PSYCHOMETRIKA—VOL. 55, NO. 4, 633–639
DECEMBER 1990

CANONICAL ANALYSIS OF CONTINGENCY TABLES WITH LINEAR CONSTRAINTS

ULF BÖCKENHOLT

UNIVERSITY OF ILLINOIS, CHAMPAIGN–URBANA

INGO BÖCKENHOLT

UNIVERSITY OF KARLSRUHE

A generalized least squares approach is presented for incorporating linear constraints on the standardized row and column scores obtained from a canonical analysis of a contingency table. The method is easy to implement and may simplify considerably the interpretation of a data matrix. The approach is compared to a restricted maximum likelihood procedure.

Key words: canonical analysis, singular value decomposition, restricted maximum likelihood.

Canonical correspondence analysis (CCA) is a useful tool for obtaining a graphical representation of the dependence between the rows and columns of a contingency table (e.g., Benzecri et al.; 1980; Gifi, 1981; Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; Nishisato, 1980). This representation is achieved by assigning scores in the form of coordinates to row and column categories. The interpretation of the multidimensional representation of the row and column categories may be greatly simplified if additional information about the row and column structure of the table is available. By incorporating this external information through linear constraints on the row and/or column scores, a representation of the data may be obtained that is not only more parsimonious but is also easier to understand. Recently, a restricted maximum likelihood (ML) strategy has been suggested for placing linear contrasts constraints on scores estimated in a canonical correlation or in an association model (Gilula & Haberman, 1986, 1988; Goodman, 1985). As discussed by Escoufier and Junca (1986), a constrained least squares approach may be more useful for *exploring* a contingency table. In this note, we present a noniterative least squares approach for incorporating linear constraints on the row and column scores. This method is easy to implement and is an extension of Rao's (1964) work on the estimation of restricted eigenvectors. In an illustrative analysis of a small data set, the least squares and the restricted maximum likelihood solution are compared. Related work on linear constraints is given in Nishisato and Lawrence (1989), ter Braak (1988), Yanai (1988), Takane (1987), and Takane and Shibayama (in press). In particular, the latter reference provides a comprehensive framework for imposing linear constraints on data matrices.

Canonical Correspondence Analysis with Lineara Constraints

Consider an $I \times J$ contingency table $\mathbf{P}$ with proportions p_{ij} describing the joint distribution of two categorical random variables, X and Y, with I and J categories, respectively. Let $\mathbf{D}_r$ and $\mathbf{D}_c$ be diagonal matrices containing the row and column sums

Requests for reprints should be addressed to Ulf Böckenholt, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820. The authors are indebted to Yoshio Takane for helpful comments on a previous draft of this manuscript.

0033-3123/90/1200-89139$00.75/0

of **P**, respectively; that is, the diagonal elements of $\mathbf{D}_r$ are $p_{i.} = \Sigma_{j=1}^{J} p_{ij}$, and the diagonal elements of $\mathbf{D}_c$ are $p_{.j} = \Sigma_{i=1}^{I} p_{ij}$. In a canonical analysis, elements of the matrix **P** are represented as

$$p_{ij} = p_{i.}p_{.j}\left(1 + \sum_{m=1}^{M-1} \rho_m x_{im} y_{jm}\right), \tag{1}$$

where M is min (I, J) and $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_{M\text{-}1} \geq 0$. Moreover, for $1 \leq m \leq m' \leq M - 1$

$$\sum_{j=1}^{J} p_{.j} y_{jm} = \sum_{i=1}^{I} p_{i.}\, x_{im} = 0, \quad \text{and}$$

$$\sum_{j=1}^{J} p_{.j} y_{jm} y_{jm'} = \sum_{i=1}^{I} p_{i.}\, x_{im}\, x_{im'} = \delta_{mm'},$$

where $\delta_{mm'}$ is the Kronecker delta. The parameters x_{im} and y_{jm} are standardized row and column scores, respectively, and the parameter ρ_m is the canonical correlation between these scores. Goodman (1985) presented an iterative maximum likelihood estimation procedure for determining the parameters of (1).

