J. Ital. Statist. Soc. (1993)
1, pp. 85-106

A PROBABILISTIC MODEL FOR NONSYMMETRIC CORRESPONDENCE ANALYSIS AND PREDICTION IN CONTINGENCY TABLES

Roberta Siciliano[1]
Università di Napoli Federico II, Italy

Ab Mooijart
Leiden University, The Netherlands

Peter G.M. van der Heijden
University of Utrecht, The Netherlands

Summary

Nonsymmetric correspondence analysis is a model meant for the analysis of the dependence in a two-way contingengy table, and is an alternative to correspondence analysis. Correspondence analysis is based on the decomposition of Pearson's Φ^2-index.

Nonsymmetric correspondence analysis is based on the decomposition of Goodman-Kruskal's τ-index for predictability.

In this paper, we approach nonsymmetric correspondence analysis as a statistical model based on a probability distribution. We provide algorithms for the maximum likelihood and the least-squares estimation with linear constraints upon model parameters.

The nonsymmetric correspondence analysis model has many properties that can be useful for prediction analysis in contingency tables. Predictability measures are introduced to identify the categories of the response variable that can be best predicted, as well as the categories of the explanatory variable having the highest predictability power. We describe the interpretation of model parameters in two examples. In the end, we discuss the relations of nonsymmetric correspondence analysis with other reduced-rank models.

1. *Address for correspondence*: Roberta Siciliano, Dipartimento di Matematica e Statistica, Università di Napoli Federico II, Monte S. Angelo - via Cintia I-80126, Napoli, Italy.

Acknowledgments: The authors are indebted to Carlo Lauro and the referees for helpful comments. This research was partially supported for the first author by MURST research funds.

Keywords: Canonical model, RC canonical correlation model, predictability measures, maximum likelihood estimation, singular value decomposition.

1. Introduction

Canonical models can be useful for studying the relation between the row variable and the column variable in a two-way contingency table. Different types of models can be used depending on the type of relation between the row and the column variables. The case in which both variables play a symmetric role has received much attention, under such names as canonical analysis (Kendall and Stuart, 1979; Gilula and Haberman, 1986), correspondence analysis (Greenacre, 1984; Nishisato, 1980; Escoufier, 1988), and the RC canonical correlation model (Goodman, 1985, 1986). However, if variables play an asymmetric role, and the column variable depends on the row variable, so-called nonsymmetric correspondence analysis (Lauro and D'Ambra, 1984; D'Ambra and Lauro, 1989) seems more appropriate.

Nonsymmetric analysis of contingency tables has a long tradition in the Italian statistical research (see, for example, Gini, 1912; for more recent overviews see D'Ambra and Lauro, 1992; Siciliano, 1992). So far nonsymmetric correspondence analysis has been studied mainly as an exploratory method for the decomposition of contingency tables (Lauro and D'Ambra, 1984; D'Ambra and Lauro, 1989; Lauro and Siciliano, 1989; Siciliano, Lauro and Mooijaart, 1990). In this paper, we approach nonsymmetric correspondence analysis as a statistical model based on a probability distribution.

Let us first introduce the nonsymmetric correspondence analysis model. Let π_{ij} be the expected probability corresponding to cell (i, j) in a contingency table with I rows and J columns $(i = 1, \ldots, I; j = 1, \ldots, J)$. The usual dot notation is used for summation: $\Sigma_j \pi_{ij} = \pi_{i.}$. For the case where the column variable depends on the row variable, nonsymmetric correspondence analysis provides the following reduced-rank decomposition of the matrix with probabilities:

$$\pi_{ij}/\pi_{i.} = \pi_{.j} + \Sigma_m \lambda_m r_{im} c_{jm}, \tag{1}$$

for $m = 1, \ldots, M \leq M^* = min(I-1, J-1)$, and $\lambda_1 \geq \ldots \geq \lambda_M \geq 0$. The score parameters r_{im} and c_{jm} satisfy the following centering and orthonormality conditions:

$$\Sigma_i r_{im}\pi_{i.} = 0, \qquad \Sigma_j c_{jm} = 0, \tag{2a}$$

$$\Sigma_i r_{im}r_{im^*} \pi_{i.} = \delta_{mm^*}, \qquad \Sigma_j c_{jm}c_{jm^*} = \delta_{mm^*}, \tag{2b}$$

where δ_{mm^*} is Kronecker's delta. Decomposition (1) is due to Lauro and D'Ambra (1984). A probability matrix can always be decomposed with (1) if $M=M^*$. Equation (1) shows that nonsymmetric correspondence analysis is concerned with the conditional probabilities $\pi_{ij}/\pi_{i.}$, and the departure from the column margin $\pi_{.j}$ is modeled as a sum of M products of the form $\lambda_m r_{im} c_{jm}$. When the column variable is independent of the row variable then $\pi_{ij}/\pi_{i.} = \pi_{.j}$.

Nonsymmetric correspondence analysis is motivated by the relation of the values λ_m with the predictability index τ proposed by Goodman and Kruskal (1954). We give attention to this index, because it plays a crucial role in nonsymmetric correspondence analysis. The τ index measures the relative increase in probability of correctly predicting the column variable when knowledge about the level of the row variable is used. Explicitly, the τ index is defined as

$$\tau = \Sigma_i \Sigma_j \pi_{i.}(\pi_{ij}/\pi_{i.} - \pi_{.j})^2/(1 - \Sigma_j \pi_{.j}^2). \qquad (3)$$

The numerator of (3) gives the variance of the conditional probabilities around the average $\pi_{.j}$, so it specifies the variance of the prediction of the column variable when we know the level of the row variable. The denominator gives the variance of the prediction of the column variable if we have no information about the level of the row variable. It is clear that $0 \leq \tau \leq 1$, with $\tau = 0$ in case of independence and $\tau = 1$ in case of perfect prediction (Goodman and Kruskal, 1954).

The relation between the values λ_m and the τ index is

$$\tau(1 - \Sigma_j \pi_{.j}^2) = \Sigma_i \Sigma_j \pi_{i.}(\pi_{ij}/\pi_{i.} - \pi_{.j})^2 = \Sigma_m \lambda_m^2. \qquad (4)$$

This shows that nonsymmetric correspondence analysis decomposes the predictability as measured by the τ index into a number of dimensions. The proportion of τ decomposed in each dimension is given by λ_m^2.

The maximum likelihood estimate of the τ index under product-multinomial sampling is equivalent to the R^2 measure of association for the analysis of variance of categorical data proposed by Light and Margolin (1971). The R^2 measure is the ratio of the 'between' and 'total' variation (the latter being proposed by Gini, 1912; also, see Margolin and Light, 1974), thus giving the 'explained' variation of the column variable attributable to the row variable.

