EXAMPLE

OPTIMUM SCORES UNDER ORDER CONSTRAINTS IN CONTINGENCY TABLES

A. Rahulji Parsa and William B. Smith
Department of Statistics
Texas A&M University
College Station, Texas 77843-3143
smith@stat.tamu.edu

Key Words: contingency tables; structural categories; optimal scoring.

ABSTRACT

Methods are developed for analyzing contingency tables which have ordered categories, *a priori*. An exact representation is obtained when differences in the scores of the categories are known (called, strong structure). When only category ordering is known (weak structure) several techniques are reduced to solutions constrained optimization problems. Numerical comparisons with other techniques are given in each case.

1. INTRODUCTION AND NOTATION

The problem of assigning scores to the rows and columns of a contingency table is an old one. Using canonical analysis is one method of solution, while correspondence analysis provides another method. Interest in correspondence analysis has been increasing in recent years, mainly due to the works of Benézecri and Lebart and the books by Greenacre (1984), Bradley et al. (1962), Goodman (1979) and Haberman (1974) have demonstrated the close relationship between canonical analysis, correspondence analysis and principal

In this paper we propose a method based on correspondence analysis to find optimum scores for rows and columns of an $I \times J$ contingency table when there is either a “strong” or a “weak” structure on the categories. In a “strong” structure, the differences in the scores of the categories are specified, while in a “weak” structure only a relative order of the categories is specified.

Let N_{ij} be the frequencies $(i = 1, \ldots, I; j = 1, \ldots, J)$ in an $I \times J$ table. Following Greenacre (1) we define the following quantities.

$$P = [p_{ij}] \text{ of order } I \times J, \text{ where } p_{ij} = n_{ij}/n \tag{1.1}$$

and

$$n = \sum_i \sum_j n_{ij}; \quad n_{i.} = \sum_j n_{ij}; \quad n_{.j} = \sum_i n_{ij}. \tag{1.2}$$

Let

$$r = Pe, \quad c = P'e \tag{1.3}$$

where e denotes a column vector of unit elements of an appropriate order. Let D_r, D_c denote diagonal matrix with elements given by those of r and c, respectively. Then

$$E = rc' = \left[\frac{n_{i.}n_{.j}}{n}\right]. \tag{1.4}$$

where A is of order $I \times k$, Λ is a diagonal matrix of order $k \times k$ with diagonal elements $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$, B is of order $J \times k$, k is the smaller of $(I-1)$ and $(J-1)$, and $\lambda_\alpha^2 (\alpha = 1, 2, \ldots, k)$ are the eigenvalues of $(P-E)(P-E)'$. It is a "generalized" SVD because A and B are normalized with respect to metrics D_r^{-1} and D_c^{-1}, that is,

2. STRONG ORDER CONSTRAINTS

Order constraints are of at least two types; one where the differences in scores of the categories is exactly specified and the other where the categories is exactly specified and the other where the categories are ordered. Since order constraints are in terms of differences of scores, they can be expressed in terms of contrast matrices of the type (illustrated for $I = 5$, say)

$$\text{(a)} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

$$\text{(b)} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix}$$

$$\text{(c)} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

We will assume that the complete order of all categories is known and without loss of generality use the matrix 0_1 of the form (d) above

3. ILLUSTRATION

As an illustration, consider the data from Goodman (3).

Table I: Cross-classification of 135 Women According to their Periodontal Condition and Calcium Intake Level.

		Level				
		1	2	3	4	Total
	A	5	3	10	11	29
Condition	B	4	5	8	6	23
	C	26	11	3	6	46
	D	23	11	1	2	37
	Total	58	30	22	25	135

The order constraint is $0_1 F_1 = \Delta_1$ where

$$0_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} ; \Delta_1 = \begin{bmatrix} .92 \\ .92 \\ .92 \\ .92 \end{bmatrix} \tag{3.1}$$

From the SVD of $R - e\,c'$, without considering the constraints (3.1), the optimal F_1 comes out as

Table II: Treatment by Categories Matrix of Response Frequencies.

	A	B	C	D	E	Total
A	9	5	9	13	4	40
B	7	3	10	20	4	44
C	14	13	6	7	0	40
D	11	15	3	5	8	42
E	0	2	10	30	2	44
Total	41	38	38	75	18	210

A = excellent, B = good, C = Fair, D = Poor, E = Terrible.

4. SUMMARY

Procedures were presented for developing scores which incorporate the natural order in the objects (or categories) that characterize the contingency table. We considered the following two possible order structures: (i) *Strong Structure* or the case in which the

BIBLIOGRAPHY

Bradley, R. A., Katti, S. K., and Coons, I. J. (1962). Optimal scaling for ordered categories. *Psychometrika*, **27**, 355–374.

Goodman, L. A. (1979). Simple models for the analysis of associations in cross classified data having ordered categories. *J. Amer. Statist. Assoc.*, **74**, 537–552.

Greenacre, M. J. (1984). *Correspondence Analysis*. New York: Academic Press.

Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classification. *Biometrics*, **30**, 589–600.

**References should be cited in the text by author and date. Please list the references in alphabetical order.

**Equations and tables may be centered.

**When sending TEX or LATEX files, please be sure to include all the input and style files. It is recommended to put the article on a disk and run it from a different computer and check for error messages before sending to *CIS*.