Alternatively, least squares estimates can be obtained by first calculating the **P** matrix of proportions under the assumption of independence as $\mathbf{E} = \mathbf{D}_r\mathbf{1}\,\mathbf{1}'\mathbf{D}_c$ (where **1** is a unit vector), and then computing a singular value decomposition (SVD) of the matrix Z,

$$\mathbf{Z} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{E})\mathbf{D}_c^{-1/2} = \mathbf{U}\mathbf{D}_\lambda \mathbf{V}', \tag{2}$$

where $\mathbf{U}'\mathbf{U} = \mathbf{I} = \mathbf{V}'\mathbf{V}$, and $\mathbf{D}_\lambda$ is a diagonal matrix with singular values λ in descending order. Standardized row scores and column scores may be computed as

$$\mathbf{X} = \mathbf{D}_r^{-1/2}\mathbf{U}, \quad \text{and} \quad \mathbf{Y} = \mathbf{D}_c^{-1/2}\mathbf{V}, \tag{3}$$

yielding $\mathbf{X}'\mathbf{D}_r\mathbf{X} = \mathbf{I} = \mathbf{Y}'\mathbf{D}_c\mathbf{Y}$, and $\mathbf{1}'\mathbf{D}_r\mathbf{X} = 0 = \mathbf{1}'\mathbf{D}_c\mathbf{Y}$ (e.g., van der Heijden, de Falguerolles, & de Leeuw, 1989).

Using a restricted maximum likelihood approach, Gilula and Haberman (1986, 1988) extended Goodman's canonical correlation model by introducing linear constraints on the canonical row and column scores. These constraints take on the form

$$\sum_{i=1}^{I} g_{iam}\, x_{im} = 0, \quad \text{and}$$

$$\sum_{j=1}^{J} h_{jbm} y_{jm} = 0,$$

where

$$\sum_{i=1}^{I} g_{iam} = \sum_{j=1}^{J} h_{jbm} = 0.$$

For each m, the a contrast vectors of the row scores, g_{iam}, and the b contrast vectors of the column scores, h_{jbm}, are linearly independent. Similarly to the unrestricted case, a least squares approach can be developed which is noniterative. This procedure is based on a straightforward extension of Rao's (1964) work on estimating linearly restricted eigenvectors. Linear row and column constraints are defined as

$$\mathbf{G}'\mathbf{X}^* = \mathbf{0} \quad \text{and} \quad \mathbf{H}'\mathbf{Y}^* = \mathbf{0}, \tag{4}$$

where $\mathbf{G}$ is a $I \times K$ matrix of rank K and $\mathbf{H}$ is a $J \times L$ matrix of rank L. The constrained standardized row and column scores are denoted by $\mathbf{X}^*$ and $\mathbf{Y}^*$, respectively. These scores are computed from the singular value decomposition of

$$\{\mathbf{I} - \mathbf{D}_r^{-1/2}\mathbf{G}(\mathbf{G}'\mathbf{D}_r^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{D}_r^{-1/2}\}\mathbf{Z}\{\mathbf{I} - \mathbf{D}_c^{-1/2}\mathbf{H}(\mathbf{H}'\mathbf{D}_c^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{D}_c^{-1/2}\} = \mathbf{M}\mathbf{D}_\eta\mathbf{B}', \tag{5}$$

with $\mathbf{M}'\mathbf{M} = \mathbf{I} = \mathbf{B}'\mathbf{B}$, and $\mathbf{D}_\eta$ is a diagonal matrix with singular values η in descending order. The standardized constrained scores are given by $\mathbf{X}^* = \mathbf{D}_r^{-1/2}\mathbf{M}$ and $\mathbf{Y}^* = \mathbf{D}_c^{-1/2}\mathbf{B}$ with $\mathbf{X}^{*\prime}\mathbf{D}_r\mathbf{X}^* = \mathbf{I} = \mathbf{Y}^{*\prime}\mathbf{D}_c\mathbf{Y}^*$ and $\mathbf{1}'\mathbf{D}_r\mathbf{X}^* = \mathbf{0} = \mathbf{1}'\mathbf{D}_c\mathbf{Y}^*$. It can be shown that (5) is equivalent to (2) in the unrestricted case by setting $\mathbf{G} = \mathbf{D}_r\mathbf{1}$ and $\mathbf{H} = \mathbf{D}_c\mathbf{1}$. These constraints ensure that the weighted average of the row and column scores equal zero. Although this procedure imposes the constraint defined by (4) on each singular vector, we may use a successive approach to define different constraints on each singular vector. For example, it may be useful to impose uniform spacing on the scores of the first singular vector but equality constraints on the scores of the second singular vector. Different constraints can be introduced by extracting the standardized row and column scores corresponding to the first singular value η_1 and computing the rank-one reduced matrix $\mathbf{Z}_1$ as

