Nonsymmetric Correspondence Analysis: A Tutorial

Pieter M. Kroonenberg
Departement Pedagogische Wetenschappen, Rijksuniversiteit Leiden, Leiden
Rosaria Lombardo
Institute of Marketing and Quantitative Methodology, Università di Napoli II, Capua, Italia

March 24, 1998

Abstract

In this paper the basic theory of nonsymmetric correspondence analysis is presented in a fairly straightforward fashion. In the Appendix the technique is illustrated in great numerical detail with a miniature example. A larger scale example is presented in the body of the paper. Both examples are concerned with the attachment bond between mothers and their infants.

Keywords:
Attachment, Biplots, Categorical Analysis of Variance, Categorical Regression, Goodman & Kruskal's Tau, Singular Value Decomposition

Acknowledgements:
The authors would like to thank Prof. D.C. van den Boom for the use of her data in the first illustration and Prof. M.H. van IJzendoorn for the use of his data in the Appendix.

Requests for reprints should be addressed to:
P.M. Kroonenberg, Department of Educational Sciences, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands
e-mail: kroonenb@rulfsw.fsw.leidenuniv.nl

INTRODUCTION

In this paper, we present the basic of the theory behind NonSymmetric Correspondence Analysis (NSCA) as developed by Lauro and D'Ambra of the University of Naples. They devised techniques especially geared towards the analysis of dependence for contingency tables with few categorical variables having a moderate to large number of categories (e.g. Lauro and D'Ambra, 1984; D'Ambra and Lauro, 1989, 1992). The explicit aim of their technique is to cater for the analysis of dependence for two or three categorical variables, but extensions to more variables exist as well. NSCA is primarily an exploratory tool, but the basic measures of increase in predictability can be tested as well.

The main purpose of the paper is to present the theory and background in a straightforward didactical manner. The original publications are primarily mathematical rather than conceptual, and not very oriented towards applications. The technique will be illustrated with data describing the changes in the interaction between mother and her infant during the first six month of the infant's life (Van den Boom, 1988; Van den Boom and Hoekstra, 1994). In the Appendix a detailed presentation is given of the type of information nonsymmetric correspondence analysis generates using a small data set about the predictive relationship of adult attachment classifications with respect to infant attachment classifications (Van IJzendoorn, 1995).

GOODMAN AND KRUSKAL'S τ

The measure of predictability which lies at the heart of nonsymmetric correspondence analysis is Goodman and Kruskal's τ (Goodman and Kruskal 1954, p. 759), but Goodman and Kruskal indicate that the measure was suggested to them by W. Allen Wallis. Light and Margolin (1971) derived independently the intra-class correlation measure for categorical data, which can be shown to be the same as Goodman and Kruskal's τ. There exist two versions of τ, viz. τ_b, when predicting from rows to columns and τ_a when predicting from columns to rows. As in any table columns and rows can be interchanged we will only consider one situation, in particular prediction from columns to rows, and thus always use τ_a, but we will refer to it simply as τ.

Predictability

We will indicate the rows variable (i.e. the criterion variable) with Y. The predictor variable will be indicated with Z. There are I rows, thus the number of categories of Y is I, and there are J columns, and thus Z has J categories. Our basic data are relative frequencies $\hat{p}_{ij}$ which estimate corresponding probabilities $p_{ij}(i = 1,\ldots,I; j = 1,\ldots,J)$, and which may be collected in the matrix $\mathbf{P}$ with $\sum_i \sum_j p_{ij} = 1$. For simplicity, we will in the sequel refer to both as probabilities even though the observed proportions are only estimates of real probabilities. A value p_{ij} thus indicates the probability that an arbitrary individual will be in row i and column j . The overall or marginal probability of a row i is the probability that an individual is in row i irrespective of the column he or she belongs to, and it is solely based on the marginal distribution of Y. It is equal to $\Pr[i] = p_{i.} = \sum_j p_{ij}$. The relative sizes of the columns are the marginal probabilities $\Pr[j] = p_{.j} = \sum_i p_{ij}$. The (conditional) probability of being row i given that the individual belongs to column j is $p_{ij}/p_{.j}$; the set of all conditional probabilities in a column j will be called a *column profile.* As per column the elements of a profile sum to one, a column profile is the probability distribution for that column. If we sum the column profiles weighting each profile by its probability $p_{.j}$ then

