$. Similarly, Table 2 depicts the results for a data matrix ($\mathbb{F}$) of eight subjects by two categorical variables (three options in Item 1; two options in Item 2). The decomposition of $\mathbb{F}^*$ yielded three nonzero singular values. A varimax transformation of the first two principal components in the person space ($\mathbb{B}$) con-

Table 2
Example Using Categorical Data

Subjects	Item 1			Item 2	
1	0	0	1	0	1
2	0	1	0	0	1
3	0	0	1	1	0
4	0	1	0	1	0
5	1	0	0	1	0
6	1	0	0	0	1
7	1	0	0	1	0
8	0	1	0	0	1

Singular-Value Decomposition

	.28	-.24	.61	.00	.00
	.45	.15	-.20	.00	.00
	-.28	.24	.61	.00	.00
	-.11	.63	-.20	.00	.00
U =	-.45	-.15	-.20	.00	.00
	.11	-.63	-.20	.00	.00
	-.45	-.15	-.20	.00	.00
	.45	.15	-.20	.00	.00

Γ = diag 2.15 , 1.54 , 1.50 , 0.00 , 0.00

	-.37	-.60	-.41	.00	.00
	.37	.60	-.41	.00	.00
W =	.00	.00	.82	.00	.00
	-.60	.37	.00	.00	.00
	.60	.37	.00	.00	.00

Varimax Solution

	.42	-.57
	.99	.14
	-.42	.57
	.14	.99
B =	-.99	.14
	-.14	-.99
	-.99	.14
	.99	-.14

	-.57	-.42
	.57	.42
Y =	.00	.00
	-.42	.57
	.42	.57

trasted Subjects 2 and 8 versus Subjects 5 and 7 on Dimension 1, and Subject 4 versus 6 on Dimension 2.

In conclusion, this note has focused upon extending the classification model of Skinner (1979) to data of rank-order and categorical scale. This extension has drawn upon recent developments in dual scaling (Nishisato, 1980). Conversely, the replication stage of Skinner (1979) might be explored as a mechanism for assessing the robustness of dual scaling solutions across samples.

References

Burger, G. K., & Cross, D. T. Personality types as measured by the California Psychological Inventory. *Journal of Consulting and Clinical Psychology,* 1979, *47,* 65–71.

de Leeuw, J. *Canonical analysis of categorical data,* Unpublished doctoral dissertation, University of Leiden, The Netherlands, 1973.

Guttman, L. An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics,* 1946, *17,* 144–163.

Jackson, D. N. *The Personality Research Form manual.* Goshen, NY: Research Psychologists Press, 1974.

Nishisato, S. Optimal scaling of paired comparison and rank order data: An alternative to Guttman's formulation. *Psychometrika,* 1978, *43,* 263–271.

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

Skinner, H. A. Dimensions and clusters: A hybrid approach to classification. *Applied Psychological Measurement,* 1979, *3,* 327–341.

Skinner, H. A., & Jackson, D. N. A model of psychopathology based on an integration of MMPI actuarial systems. *Journal of Consulting and Clinical Psychology,* 1978, *46,* 231–238.

Skinner, H. A., Jackson, D. N., & Hoffmann, H. Alcoholic personality types: Identification and correlates. *Journal of Abnormal Psychology,* 1974, *83,* 658–666.

Skinner, H. A., & Lei, H. Modal Profile Analysis: A computer program for classification research. *Educational and Psychological Measurement,* 1980, *40,* 769–772.

Stewart, G. W. *Introduction to matrix computations.* New York: Academic Press, 1973.

Woodward, J. A., & Overall, J. E. Factor analysis of rank-ordered data: An old approach revisited. *Psychological Bulletin,* 1976, *83,* 864–867.

Acknowledgments

This research was supported by the Addiction Research Foundation of Ontario.

Author's Address

Send requests for reprints or further information to Harvey A. Skinner, Addiction Research Foundation, 33 Russell Street, Toronto, Canada M5S 2S1.