$$\mathbf{Z}_1 = (\mathbf{I} - \mathbf{m}_1\mathbf{m}_1')\mathbf{Z}(\mathbf{I} - \mathbf{b}_1\mathbf{b}_1'),$$

where $\mathbf{m}_1$ and $\mathbf{b}_1$ are the singular vectors corresponding to η_1. In the next step, we substitute $\mathbf{Z}_1$ for $\mathbf{Z}$ in (5) and, in addition to the new constraints, we augment the matrices $\mathbf{G}$ and $\mathbf{H}$ in (4) by $\mathbf{D}_r^{1/2}\mathbf{m}_1$ and $\mathbf{D}_c^{1/2}\mathbf{b}_1$, respectively. These additional constraints ensure that the relationships, $\mathbf{X}^{*\prime}\mathbf{D}_r\mathbf{X}^* = \mathbf{I} = \mathbf{Y}^{*\prime}\mathbf{D}_c\mathbf{Y}^*$, are satisfied. We may extract further singular values subject to certain constraints by following this successive procedure.

Finally, similar results can be derived for the association model. Both Escoufier (1988) and Gilula (1982) present an approach to transform a contingency table such that its structure conforms to the association model. Linearly constrained row and column scores of the transformed table may be obtained by applying the methods discussed above.

Application

To illustrate the approach, we analyze a 9 × 3 table (Table 1) from Haberman (1979, p. 399). The nine rows of this table are a result of a 3 × 3 cross-classification of religion and years of education, while the column variable describes attitudes towards nontherapeutic abortion. Gilula and Haberman (1988) analyzed this small data set using their restricted maximum likelihood approach.

An unrestricted CA yields the singular values $\lambda_1 = 0.2769$ and $\lambda_2 = 0.0653$. Clearly, a one-dimensional solution is sufficient for representing this table. In Table 2, the rows and column scores obtained from (2) are presented. In their analysis, Gilula and Haberman (1988) noted that given religion, the scores appear to be approximately linear in the three educational levels. Moreover, the linear spacing seems to be similar

TABLE 1

Attitudes Towards Nontherapeutic Abortion

Cross classified by Religion and Education of Respondents

		Attitude		
Religion	Education (Years)	Positive	Neutral	Negative
N. Protestants	≤8	49	46	115
	9-12	293	140	277
	≥13	244	66	100
S. Protestants	<8	27	34	117
	9-12	134	98	167
	≥13	138	38	73
Catholic	≤8	25	40	88
	9-12	172	103	312
	≥13	93	57	135

for the North and South Protestant groups, but different for the Catholic group. Thus, a constrained solution was computed by setting $\mathbf{H} = \mathbf{D}_c\mathbf{1}$ and $\mathbf{G} = (\mathbf{D}_r\mathbf{1} | \mathbf{G}_1)$ with

$$\mathbf{G}_1' = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

While the first three rows of $\mathbf{G}_1'$ ensure the linear spacing of the standardized row scores, the last row restricts the spacing to be the same for both Protestant groups. The standardized column and row scores corresponding to the first singular value of the unrestricted and restricted SVD are presented in Table 2. The scores indicate only subtle differences between the two solutions. Using equally-spaced coefficients for the rows, we can account for most of the variations in the residuals. The singular values obtained from the restricted SVD are $\eta_1 = 0.2759$ and $\eta_2 = 0.0252$. Moreover, the row and column scores agree closely with the row and column scores obtained by Gilula and Haberman's restricted ML solution presented in the last column of Table 2.