$$\text{(1)} \qquad \sum_{j=1}^{J} p_{.j} \left(\frac{p_{ij}}{p_{.j}} \right) = p_{i.} \quad (i = 1, \ldots, I),$$

so that we see that the marginal profile is the weighted average of the column profiles. If the column profiles are similar to the marginal profile then the distribution in a column does not deviate much from the overall distribution. If the difference is large, then some rows occur proportionally more often in this column than one would expect from the overall proportion of the row, and some rows will be lower then one would expect from the marginal distribution. Thus to investigate the degree of deviation from the marginal distribution, the column profiles have to be compared with the overall profile. The difference between a column profile and the marginal profile, or the difference profile, will be indicated by Π_{ij},

$$\text{(2)} \qquad \Pi_{ij} = Pr[i|j] - Pr[i] = \frac{p_{ij}}{p_{.j}} - p_{i.}\,.$$

It is useful to have an overall measure which indicates the overall predictability of the rows by the columns. A very common class of such measures are based on the *proportional reduction of error*, and they generally take the form of the reduction in error variance over a total error variance. In the present case we predict that an individual will be in a row i with probability $p_{i.}$ if his or her column is unknown, but we predict $p_{ij}/p_{.j}$ when we know that the column is j. Thus for each row-column combination the reduction in error of prediction due to the knowledge of an individual's column is Π_{ij}. By squaring this and weighting each cell with $p_{.j}$, i.e. the probability of belonging to column j, we obtain the expression for the 'reduction of error' variance or between-columns variance

$$\begin{aligned}
\text{(3)} \qquad Var_{between} &= \frac{1}{2} \sum_i \sum_j p_{.j} \Pi_{ij}^2 \\
\text{(4)} \qquad &= \frac{1}{2} \sum_i \sum_j p_{.j} \left(Pr[i|j] - Pr[i] \right)^2 \\
\text{(5)} \qquad &= \frac{1}{2} \sum_i \sum_j p_{.j} \left(\frac{p_{ij}}{p_{.j}} - p_{i.} \right)^2,
\end{aligned}$$

where the half is needed to make it a proper variance (see Light and Margolin 1971). The total error of prediction when the individual columns are unknown can be derived as follows. Given that an individual is in a row, we predict with a probability $p_{i.}$ that this row is i, and the probability that the prediction is correct is also $p_{i.}$, thus the probability of predicting that an individual is in row i and that this prediction is correct is $p_{i.}^2$. The total proportion of correct predictions is $\sum_i p_{i.}^2$, and therefore the total proportional error of predictions is $1 - \sum_i p_{i.}^2$, multiplying by $\frac{1}{2}$ gives again a proper variance. Combining this information, the measure for the proportional reduction of error in prediction or proportional increase in predictability becomes

$$(6) \qquad \tau \; = \; \frac{\sum_i \sum_j p_{.j} \Pi_{ij}^2}{1 - \sum_i p_{i.}^2} = \sum_j p_{.j} \left[\frac{\sum_i \Pi_{ij}^2}{1 - \sum_i p_{i.}^2} \right] = \sum_j \tau_j \ .$$

This measure is known in the literature as Goodman and Kruskal's τ, and the τ_j are the contributions of the j^{th}-column to the value of τ. Note that the contributions of the columns to τ involve the weights $p_{.j}$. Thus, columns with deviating difference profiles contribute more to the measure of increase in predictability if there are many individuals in a column. These contributions to τ can be used to assess whether a column is worth concentrating on, because only deviating columns which form a sizeable part of the population will have large contributions.