The nonsymmetric correspondence analysis is closely related to the RC canonical correlation model (Gilula and Haber, 1986; Goodman, 1986). The two models differ in the weighting systems of row and column scores. In nonsymmetric correspondence analysis, the scores of the response variable are weighted in a different way from the scores of the explanatory variable. This is motivated by a geometrical consideration by Lauro and D'Ambra (1984). Consequently, nonsymmetric correspndence analysis decomposes the predicability τ index whereas canonical analysis decomposes Φ^2 coefficient of contingency. Gilula and Haberman (1988) also consider asymmetric versions of canonical analysis by simply defining the model (1) for the conditional probabilities with the same weighting systems as in canonical analysis. But their approach leads to a decomposition of the Φ^2 contingency coefficient which is used for symmetric relationships and so it is not a nonsymmetric analysis.

This distinct between canonical analysis and nonsymmetric correspondence analysis justifies further interest in nonsymmetric correspondence analysis as a model.

Whereas linear restrictions upon canonical scores have been considered in canonical analysis (Gilula and Haberman, 1988; Wasserman and Faust, 1989; Böckenholt and Böckenholt, 1990; Takane, Yanai and Mayekawa, 1991), they have not been considered in nonsymmetric correspondence analysis so far. In this paper, we describe how to incorporate linear constraints in nonsymmetric correspondence analysis under both the maximum likelihood method and the least-squares criterion. Using the maximum likelihood method allows to test restricted versions of model (1) if certain assumptions about the observations are fulfilled. In section 2 we provide an algorithm for maximum likelihood estimation of nonsymmetric correspondence analysis that is different from algorithms used in similar estimation situations, such as those used for canonical analysis (Goodman, 1985; Gilula and Haberman, 1986). In section 3 we give a least-squares estimation procedure to impose linear constraints upon row and column scores.

Apart from the relation to the τ index, nonsymmetric correspondence analysis has many more aspects that make it a useful model for prediction analysis (for a general introduction to prediction analysis we refer to Hildebrand, Laing and Rosenthal, 1977). This is discussed in detail in section 4 where we introduce some predictability measures and we show the interpretation of model parameters in terms of prediction analysis in two examples.

We end with a comparison of nonsymmetric correspondence analysis with canonical analysis and latent class analysis, finding that in important classes of applications they are equivalent.

2. Maximum likelihood estimation with linear constraints

There is more than one way to estimate the nonsymmetric correspondence analysis model by maximum likelihood. We propose an estimation procedure with two parts. In the first part an iterative algorithm (Siciliano, Mooijaart and van der Heijden, 1990; Siciliano and Moojaart, 1991) gives unique estimates of a rank M matrix with elements $z_{ij} = (\pi_{ij}/\pi_{i.} - \pi_{.j})$. In the cycles of the algorithm we use parameters x_{im} and y_{jm} that satisfy $z_{ij} = \sum_m x_{im}y_{jm}$. These parameters satisfy centering conditions similar to (2a) as the model parameters r_{im} and c_{jm}, but do not follow orthonormality conditions. Thus x_{im} and y_{jm} are not identified, whereas the scalar product $z_{ij} = \sum_m x_{im}y_{jm}$ is identified. Experience with the algorithm in many examples has shown that convergence will be reached without imposing constraints during the iteration process. In the second part of the estimation procedure, after convergence, we identify the estimates of the parameters λ_m, r_{im} and c_{jm} from the estimated reduced-rank matrix z_{ij} using a generalized singular value decomposition.

The algorithm minimizes the objective function $F = -log\ L$, where $log\ L$ is the kernel of the loglikelihood under product multinomial sampling:

$$Log\ L = \sum_i \sum_j p_{ij}\ log\ \pi_{ij}/\pi_{i.} = \sum_i \sum_j p_{ij}\ log(\pi_{.j} + \sum_m x_{im}y_{jm}), \qquad (5)$$

where p_{ij} are the observed proportions. Using the Lagrange multipliers method it can be proved that the maximum likelihood estimate of $\pi_{.j}$ is equal to the observed margin $p_{.j}$. This simplifies the task of finding estimates for the parameters x_{im} and y_{jm}. The parameters are not estimated simultaneously, that is an alternating method is used. In the first step, the x_{im} are estimated for given y_{jm} scores, and then the y_{jm} are estimated for given x_{im} scores. This process will be repeated until convergence has been reached. The first and second derivatives of F are

$$\partial F/\partial x_{il} = -\sum_j p_{ij}y_{jl}/t_{ij}, \qquad (6a)$$

$$\partial F/\partial y_{jl} = -\sum_i p_{ij}x_{il}/t_{ij}, \qquad (6b)$$

$$\partial^2 F/\partial (x_{il})^2 = \sum_j p_{ij}y_{jl}^2/t_{ij}^2. \qquad (6c)$$

$$\partial^2 F/\partial (y_{jl})^2 = \sum_i p_{ij}x_{il}^2/t_{ij}^2, \qquad (6d)$$

for $l = 1, \ldots M$, where $t_{ij} = p_{.j} + \sum_m x_{im}y_{jm}$. The interesting point is that $\partial^2 F/\partial x_{il}\partial x_{i'l} = 0$ for $i \neq i'$, and that $\partial^2 F/\partial y_{jl}\partial y_{j'l} = 0$ for $j \neq j'$ whereas

$\partial^2 F/(\partial x_{il})^2 > 0$ and $\partial^2 F/(\partial y_{jl})^2 > 0$. This means that for each dimension m of the row parameters and for each dimension m of the column parameters the corresponding block-diagonal matrices of the Hessian are diagonal. Each block-diagonal matrix of second derivatives is positive-definite since it is diagonal with positive elements and thus a necessary condition for the convergence is fulfilled. We can make use of this property by choosing a multidimensional Newton algorithm to estimate the parameters. The parameters of each block are estimated alternatingly while the parameters of the other blocks remain fixed. By choosing each time another block of parameters to be estimated convergence to a (local) optimum is guaranteed.

Let **p** be the column vector of one block of parameters to be estimated. Each block of parameters should satisfy centering conditions (2a) and eventual linear restrictions. These side conditions are defined by the set of equations **A**′**p** = **s**, where the number of rows of the matrix **A** depend on the number of parameters to be estimated, and the number of columns of **A** is equal to the number of restrictions. Since the centering conditions (2a) should always be included, the first column of **A** consists of unitary elements and the first element of s is equal to zero. The remaining columns of **A** and corresponding elements of s are fixed in such a way as to obtain linear restrictions.

Linear restrictions often concern equality between two row/column scores for each dimension. This is used as a criterion to test collapsibility of rows/columns (see, for more details, Breiger, 1981; Goodman, 1981; Gilula, 1986; Gilula and Krieger, 1989). For a restriction such as $x_{im} = x_{i'm}$, for $m = 1, \ldots, M$, the matrix **A** has two columns of length I, column 1 consisting of ones and column 2 being zero except for elements i and i' that are 1 and −1 respectively; s is a column vector with two zero elements.