TABLE 2

Column and Row Scores of SVD Decomposition of Table 1

a) Row Scores Corresponding to η_1

Singular Value Decomposition		
Unconstrained	Constrained	Constrained (ML)
-1.033	-1.121	-1.110
+0.349	+0.338	+0.342
+1.724	+1.797	+1.794
-1.749	-1.620	-1.624
-0.154	-0.161	-0.172
+1.378	+1.299	+1.280
-1.495	-1.299	-1.349
-0.655	-0.764	-0.764
-0.334	-0.228	-0.180

b) Column Scores Corresponding to η_1

Singular Value Decomposition		
Unconstrained	Constrained	Constrained (ML)
+1.271	+1.269	+1.274
-0.309	-0.296	-0.324
-0.940	-0.944	-0.935

Discussion

Introducing linear constraints on the row and column coordinates of a CA representation is useful in a variety of applications. For example, a CA of paired comparison data (Nishisato, 1980) may be easier to understand by imposing row constraints that contain external information about the choice alternatives. Other important applications where external information is available are, for example, the analysis of confusion or transition matrices (van der Heijden, & de Leeuw, 1985). The implementation of linear constraints is both straightforward and may greatly simplify the interpretation of the data matrix. Moreover, the possibility of imposing different constraints for each singular value may be useful in developing a parsimonious representation of a frequency table.

In the exemplary application, the scores derived from the SVD and the restricted ML procedure agreed closely. This result is not too surprising when considering the low dimensionality of the solutions. Although a statistical modeling approach can rigorously address certain questions, for instance, the dimensionality of the dependence or goodness-of-fit, there are certain advantages in utilizing a least squares approach that are discussed in detail by Gilula (1982). For example, the estimates obtained from the SVD are consistent estimates and when used as starting values in an iterative ML estimation procedure, the least squares estimates lead to convergence of the ML procedure with probability 1 (see Goodman, 1974). A more important advantage of the least squares approach is that canonical correspondence analysis is generally performed on large data tables. Clearly, an iterative maximum likelihood approach for estimating row and column scores may be quite costly and possibly infeasible if a large number of constraints are incorporated. In contrast, the SVD is numerically very stable and does not require any iterations. Thus, particularly in large scale applications with sparse matrices and matrices with some low cell frequencies, the least squares approach with linear constraints provides an attractive alternative to the restricted ML approach.

References

Benzécri, J. P. et al. (1980). *Practique de l'analyse des donnees* [Practice of data analysis] (Vols. 1–3). Paris: Dunod.

Escoufier, Y. (1988). Beyond correspondence analysis. In H. H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 505–514). Amsterdam: North-Holland.

Escoufier, Y., & Junca, S. (1986). Least-squares approximation of frequencies and their logarithms. *International Statistical Review, 54*, 279–283.

Gifi, A. (1981). *Non-linear multivariate analysis.* Leiden: University of Leiden.

Gilula, Z. (1982). A note on the analysis of association in cross-classifications having ordered categories. *Communications in Statistics, Part A: Theory and Methods, 11*, 1233–1240.

Gilula, Z., & Haberman, S. J. (1986). Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association, 81*, 780–788.

Gilula, Z., & Haberman, S. J. (1988). The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association, 83*, 760–771.

Goodman, L. (1974). The analysis of systems of qualitative variables when some variables are unobservable, Part I: A modified latent structure approach. *American Journal of Sociology, 75*, 1179–1259.

Goodman, L. (1985). The analysis of cross-correlated data having ordered and/or unordered categories: Association models, correlatin models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics, 13*, 10–69.

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

Haberman, S. J. (1979). *Analsysis of qualitative data* (Vol. 2). New York: Academic Press.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis: Correspondence analysis and related techniques for large matrices.* New York: Wiley.

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

Nishisato, S., & Lawrence (1989). Dual scaling of multiway data matrices:Several variants. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 317–326). Amsterdam: North Holland.

Rao, C. R. (1964). The use and interpretation of principal components analysis in applied research. *Sankhya, Series A, 26*, 329–358.

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

Takane, Y., & Shibayama. (in press). Principal component analysis with external information on both subjects and variables. *Psychometrika.*

ter Braak, C. J. F. (1988). Partial canonical correspondence analysis. In H. H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 551–558). Amsterdam: North-Holland.

van der Heijden, P. G. M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, *50*, 429–447.

van der Heijden, P. G. M., de Falguerolles, A., & de Leeuw, J. (1989). A combined approach of contingency tables analysis using correspondence analysis and loglinear analysis. *Applied Statistics*, *38*, 249–292.

Yanai, H. (1988). Partial correspondence analysis and its properties. In C. Hayashi, M. Jambu, E. Diday, & N. Ohsumi (Eds.), *Recent developments in clustering and data analysis* (pp. 259–266). Boston: Academic Press.

Manuscript received 7/16/89
Final version received 10/11/89