

# Tests of ignoring and eliminating in nonsymmetric correspondence analysis (NSCA)

Yoshio Takane and Sunho Jung

Department of Psychology  
McGill University

July 14 2008

# Analysis of contingency tables with an asymmetric role of rows and columns

- Table 1. Cross-classification of patients in terms of Treatment, Gender, and Prognosis

|            |        | Recovery | Non-recovery |
|------------|--------|----------|--------------|
| Medication | Male   | 180      | 120          |
| Control    |        | 70       | 30           |
| Medication | Female | 20       | 80           |
| Control    |        | 120      | 180          |

- Nonsymmetric Correspondence Analysis (NSCA; Lauro & D'Ambra, 1984)

# Nonsymmetric Correspondence Analysis

Let  $F$  denote  $r$  by  $c$  contingency table

Let  $P = F/n$

Let  $P_R$  and  $P_C$  represent diagonal matrices of row and column marginal probabilities of  $P$ , respectively

- **Measuring predictive power**

$$\begin{aligned} A &= P_R^{-1}P - 1_R 1'_C P_C \\ &= Q_{1/P_R} P_R^{-1}P \quad (Q_{1/P_R} = I_R - 1_R 1'_R P_R) \end{aligned}$$

- **Solution**

$$\text{GSVD}(A)_{P_R, I}$$

- **Goodman-Kruskal (1954)'s  $\tau$  index**

$$\text{BSS} = \text{SS}(A)_{P_R, I} = \text{tr}(A'P_R A)$$

$$\text{TSS} = 1 - \text{tr}(P_C^2)$$

$$\tau = \text{BSS} / \text{TSS}$$

- **CATANOVA test (Light & Margolin, 1971)**

$$C = (n - 1)(c - 1)\tau \sim \chi^2_{(c-1)(r-1)}$$

# Constrained NSCA

- Table 1.

|            |        | Recovery | Non-recovery |
|------------|--------|----------|--------------|
| Medication | Male   | 180      | 120          |
| Control    |        | 70       | 30           |
| Medication | Female | 20       | 80           |
| Control    |        | 120      | 180          |

- Orthogonal contrasts  $T_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$ ,  $T_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & 1 \end{pmatrix}$

- **Constrained NSCA** (Takane & Jung, in press)

Let  $H$  denote a known linear constraint matrix

Let  $X = Q_{1/P_R}H$

$$A^{(C)} = X(X'P_RX)^{-1}X'P$$

# Decomposition of $Q_{1/P_R} P_R^{-1} P$

- Khatri (1966)'s lemma

Let  $S$  denote a  $p$  by  $p$  positive definite symmetric matrix

Let  $B$  and  $C$  be such that  $B'C = 0$  and  $\text{rank}(B) + \text{rank}(C) = p$

$$S^{-1} = B(B'SB)^{-1}B' + S^{-1}C(C'S^{-1}C)^{-1}C'S^{-1}$$

- **Decomposition of  $Q_{1/P_R} P_R^{-1} P$**

Let  $T_1$  and  $T_2$  be a matrix of contrast vectors such that  $\text{Sp}[T_1, T_2] = \text{Ker}(1'_R)$  and  $T_1'T_2 = 0$

$$\begin{aligned} Q_{1/P_R} P_R^{-1} P &= Q_{1/P_R} T_1 (T_1' P_R Q_{1/P_R} T_1)^{-1} T_1' Q_{1/P_R}' P \\ &\quad + Q_{1/P_R} P_R^{-1} T_2 (T_2' P_R^{-1} T_2)^{-1} T_2' P_R^{-1} P \end{aligned}$$

# Constrained NSCA and analysis of its residuals

- $A = [A_{T_1 \text{ ignoring } T_2} + A_{T_2 \text{ eliminating } T_1}]$

- Constrained NSCA and analysis of its residuals

## Constrained NSCA of contingency table with $T_1$ ignoring $T_2$

$$A_{T_1 \text{ ignoring } T_2} = Q_{1/P_R} T_1 (T_1' P_R Q_{1/P_R} T_1)^{-1} T_1' Q_{1/P_R}' P$$