Another type of restriction is that a row/column score is equal to zero for each dimension. This type of restrictions is used to test independence of the conditional probabilities of a row/column. For a restriction as $x_{im} = 0$ for $m = 1, \ldots, M$, the first column of **A** consists of unitary elements, and the second column is zero, except for element i which is one; s is a column vector with two zero elements.

A third and last important type of restriction is that the row parameters and/or the column parameters are equally spaced in a one-dimensional model. We cannot restrict further dimensions in this way because the scores must be orthonormal due to (2b). With this type of restriction the interpretation simplifies considerably: it can be tried out if the categories of the row or the column variable follow some order that is a priori known. Equal spacing means that for instance the parameters x_{il} are known. In the algorithm each cycle then consists of one step.

In iteration k let $\mathbf{p}^{(k)}$ be the current estimate of $\mathbf{p}$ for which the appropriate restrictions hold, led $\mathbf{q}^{(k)}$ be the search vector to be determined, and let $\sigma^{(k)}$ be a step-size parameter. A new estimate of $\mathbf{p}$ is found by $\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} + \sigma^{(k)}\mathbf{q}^{(k)}$. Without loss of generality we fix $\sigma^{(k)} = 1$ although it can be reduced in case the function value increases or improper solutions are found. The search vector $\mathbf{q}^{(k)}$ simply gives the direction in which the estimates of $\mathbf{p}^{(k)}$ have to be improved. In iteration k an estimate $\mathbf{p}^{(k)}$ for which the side conditions holds is found, so that $\mathbf{A}'\mathbf{p}^{(k+1)} = \mathbf{s}$ iff $\mathbf{A}'\mathbf{q}^{(k)} = \mathbf{0}$. Using the Taylor expansion the function near the optimum can be written as

$$F(\mathbf{p}^{(k)} + \mathbf{q}^{(k)}) = F(\mathbf{p}^{(k)}) + \mathbf{g}^{(k)\prime}\mathbf{q}^{(k)} + (1/2)\mathbf{q}^{(k)\prime}\mathbf{H}^{(k)}\mathbf{q}^{(k)}, \qquad (7)$$

where $\mathbf{g}^{(k)}$ and $\mathbf{H}^{(k)}$ are respectively the gradient vector and the Hessian matrix of F in the point $\mathbf{p}^{(k)}$. The procedure is now to find a $\mathbf{q}^{(k)}$ which minimizes (7) under the side conditions $\mathbf{A}'\mathbf{q}^{(k)} = \mathbf{0}$. By introducing a vector of Lagrange multipliers, the following function is minimized

$$\mathbf{g}^{(k)\prime}\mathbf{q}^{(k)} + (1/2)\mathbf{q}^{(k)\prime}\mathbf{H}^{(k)}\mathbf{q}^{(k)} - \boldsymbol{\varepsilon}'^{(k)\prime}(\mathbf{A}'\mathbf{q}^{(k)}), \qquad (8)$$

where $\boldsymbol{\varepsilon}'^{(k)}$ is the column vector of Lagrange multipliers in the step k. In the optimum the first derivatives of (8) are equal to zero. As a result, the set of equations $\mathbf{H}^{(k)}\mathbf{q}^{(k)} - \boldsymbol{\varepsilon}'^{(k)}\mathbf{A} = -\mathbf{g}^{(k)}$ has to be solved together with the equation $\mathbf{A}'\mathbf{q}^{(k)} = \mathbf{0}$. A solution for $\mathbf{q}^{(k)}$ can be shown to be equal to $\mathbf{q}^{(k)} = -\mathbf{W}^{(k)}(\mathbf{H}^{(k)})^{-1}\mathbf{g}^{(k)}$, where $\mathbf{W}^{(k)} = \mathbf{I} - (\mathbf{H}^{(k)})^{-1}\mathbf{A}((\mathbf{A}'(\mathbf{H}^{(k)})^{-1}\mathbf{A})^{-1}\mathbf{A}'$. Because $\mathbf{H}^{(k)}$ is diagonal the elements of $\mathbf{q}^{(k)}$ can be easily written as $(-g_s{}^{(k)}/h_s{}^{(k)} + w_s{}^{(k)})$ for $s = i$ or j, where $g_s{}^{(k)}$, $h_s{}^{(k)}$ and $w_s{}^{(k)}$ are the general terms of $\mathbf{g}^{(k)}$, $\mathbf{H}^{(k)}$ and $\mathbf{W}^{(k)}$ respectively. Consequently, the updating formula for the elements $p_s{}^{(k+1)}$ of $\mathbf{p}^{(k+1)}$ includes the adjustment coefficient $w_s{}^{(k)}$ that guarantees the side conditions to be satisfied for the step $(k + 1)$.

The algorithm requires an initial point from which a feasible search direction can be computed. Starting values of parameters which satisfy the side conditions and eventual linear restrictions can be the first M rescaled eigenvectors of the constrained least-squares estimation of nonsymmetric correspondence analysis (see section 3). But for $M < M^*$ the approximation of the observed conditional proportions can be negative, in which case the loglikelihood function is not defined. A solution to this problem is to make the singular values somewhat smaller in these instances.

The algorithm described in this section is different from algorithms used in similar estimation situations, such as those used for canonical analysis. Goodman (1979, 1985) uses a unidimensional Newton approach for the estimation of the parameters in the one-dimensional canonical model. Gilula and

Haberman (1986; 1988) propose a scoring method for the canonical correlation model. The former is rather simple although convergence of the procedure is slow. The latter method is more complicated but it has a higher convergence rate than Goodman's method. The disadvantage of the latter method is that in each step a (large) matrix has to be inverted. We propose a method which is a higher dimensional extension of the method used by Goodman that uses the fact that the matrices to be inverted are diagonal. This makes inverting a complete matrix not necessary. The difference with Goodman's algorithm is that we give an updating formula that includes an adjustment coefficient that guarantees the side conditions to be satisfied at each iteration.