When all column distributions are identical to overall distribution, then there is no relative increase in predictability, and thus τ is zero (see also Lauro and D'Ambra 1984). Similarly, if knowing that an individual belongs to column j implies knowing which row he or she is in, then τ is at its maximum value.

Strength of dependence: τ as intra-class coefficient

Margolin and Light (1974) (see also Lauro and D'Ambra, 1984) discuss several other properties of τ (which is also called the *concentration coefficient*, see Agresti, 1990, p. 24), and its differences with the Pearson's mean-square contingency coefficient or inertia which is at the heart of regular correspondence analysis. Following Light and Margolin (1971), we now show in a different way that τ can be seen as a ratio of the between-column variability and the total variability in the contingency table, and thus as an intra-class coefficient.

Whereas regression analysis assumes that both the criterion and the predictor are continuous, analysis of variance treats the dependence of a continuous criterion on a categorical predictor. The improvement of knowledge about the criterion based on the values of the predictor is measured by evaluating the relative size of the variance between the categories of Y with that of the total variance of Y. To transfer this view to the case of two categorical variables, one has to decide how to measure variation. For this purpose, as worked out by Light & Margolin (1971) who called their approach CATegorical ANAlysis of variance (CATANA), one should not use the usual formula for the variance based on squared deviations from a mean, but a formula, which apparently goes back to Huygens, based on the squared differences between all different pairs of observations (see Gini, 1912; e.g. cited in Light & Margolin, 1971, and D'Ambra & Lauro, 1989). In particular, one can define the difference between two categorical observations as 0 if they are in the same category and 1 if they are in different categories. Using this definition of difference, the (total) variability of Y, Var_{tot} can be expressed as

$$(7) \qquad Var_{tot} = \frac{1}{2}\sum_i p_{i.}(1 - p_{i.}) = \frac{1}{2}\left(1 - \sum_i p_{i.}^2\right).$$

Using the same measure of variation for each column and adding all columns gives the within variability of the set of columns.

$$(8) \qquad Var_{within} = \frac{1}{2}\left(1 - \sum_i \sum_j \frac{p_{ij}^2}{p_{.j}}\right).$$

Subtracting the Var_{within} from Var_{tot} the desired between variability $Var_{between}$ becomes

$$(9) \qquad Var_{between} = \frac{1}{2}\left(\sum_i \sum_j \frac{p_{ij}^2}{p_{.j}} - \sum_i p_{i.}^2\right),$$

which equation is identical to equation (3). The desired intra-class coefficient for measuring the dependence between Y and X is then $Var_{between}/Var_{tot}$. And thus we end up again with Goodman and Kruskal's τ.

Sensitivity of τ

As Agresti (1990, p. 25) indicates there are some problems with the sensitivity of τ to different probability distributions.

> "[a] difficulty with [τ] is in determining how large a value constitutes a "strong" association. When the response variable has several possible categorizations, these measures tend to take smaller values as the number of categories increases. For instance, for τ the variation measure is the probability that two independent observations occur in different categories. Often this probability approaches 1.0 for both the conditional and marginal distributions as the number of response categories grows larger, in which case τ decreases towards 0."

Such considerations should be kept in mind when one is using τ in an analysis. However, there is some guidance to be had from an asymptotic test of τ against independence. In particular, $U^2 = (N-1)(I-1)\tau$ is under the null hypothesis of independence asymptotically chi-square distributed with $(I-1)(J-1)$ degrees of freedom, assuming the observations in the cells are independent (Light and Margolin 1971). This at least gives some protection against embarking on a detailed analysis of the interaction when there is none. As Light and Margolin (1971, p. 540) and Bishop, Fienberg, and Holland (1975, p. 392) indicate, there might be a statistically significant association in large data sets, which only involves a very small amount of explained variation. Thus, if the values of τ are low but significant the predictability or dependence structure should be analysed further. Of course, this does not mean that a meaningful result will necessarily obtain. However, nonsignificance indicates that there is no point to such an analysis.