## Constrained NSCA with $T_2$ eliminating $T_1$

$$A_{T_2 \text{ eliminating } T_1} = P_R^{-1} T_2 (T_2' P_R^{-1} T_2)^{-1} T_2' P_R^{-1} P$$

- Interchange the roles of  $T_1$  and  $T_2$

$$A = [A_{T_2 \text{ ignoring } T_1} + A_{T_1 \text{ eliminating } T_2}]$$

# Simpson's paradox

- Table 1.

|            |        | Recovery | Non-recovery | Recovery rate |
|------------|--------|----------|--------------|---------------|
| Medication | Male   | 180      | 120          | .6            |
| Control    |        | 70       | 30           | .7            |
| Medication | Female | 20       | 80           | .2            |
| Control    |        | 120      | 180          | .4            |

- Table 2. The marginal table within Table 1, collapsing over Gender

|            | Recovery | Non-recovery | Recovery rate |
|------------|----------|--------------|---------------|
| Medication | 200      | 200          | .5            |
| Control    | 190      | 210          | .475          |

# Decomposition of $\tau$ index and CATANOVA test statistics

- Decomposition of  $\tau$  index

$$\text{BSS} = \underline{\text{BSS}_{T_1 \text{ ignoring } T_2}} + \text{BSS}_{T_2 \text{ eliminating } T_1}$$

- Decomposition of CATANOVA test statistics

Let  $r_1^*$  and  $r_2^*$  denote rank( $T_1$ ) and rank( $T_2$ ), respectively

$$(r_1^* + r_2^* = r - 1)$$

$$(n - 1)(c - 1) \text{BSS}_{T_1 \text{ ignoring } T_2} / \text{TSS} \sim \chi_{(c-1)r_1^*}^2$$

$$(n - 1)(c - 1) \text{BSS}_{T_2 \text{ eliminating } T_1} / \text{TSS} \sim \chi_{(c-1)r_2^*}^2$$

- Interchange the roles of  $T_1$  and  $T_2$

$$\text{BSS} = \text{BSS}_{T_2 \text{ ignoring } T_1} + \underline{\text{BSS}_{T_1 \text{ eliminating } T_2}}$$

# An illustration

- Table 1.

|            |        | Recovery | Non-recovery |
|------------|--------|----------|--------------|
| Medication | Male   | 180      | 120          |
| Control    |        | 70       | 30           |
| Medication | Female | 20       | 80           |
| Control    |        | 120      | 180          |

- Main effect of Treatment (T)

$$T_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

- Main effect of Gender (G) and Interaction effect b/w T and G (TG)

$$T_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & 1 \end{pmatrix}$$

# CATANOVA test results

- $C_{T_1 \text{ eliminating } T_2}: \chi^2(1) = 15.05$

$$C_{T_2 \text{ ignoring } T_1}: \chi^2(2) = 60.50$$

$$C_{T_2 \text{ ignoring } T_1} + C_{T_1 \text{ eliminating } T_2} = 75.55 \quad (3)$$

- $C_{T_1 \text{ ignoring } T_2}: \chi^2(1) = 0.48$

$$C_{T_2 \text{ eliminating } T_1}: \chi^2(2) = 75.07$$

$$C_{T_1 \text{ ignoring } T_2} + C_{T_2 \text{ eliminating } T_1} = 75.55 \quad (3)$$

- Constrained NSCA of Table 1 with the T eliminating G and TG

$$A_{T_1 \text{ eliminating } T_2} = \begin{pmatrix} -.0375 & .0375 \\ .1125 & -.1125 \\ -.1125 & .1125 \\ .0375 & -.0375 \end{pmatrix}$$

- Constrained NSCA of Table 1 with the T ignoring G and TG

$$A_{T_1 \text{ ignoring } T_2} = \begin{pmatrix} .0125 & -.0125 \\ -.0125 & .0125 \\ .0125 & -.0125 \\ -.0125 & .0125 \end{pmatrix}$$

- Constrained NSCA
- Analysis of the residuals from Constrained NSCA
- Simpson's paradox in the analysis of contingency table