In the second stage of the estimation procedure, we can identify the estimates of the parameters λ_m, r_{im} and c_{jm} from the estimated reduced-rank matrix z_{ij} by a suitable singular value decomposition. This is done as follows. Collect the estimates of the elements $(\pi_{ij}/\pi_{i.} - \pi_{.j})$ into a matrix $\mathbf{Z}$. Collect the estimates of the parameters x_{im} and y_{jm} in the matrices $\mathbf{X}$ and $\mathbf{Y}$. Collect the estimates of the parameters r_{im} and c_{jm} in the matrices $\mathbf{R}$ and $\mathbf{C}$, the estimates of the parameters λ_m in the diagonal matrix $\boldsymbol{\Lambda}$, and the observed margins $p_{i.}$ in the diagonal matrix $\mathbf{D_r}$. We aim to find that $\mathbf{Z} = \mathbf{R\Lambda C}'$ with the conditions in (2). This can be accomplished by taking the singular value decomposition

$$\mathbf{D_r}^{1/2}\mathbf{Z} = \mathbf{U\Lambda V}', \qquad \mathbf{U}'\mathbf{U} = \mathbf{I}, \qquad \mathbf{V}'\mathbf{V} = \mathbf{I}, \tag{9}$$

and we find $\mathbf{R}$ as $\mathbf{R} = \mathbf{D_r}^{-1/2}\mathbf{U}$ and $\mathbf{C}$ as $\mathbf{C} = \mathbf{V}$. This can be seen immediately by bringing $\mathbf{D_r}^{1/2}$ to the right side of equation (9). Then $\mathbf{Z} = \mathbf{D_r}^{-1/2}\mathbf{U\Lambda V}' = \mathbf{R\Lambda C}'$. It follows that the conditions (2b) are satisfied since $\mathbf{U}'\mathbf{U} = \mathbf{I} = \mathbf{R}'\mathbf{D_r}^{1/2}\mathbf{D_r}^{1/2}\mathbf{R} = \mathbf{R}'\mathbf{D_r}\mathbf{R}$.

Although the columns of $\mathbf{R}$ and $\mathbf{C}$ are orthonormal with respect to $\mathbf{D_r}$ and $\mathbf{I}$, this is not imposed on the columns of $\mathbf{X}$ and $\mathbf{Y}$ to be used in the algorithm. This is not a problem for the linear restrictions that we can impose on model parameters. That is, we can impose such restrictions upon the parameters $\mathbf{X}$ and $\mathbf{Y}$, and then these restrictions will hold for $\mathbf{R}$ and $\mathbf{C}$ if we derive these from $\mathbf{X}$ and $\mathbf{Y}$ through (9).

The goodness of fit of the model (1) with only conditions (2) against the data can be assessed by the usual likelihood-ratio statistic G^2 (see for example Agresti, 1990) which is asymptotically χ^2-distributed with degrees of freedom $df = (I - M - 1)(J - M - 1)$. The number of degrees of freedom in models (1) is equal to the difference between the number of independent cells, being $I(J - 1)$, minus the number of independent parameters, being $[(J - 1) + M(I - 1) + M(J - 1) + M - M(M + 1)]$. It is clear that when

$M = M^* = min(I-1)(J-1)$ the model is saturated $(df = 0)$ and when $M = 0$ (i.e. $\lambda_m = 0$ for each m) there is statistical independence $(df = (I-1)(J-1))$.

We can test model restrictions against the data by using the likelihood-ratio statistic G^2 with degrees of freedom $df = [(I - M - 1)(J - M - 1) + M(s_I + s_J)]$ where s_I and s_J are the number of restrictions upon row and column scores respectively. We can also assess the difference of the goodness of fit between two differently restricted models by using the conditional likelihood-ratio statistic (see for example Agresti, 1990).

3. Least-squares estimation with linear constraints

Nonsymmetric correspondence analysis has been initially introduced as a geometrical model to be used to represent an observed matrix graphically (Lauro and D'Ambra, 1984; D'Ambra and Lauro; Lauro and Siciliano, 1989). A generalized singular value decomposition of the observed matrix with elements $(p_{ij}/p_{i.} - p_{.j})$ in the metrics $\mathbf{D_r}$ and $\mathbf{I}$ allows factorial representations of row and column categories to be made. This is equivalent to (9) (where the estimates of the model parameters are now least-squares estimates). We can make M-dimensional graphical representation of the row points by using the rows of $\mathbf{R\Lambda}$ as coordinates, and for the column points by using the rows of $\mathbf{C\Lambda}$ as coordinates. These lower-dimensional representations are optimal in the sense that $\mathbf{D_r}^{1/2}\mathbf{Z}$ is approximated in a least-squares sense by $\mathbf{U\Lambda V'}$ (compare with Escoufier, 1988). If all the dimensions are considered, the least-squares estimates of the scores obtained with this geometrical approach are identical to the maximum likelihood estimates for the saturated model. Of course, we can also make graphical representations using the maximum likelihood estimates for non-saturated models (Goodman, 1991).

The main difference between the graphical representations made from least-squares estimates of expected probabilities and those made from maximum likelihood estimates of expected probabilities is that in the latter case the dimensionality of the subspaces spanned by the clouds of points is chosen for $M \leqslant M^* = min(I-1, J-1)$ so that it gives an acceptable fit to the observed proportions. Another difference is that maximum likelihood estimation allows to test the fit of restricted models where the restrictions concern the model parameters.

Linear constraints on model parameters can be considered not only in the maximum likelihood estimation but also in the least-squares estimation. Böckenholt and Böckenholt (1990) have recently proposed a generalized least-squares approach without using an iterative algorithm to incorporate linear constraints on the canonical scores of correspondence analysis. We use a similar approach in nonsymmetric correspondence analysis. Linear constraints can be defined as

$$\mathbf{G}'\mathbf{R}^* = \mathbf{0}, \qquad \mathbf{H}'\mathbf{C}^* = \mathbf{0}, \tag{10}$$

where **G** has I rows and K columns, and **H** has J rows and L columns. The matrices **G** and **H** have rank respectively equal to K and L, being K and L the number of linear constraints upon row and column scores respectively. The estimates of the constrained row scores and column scores are collected in the matrices $\mathbf{R}^*$ and $\mathbf{C}^*$ respectively. These scores can be obtained from the singular value decomposition of

$$\{\mathbf{I} - \mathbf{D}_r^{-1/2}\mathbf{G}(\mathbf{G}'\mathbf{D}_r^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{D}_r^{-1/2}\}\mathbf{D}_r^{-1/2}\mathbf{Z}\{\mathbf{I} - \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\} = \mathbf{M}\mathbf{D}_\lambda\mathbf{B}', \tag{11}$$

with $\mathbf{M}'\mathbf{M} = \mathbf{I} = \mathbf{B}'\mathbf{B}$, and $\mathbf{D}_\lambda$ is a diagonal matrix with singular values in descending order. The constrained scores are given by $\mathbf{R}^* = \mathbf{D}_r^{-1/2}\mathbf{M}$ and $\mathbf{C}^* = \mathbf{B}$. Note that the orthonormality conditions (2b) are satisfied since $\mathbf{R}^{*\prime}\mathbf{D}_r\mathbf{R}^* = \mathbf{I}$. It can be shown that (11) is equivalent to (9) in the unrestricted case by setting $\mathbf{G} = \mathbf{D}_r\mathbf{1}$ and $\mathbf{H} = \mathbf{I1}$ (where **1** is a unit vector).

