New developments of Non Symmetrical Correspondence Analysis for two and three way contingency table

LUIGI D'AMBRA *(Dept. Biological Science, dambra@unina.it)*
ANTONELLO D'AMBRA *(Department of Accounting Management and Quantitative Methods- Second University of Naples, antonello.dambra@unina2.it)*
PASQUALE SARNACCHIARO *(University of Naples Federico II, sarnacch@unina.it)*

The identification of meaningful relationships between two or more categorical variables is an important element to the analysis of contingency tables. It involves detecting categories of the variables that are similar and/or different to other categories of the same, or of a different, variable. Generally, such relationships are identified by first isolating those categories that do, and not, exhibit characteristics consistent with the departure from independence.
For the graphical analysis of these departures, correspondence analysis is a very popular tool that can represent the strength of dependence between the rows and columns of a two-way contingency table. When two categorical variables are assumed to have a two-way relationship, symmetric or simple, correspondence analysis can be performed. When it is evident that there is a one-way relationship between the variables, then non-symmetrical correspondence analysis can be used (D'Ambra L. and Lauro, 1989; Takane and Jung, 2009, Sarnacchiaro and D'Ambra A., 2007).
When, there is a directional association the Pearson chi-squared statistic can't be use and the Goodman Kruskal or Light-Margolin (1971) index are preferred to analyse the non symmetrical relationship.
Let N and $P = N/n = \left[n_{ijk}/n = p_{ijk} \right]$ be the absolute and relative three-way contingency tables, respectively, of dimension $I \times (J \times K)$, with I categories of the response variable, $J \times K$ categories of the composite predictor variable generated matching the two predictor variables and $n = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}$.

Let $p_{i\bullet} = \sum_{j=1}^{J}\sum_{k=1}^{K} p_{ijk}$ and $p_{\bullet jk} = \sum_{i=1}^{I} p_{ijk}$ be the marginal row and column frequencies, respectively.
In order to analyze this structure of data, we use Multiple Non Symmetrical Correspondence Analysis (D'Ambra and Lauro, 1989), particularly we decompose the Grey-Williams index.
Starting from Multiple Non Symmetrical Correspondence Analysis (MNSCA), in this paper we show some statistical tools (Confidence regions, statistical tests and graphical instruments) that allow to have further information about the statistical significance of the relationships between the response and predictor variables.
In MNSCA is possible to reconstruct the original frequency table by using the following formula:

$$p_{ijk} = p_{\bullet jk}\left[p_{i\bullet\bullet} + \sum_{\alpha}\left(\frac{1}{\sqrt{\lambda_{\alpha}}}\right)\pi_{\alpha i}\varphi_{jk\alpha}\right] \tag{1}$$

where, λ_{α} is the α^{th}, with $\alpha = \{1,\dots,\min[(I-1);(J\times k)-1]\}$, eigenvalue, $\pi_{\alpha i}$ and $\varphi_{jk\alpha}$ are the row and column coordinates, respectively.

The $\varphi_{jk\alpha}$ represents the coordinates of the $J \times K$ categories on the α^{th} axis and each value includes the interaction effect between the modalities.
In order to eliminate the interaction effect we use the two-way Analysis of Variance without interaction to compute a $\hat{\varphi}_{j,k(\alpha)}$. Then we substitute these last values in (1) and we obtain an estimate matrix $\hat{P}$.
The matrices P and $\hat{P}$ have the same marginal frequencies and the inertias are linked by the following important relationship:

$$In\left[\left(\frac{p_{ijk}}{p_{\bullet jk}} - p_{i\bullet\bullet}\right)p_{\bullet jk}\right] = In\left[\left(\frac{\hat{p}_{ij,k}}{p_{\bullet jk}} - p_{i\bullet\bullet}\right)p_{\bullet jk}\right] + In\left[\left(\frac{p_{ijk}}{p_{\bullet jk}} - \frac{\hat{p}_{ij,k}}{p_{\bullet jk}}\right)p_{\bullet jk}\right]$$

An application with real data will end the paper.