This procedure imposes the same constraints on the row and column scores of each dimension. Another procedure suggested by Böckenholt and Böckenholt (1990) can be used to define for each dimension a different set of linear constraints upon row and column scores. This can be done in nonsymmetric correspondence analysis in the following way. Suppose that we impose different constraints on the scores of the first singular value with respect to the scores of the second singular value. We apply initially the above procedure and we extract the orthonormalized row and column scores associated to the first singular value λ_1. Then we compute the rank-one reduced matrix $\mathbf{Z}_1$ as

$$\mathbf{Z}_1 = (\mathbf{I} - \mathbf{m}_1\mathbf{m}_1')\mathbf{D}_r^{-1/2}\mathbf{Z}(\mathbf{I} - \mathbf{b}_1\mathbf{b}_1'), \tag{12}$$

where $\mathbf{m}_1$ and $\mathbf{b}_1$ are the singular vectors corresponding to λ_1. The matrix $\mathbf{Z}_1$ is substituted to the matrix **Z** in (11) and the new constraints are defined in the matrices **G** and **H**. These latter matrices are augmented by $\mathbf{D}_r^{1/2}\mathbf{m}_1$ and $\mathbf{b}_1$ respectively in order to assure that the new scores are orthonormalized. In this way we extract the row and column scores by applying the (11) again. The procedure can be iterated when more dimensions are considered.

4. Parameter interpretation and prediction analysis

Nonsymmetric correspondence analysis model has some properties that can be useful for prediction analysis, and we give some new results for the interpretation of model parameters. If we consider the weighted average of the conditional probabilities $\pi_{ij}/\pi_{i.}$ for row i, where the weights are given by the column scores c_{jm}, using (2b), we find for $m = 1, \ldots, M$:

$$\Sigma_j(\pi_{ij}/\pi_{i.})c_{jm} = \Sigma_j \pi_{.j}c_{jm} + \lambda_m r_{im}. \tag{13}$$

The left part $\Sigma_j(\pi_{ij}/\pi_{i.})c_{jm}$ of (13) can be interpreted as the predictability of the column variable attributed to the ith row category. When the column variable is not predicted by the row category, $\pi_{ij}/\pi_{i.} = \pi_{.j}$ so that $\Sigma_j(\pi_{ij}/\pi_{i.})c_{jm} = \Sigma_j \pi_{.j} c_{jm}$. As a result the first term on the right hand side of (13), $\Sigma_j \pi_{.j}c_{jm}$, is a constant term that all rows have in common. The effect of the row category in the predictability of the column variable is measured by the second term of (13). The λ_m measures the increase in the predictability of the weighted average of the conditional probabilities given by the row category. In this respect, it can be interpreted as some type of regression coefficient.

The row and column coordinates can be shown to be related to the predictability τ-index as follows:

$$\tau(1 - \Sigma_j\pi_{.j}^2) = \Sigma_m \lambda_m^2 = \Sigma_i \pi_{i.} \Sigma_m(\lambda_m r_{im})^2 = \Sigma_i\Sigma_m(\lambda_m c_{jm})^2. \tag{14}$$

Since the τ-index is a measure for the predictability power of the row variable in a two-way table, and τ is proportional to $\Sigma_m \lambda_m^2$ through (4), (14) shows that we can partition this predictability power over the row categories, over the column categories, and over dimensions. These relations allow us to define the following predictability measures:

$$pred(R_i) = \pi_{i.}\Sigma_m(\lambda_m r_{im})^2/\Sigma_m \lambda_m^2, \qquad \Sigma_i\, pred(R_i) = 1, \tag{15}$$

$$pred(C_j) = \Sigma_m(\lambda_m c_{jm})^2/\Sigma_m \lambda_m^2, \qquad \Sigma_j\, pred(C_j) = 1, \tag{16}$$

$$pred(R_{im}) = \pi_{i.}(\lambda_m r_{im})^2/\Sigma_m \lambda_m^2, \qquad \Sigma_m\, pred(R_{im}) = pred(R_i), \tag{17}$$

$$pred(C_{jm}) = (\lambda_m c_{jm})^2/\Sigma_m \lambda_m^2, \qquad \Sigma_m\, pred(C_{jm}) = pred(C_j), \tag{18}$$

$$pred(D_m) = \lambda_m^2/\Sigma_m \lambda_m^2, \qquad \Sigma_m\, pred(D_m) = 1, \tag{19}$$

where R_i and C_j denote the ith row category and the jth column category, respectively; R_{im} and C_{jm} denote the ith row category and the jth column category respectively represented in the mth dimension; and D_m denote the mth dimension. Using (15) we can distinguish which row categories have most predictability power, whereas using (16) we can distinguish which column categories are the best predicted. Such measures differ from the common contributions used in correspondence analysis techniques (see, for example, Greenacre, 1984). Equations (17) and (18) show that we can measure the percentage of predictability power of each row/column category attributed to each dimension $m = 1, \ldots, M$. Using (19) we can distinguish which dimension retains the highest percentage of the predictability power measured by the τ-index. Note that equations from (17) to (19) are only used for models with more than one dimension.

We describe the interpretation of parameters in terms of prediction analysis in two examples. As a first example we use a data set analysed by Agresti (1984) with logit-linear models. The example considers a cross-classification of police classification of a homicide, court classification of a homicide, race of defendant, and race of victim. The sample consists of 1017 individuals indicted for homicide in Florida between 1973 and 1977. Each case is classified by the police and successively by the court as 'felony', 'possible felony', or 'nonfelony'. This refers to the judgement about whether the homicide was committed concurrently with another felony, such as robbery or rape. The race of the defendant and the race of the victim can be black or white. The matrix is given in Table 1. The rows are the categories of the explanatory variable (that is, a compound variable formed by two variables, namely the race of defendent/victim and the police judgement) and the columns are the categories of the dependent variable (namely the court judgement).

We want to analyse whether the final judgement of the court is influenced by the previous report of the police as well as by the race of the defendant and the race of the victim. For this analysis we apply the nonsymmetric correspondence analysis model in (1), and find that a model with $M = 1$ dimension fits the data adequately since the likelihood-ratio statistic is equal to $G^2 = 15.24$ $(df = 10)$. The τ-index for the estimated table is $\tau = .512$ (for the observed table is .518). In Table 2 we show the estimates of row coordinates $(r_{i1}\lambda^1)$ and of column coordinates $(c_{j1}\lambda_1)$ where $\lambda_1 = .496$ and the corresponding predictability power proportions (compare with (8) and (9)). If we look at the proportions we see that the judgement about possible felony has a very low predictability power when it is expressed by the police (their proportions being .019, .039, .000, .002) and has no influence on the court decision being not predicted (with proportion .000); the police judgement about felony has an higher predictability power when the race of the victim is white in contrast

Table 1

Police and Court Classifications of 1017 Homicides, by Races of Defendent and Victim (Agresti, 1984, p. 133)

Race of Defendent/Victim	Police Classification	Court Classification No Felony	Possible Felony	Felony	*Total*
Black/White	No Felony	7	1	3	11
	Possible Felony	0	2	6	8
	Felony	5	5	109	119
	Total	12	8	118	138
White/White	No Felony	236	11	26	273
	Possible Felony	7	2	21	30
	Felony	25	4	101	130
	Total	268	17	148	433
Black/Black	No Felony	328	6	13	347
	Possible Felony	7	2	3	12
	Felony	21	1	36	58
	Total	356	9	52	417
White/Black	No Felony	14	1	0	15
	Possible Felony	6	1	1	8
	Felony	1	0	5	6
	Total	21	2	6	29

to being black (compare proportions .337 and .209 with proportions .038 and .013); when the victim is black, the police judgement about nonfelony has some influence on the court judgement about nonfelony only when the defendant also is black (compare proportion .225 with proportion .008). If we look at the signs and the absolute values of the coordinates we can see two aspects: first, the court judgement tends to confirm the police judgement

Table 2

Maximum Likelihood Estimation of Nonsymmetric Correspondence Analysis of Data in Table 1

Race of Defendent/Victim	Police Classification	*Unrestricted* Dim 1	*Unrestricted* Pred(R_i)	*Restricted* Dim 1	*Restricted* Pred(R_i)
Black/White	No Felony	−.030	.000	*.000	*.000
	Possible Felony	.728	.019	.710	.018
	Felony	.843	.337	.846	.339
	Total		.356		.357
White/White	No Felony	−.317	.110	−.322	.113
	Possible Felony	.571	.039	.563	.038
	Felony	.636	.209	.634	.208
	Total		.358		.359
Black/Black	No Felony	−.404	.225	−.405	.226
	Possible Felony	−.011	.000	*.000	*.000
	Felony	.404	.038	.401	.037
	Total		.263		.263
White/Black	No Felony	−.361	.008	−.366	.009
	Possible Felony	−.244	.002	*.000	*.000
	Felony	.690	.013	.677	.013
	Total		.023		.022
Total			1.000		1.000

Court Classification		*Unrestricted* Dim 1	*Unrestricted* Pred(C_j)	*Restricted* Dim 1	*Restricted* Pred(C_j)
	No Felony	−.355	.512	−.351	.500
	Possible Felony	.008	.000	*.000	*.000
	Felony	.347	.488	.351	.500
	Total		1.000		1.000

Table 3

Principal Worries of Israeli Adults (Guttman, 1971)

Place of residence/ residence father	Principal Worries ENR	SAB	MIL	POL	ECO	OTH	MTO	PER	Total
Asia/Africa	61	70	97	32	4	81	20	104	469
Europe/America	104	117	218	118	11	128	42	48	786
Israel: father Asia/Africa	8	9	12	6	1	14	2	14	66
Israel: father Europe/America	22	24	28	28	2	52	6	16	178
Israel: father Israel	5	7	14	7	1	12	0	9	55
Total	200	227	369	191	19	287	70	191	1554

ENR = Enlisted relative; SAB = Sabotage; MIL = Military situation; POL = Political situation; ECO = Economic situation; OTH = Other; MTO = More than one worry; PER = Personal economics.

when the decision is about felony or nonfelony; second, the police judgement varies with respect to the cases that the victim is white or black and it gives some preferences to felony.

As some proportions of predictability are nearly equal to zero, we can test a one-dimensional model with restrictions $r_{1,1} = 0$, $r_{8,1} = 0$, $r_{11,1} = 0$, $c_{2,1} = 0$. We find $G^2 = 18.43$ ($df = 14$), so this model still has a good fit and these restrictions may be essentially imposed.

As a second example we discuss an analysis of data published by Guttman (1971) on the principal worries of Israeli adults. Guttman (1971; see also Greenacre, 1984) analysed these data with canonical analysis. The matrix is given in Table 3. The rows consist of the categories of the explanatory variable; five groups of adults are specified by their place of residence and that of their fathers. The column variable is the dependent variable; eight possible principal worries of the adults are specified.

The model with $M = 1$ does not fit well: $G^2 = 29.33$ ($df = 18$). The model with $M = 2$ fits adequately: $G^2 = 6.47$ ($df = 10$). The τ-index for the estimates of expected probabilities as well as for the observed table is $\tau = .013$.

The parameter estimates are given in Table 4. We present a plot of the parameters in Figure 1. We have used $(r_{im}\lambda_m)$ as coordinates for a row point and $(c_{jm}\lambda_m)$ as coordinates for a column point, where $\lambda_1 = .088$ and $\lambda_2 = .053$. Thus, the distances between the row points in the plot are also equal to the distances between the rows in the estimated table of general term $(\pi_{ij}/\pi_{i.} - \pi_{.j})$, and the distances between the column points in the plot are also equal to the distances between the columns in the estimated table. On the other hand, Figure 1 cannot be considered as a biplot (Gabriel, 1971), since the singular values λ_m are 'used' twice, once in $(r_{im}\lambda_m)$ and once in $(c_{jm}\lambda_m)$. Thus we cannot reconstitute the estimated table from Figure 1 (for more details about graphical displays in contingency table analysis, see Goodman, 1991). In Table 4 we give the proportions of predictability associated with the row and column points. This shows that the predictability power of rows 1 (Asia/Africa), 2 (Europe/America), and 4 (Israel: father; Europe/America) is large, their proportions being .340, .324 and .214. Row categories 3 and 5 have a relatively small predictability power due to the fact that their weights $p_{i.}$ are relatively small. Column 8 (personal economics) is predicted best, with a proportion of τ of .514, and columns 3 (military situations), 4 (political situation), and 6 (other) are also relatively well-predicted by the place of residence of the adults and their fathers. Compared to the other categories, categories 1 (enlisted relative), 2 (sabotage), and 5 (economic situation) and not well-predicted by the residence of the adults and their fathers. The way in which the place of residence of the adults and their fathers predicts the worries is clearly shown in Figure 1. Living in Asia/Africa

Table 4

Maximum Likelihood Estimation of Nonsymmetric Correspondence Analysis of Data in Table 3

	Unrestricted Row Coordinates			*Restricted Row Coordinates*		
	Dim1	Dim2	Pred(R_i)	Dim1	Dim2	Pred(R_i)
1	.115	−.028	.340	.116	−.025	.410
2	−.080	−.021	.324	−.081	−.019	.331
3	.108	.032	.051	.106	.026	.049
4	−.004	.140	.214	.000	.138	.207
5	.042	.041	.011	.029	.000	.003

	Unrestricted Column Coordinates			*Restricted Column Coordinates*		
	Dim1	Dim2	Pred(C_j)	Dim1	Dim2	Pred(C_j)
1	−.002	−.004	.002	*.000	*.000	*.000
2	−.001	−.005	.003	*.000	*.000	*.000
3	.032	−.029	.178	−.037	−.032	.233
4	−.036	.013	.139	−.038	.011	.148
5	−.002	.000	.000	*.000	*.000	*.000
6	.007	.040	.158	.000	.035	.120
7	−.006	−.006	.007	*.000	*.000	*.000
8	.073	−.011	.514	.071	−.014	.500

Unrestricted model

	pred(R_{i1})	pred(R_{i2})	pred(C_{j1})	pred(C_{j2})
1	.377	.022	.000	.001
2	.303	.020	.000	.002
3	.046	.004	.010	.078
4	.000	.214	.122	.017
5	.006	.005	.000	.000
6			.004	.154
7			.004	.003
8			.503	.010
pred(D_m)	.732	.266	.643	.265

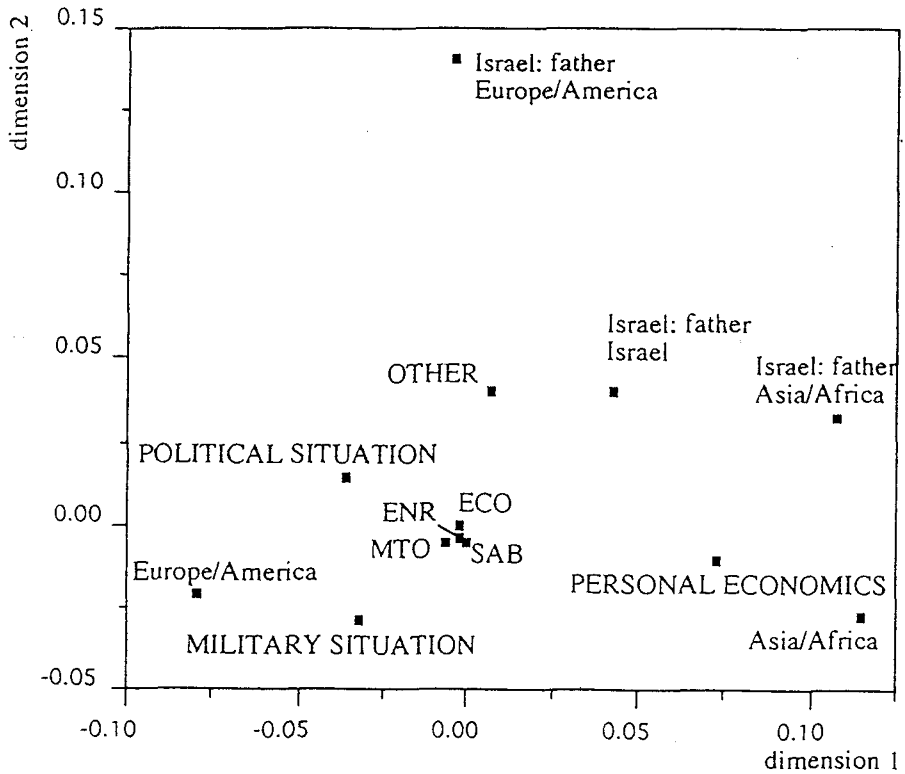

Fig. 1 – Principal worries of Israeli adults (Guttman, 1971), Nonsymmetric correspondence analysis model with two dimensions, explanatory categories lower-case.

or having a father living there predicts the adults to be worried about their personal economy with a relatively larger probability, whereas living in Europe/America predicts the adults to be worried about the military and the political situations with a relatively larger probability. Living in Israel and having a father living in Europe/America (and, to a lesser extent, living in Israel and having a father also living there) predicts 'other' worries with a relatively larger probability. The graphical display clearly shows that the column categories ENR (enlisted relative), ECO (economic situation), MTO (more than one worry), and SAB (sabotage) are not well predicted in terms of the τ-index (i.e., they do not contribute much to its value). For 'economic situation' this is also the result from its low marginal proportion.

The previous four column categories are close to the origin in Figure 1. We can test a two-dimensional model with restrictions $c_{1m} = 0$, $c_{2m} = 0$, $c_{5m} = 0$

and $c_{7m} = 0$ to verify whether these short distances to the origin are significant, or in other words, the prediction of these column categories is independent from the information about row categories. We find $G^2 = 10.67$ $(df = 18)$, so this model has still a good fit and the imposition of these restrictions may be reasonable. The column categories ENR (enlisted relative), ECO (economic situation), MTO (more than one worry), and SAB (sabotage) are not well predicted.

5. Relation with canonical analysis and latent class analysis

We end this paper by discussing the relationship between nonsymmetric correspondence analysis and other reduced-rank models for contingency tables, namely canonical correlation model and latent class model.

The nonsymmetric correspondence analysis model is closely related to the canonical correlation model. As we have discussed in section 1 the two models differ in the weighting systems. Consequently, reduced-rank approximations of the observed proportions under the least-squares criterion are different. Instead, the maximum likelihood estimates of expected probabilities are the same under the two models, and thus the nonsymmetric correspondence analysis model and the canonical correlation model provide the same fit to the data.

Due to different weighting systems the two models differ in the orthonormality conditions used to identify the model parameters. The parameter estimates of the canonical correlation model can be identified by a singular value decomposition different from (9) at which we should add a diagonal matrix with the elements $1/\sqrt{p_{.j}}$. Therefore, the identified estimates of the model parameters and thus the geometrical representation of the relations between row and column categories will be in general different in the two models. They can be equivalent only in two cases: when $M=1$ since in such case the orthonormality conditions are not needed to identify the parameter estimates, and when $p_{.j}$ are equal to $1/J$ for each j so that the uniform marginal distribution does not provide any difference in the singular value decompositions.

Both canonical analysis and nonsymmetric correspondence analysis are also related to latent class analysis and latent budget analysis (for an overview of these models see Clogg, 1982; van der Heijden, Mooijaart and de Leeuw, 1989). The latent budget model is a reparametrization of the latent class model to be used for the analysis of the dependence.

All these models provide reduced-rank decompositions of a matrix with probabilities (Good, 1969; de Leeuw and van der Heijden, 1991; Siciliano, 1992). The latent class models and the latent budget models provides a non-negative rank-M approximation of the probability matrix. The canonical model and the nonsymmetric correspondence analysis model provide a rank-M approximation of the probability matrix using parameters that can also be negative. De Leeuw and van der Heijden (1991) show that for rank 2 the latent class model and the canonical model imply each other, that is they provide identical maximum likelihood estimates of expected probabilities; for rank larger than 2, if the latent class model is true, then the canonical model is true, but the reverse does not necessarily hold. Because nonsymmetric correspondence analysis is equivalent to canonical analysis, these results are directly relevant for nonsymmetric correspondence analysis also.

So we conclude that there are four models providing a reduced-rank approximation of probabilites. Nonsymmetric correspondence analysis and latent budget analysis are models for the situation where the column variable is a response variable, and the row variable is an explanatory variable; the canonical model and latent class analysis are models for the situation where both variables play a simmetric role – although the latter models are regularly used in practice for the analysis of asymmetric tables. A choice between these models should be made, first, on the basis of the type of relation between the manifest variables, and, second, on the type of parameters that is considered to be most useful for the particular application.

REFERENCES

AGRESTI, A. (1984) *Analysis of ordinal categorical data*. New York: Wiley.

AGRESTI, A. (1990) *Categorical Data Analysis*. New York: Wiley.

BÖCKENHOLT, U. and BÖCKENHOLT, I. (1990) Canonical analysis of contingency table with linear constraints. *Psychometrika*, 55, 633-639.

BREIGER, R. L. (1981) The social class structure of occupational mobility. *American Journal of Sociology*, 87, 578-611.

CLOGG, C. C. (1982) Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, 79, 762-771.

D'AMBRA, L. and LAURO, N. C. (1989) Nonsymmetrical analysis of three-way contingency tables. *Multiway data analysis* (eds. R. Coppi and S. Bolasco), 301-315. Amsterdam: North Holland.

D'AMBRA, L. and LAURO, N. C. (1992) Non symmetrical exploratory data analysis. *Statistica Applicata. Italian Journal of Applied Statistics*, 4, 4, Napoli: Curto.

DE LEEUW, J. and VAN DER HEIJDEN, P. G. M. (1991) Reduced-rank models for contingency tables. *Biometrika*, 78, 239-242.

ESCOUFIER, Y. (1988) Beyond correspondence analysis. *Classification and related methods of data analysis* (ed. H. H. Bock), 505-514. Amsterdam: North Holland.

GABRIEL, K. R. (1971) The biplot-graphic display of matrices with applications to principal component analysis. *Biometrika*, 58, 453-467.

GILULA, Z. (1986) Grouping and association in contingency tables: an exploratory canonical correlation approach. *Journal of the American Statistical Association*, 81, 773-779.

GILULA, Z. and HABERMAN, S. J. (1986) Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association*, 81, 780-788.

GILULA, Z. and 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.

GILULA, Z and KRIEGER, A. M. (1989) Collapsed two-way contingency tables and the chi-square reduction principle. *Journal of the Royal Statistical Society, Series B*, 51, 425-433.

GINI, C. (1912) Variabilità e mutabilità. Contributi allo studio delle relazioni e delle distribuzioni statistiche. *Studi Economico-Giuridici della Università di Cagliari.*

GOOD, I. J. (1969) Some applications of the singular value decomposition of a matrix. *Technometrics*, 11, 823-831.

GOODMAN, L. A. (1979) Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 70, 755-768.

GOODMAN, L. A. (1981) Criteria for determining whether certain categories in a cross-classification table should be combined with special reference to occupational categories in occupational mobility tables. *American Journal of Sociology*, 87, 612-650.

GOODMAN, L. A. (1985) The analysis of cross-classified data having ordered and/or unordered categories: association modes, correlation models and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics*, 13, 10-69.

GOODMAN, L. A. (1986) Some useful extensions to the usual correspondence analysis approach and the usual loglinear approach in the analysis of contingency tables (with comments). *International Statistical Review*, 54, 243-309.

GOODMAN, L. A. (1987) New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *American Journal of Sociology*, 93, 529-583.

GOODMAN, L. A. (1991) Models, measures, and graphical displays in the analysis of contingency tables (with discussion). *Journal of the American Statistical Association*, 86, 1085-1138.

GOODMAN, L. A. and KRUSKAL, W. H. (1954) Measures of association for cross-classifications. *Journal of the American Statistical Association*, 49, 732-764.

GUTTMAN, L. (1971) Measurement as structural theory. *Psychometrika*, 36, 329-347.

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

HILDEBRAND, D. K., LAING, J. D. and ROSENTHAL, H. (1977). *Prediction analysis of cross-classifications*. New York: Wiley.

KENDALL, M. G. and STUART, A. (1979) *The advanced theory of statistics*, Vol. 2 (4th Ed.). New York: Hafner.

LAURO, N. C. and D'AMBRA, L. (1984) L'analyse non symmetrique des correspondances. *Data analysis and informatics III* (eds. E. Diday et al.), 433-446. Amsterdam: North Holland.

LAURO, N. C. and SICILIANO, R. (1989) Exploratory methods and modelling for contingency tables analysis: an integrated approach. *Statistica Applicata. Italian Journal of Applied Statistics*, 1, 5-32.

LIGHT, R. J. and MARGOLIN, B. H. (1971) An analysis of variance for categorical data. *Journal of the American Statistical Association*, 66, 534-544.

MARGOLIN, B. H. and LIGHT, R. J. (1974) An analysis of variance for categorical data, II: small sample comparisons with chi square and other competitors. *Journal of the American Statistical Association*, 69, 755-764.

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

SICILIANO R. (1992). Reduced-rank models for dependence analysis of contingency tables. *Statistica applicata. Italian Journal of Applied Statistics*, 4, 4, Napoli, Curto.

SICILIANO R. and MOOIJAART, A. (1991) *An algorithm for maximum likelihood estimation of generalized reduced-rank models for contingency table analysis*. Technical Report - PRM 04-91, University of Leiden.

SICILIANO, R., LAURO N. C. and MOOIJAART, A. (1990) Exploratory approach and maximum likelihood estimation of models for nonsymmetrical analysis of two-way multiple contingency tables. *Compstat '90*, 157-162.

SICILIANO, R. MOOIJAART, A. and VAN DER HEIJDEN, P. G. M. (1990) *Non-symmetric correspondence analysis by maximum likelihood*. Technical Report - PRM 05-90, University of Leiden.

TAKANE, Y., YANAI, H. and MAYEKAWA, S. (1991) Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, in press.

VAN DER HEIJDEN, P. G. M., MOOIJAART, A. and DE LEEUW, J. (1989) Latent budget analysis. *Statistical Modelling* (eds. A. de Carli et al.). Berlin: Springer Verlag.

WASSERMAN, S. and FAUST, K. (1989) Canonical analysis of the composition and structure of social networks. *Sociological methodology 1989* (ed. C. C. Clogg), 1-42. London, Basil Blackwell.