NONSYMMETRIC CORRESPONDENCE ANALYSIS

The two primary purposes of nonsymmetric correspondence analysis in the present setting are (1) to find a reordering of the rows and columns of the contingency table to bring its structure to light and (2) to construct a display in a lower dimensional space which portrays the principal dependence of the rows (criterion variable) Y on the columns (predictor variable) Z in such a way that most of the relationship between the two variables is in the display.

In the required display we will portray the rows and columns, and the way they are arranged in the graph will be in accordance with the changes in predictability that are of prime interest. Thus the graph needs to be constructed in such a way that it becomes obvious which column distributions deviate considerably from the overall or marginal distribution, and which columns have more or less the same distribution as the overall one. Nonsymmetric correspondence analysis will provide the appropriate coordinates for the rows ($\mathbf{Y}_i$), and those for the columns ($\mathbf{Z}_j$). In this section we derive these coordinates and show some of their most important properties.

Co-ordinates for rows and columns

To find a representation for the increase in predictability in such a way that most of the explained variability is concentrated in a low-dimensional Euclidean space (i.e. a space with the standard Euclidean distance function), the numerator of τ will be decomposed, which is the absolute rather than the relative increase in predictability. To see how one can arrive at such a low-dimensional representation, we consider the numerator of equation (6) in the form

$$N_\tau = \sum_i \sum_j p_{.j}\Pi_{ij}^2, \tag{10}$$

so that it is seen that N_τ is the weighted Euclidean norm (or variance) of $\mathbf{\Pi} = (\Pi_{ij})$ with the probability that an individual belongs to column j, $p_{.j}$, as weights. Marcotorchino (e.g. 1984) refers to this numerator as the "Light-Margolin Haldane's criterion". Note that it is $2 \times Var_{between}$ of equation (3). In order to find the (full) rank S_o representation of the columns and the rows, we can fruitfully make use of a generalisation of the singular value decomposition (SVD; Eckart and Young 1936). In a similar manner, Greenacre (1984, p. 39) describes a form of generalised SVD for ordinary correspondence analysis using weighted metrics for both rows and columns. A variant of that generalisation for nonsymmetric correspondence analysis, referred to here as GSVD, has the following form.

$$\mathbf{\Pi} = \sum_{s=1}^{S_0} \lambda_s \mathbf{a}_s \mathbf{b}'_s\ , \tag{11}$$

where the scalar λ_s is the singular value, and the $\mathbf{a}_s$, the $\mathbf{b}_s$ are orthonormal singular vectors in an unweighted and weighted metric, respectively, i.e.

$$\sum_{i=1}^{I} a_{is}a_{is'} = 0 \text{ if } s \neq s', \text{ and } = 1 \text{ if } s = s' , \tag{12}$$

and

$$\sum_{j=1}^{j} p_{.j}b_{js}b_{js'} = 0 \text{ if } s \neq s', \text{ and } = 1 \text{ if } s = s' . \tag{13}$$

The difference with the correspondence analysis generalisation is that the weighted metric is only used for the columns and not for the rows. One effect of this is that rows with small proportions will have little influence on the analysis in contrast with ordinary correspondence analysis. Note that the Π_{ij}, i.e the changes in prediction probabilities, are approximated by the singular value decomposition, and not the original probabilities p_{ij}.

Low-rank approximation of $\mathbf{\Pi}$. Similarly to the SVD, the best rank S approximation of $\mathbf{\Pi}$, say $\widehat{\mathbf{\Pi}}^{(S)}$, can be obtained from the GSVD of $\mathbf{\Pi}$ by summing only the first S terms of equation (11),

$$\widehat{\mathbf{\Pi}}^{(S)} = \sum_{s=1}^{S} \lambda_s \mathbf{a}_s \mathbf{b}'_s . \tag{14}$$

Such an approximation leads to the so-called *reconstruction formula*

$$\widehat{p_{ij}}^{(S)} = p_{i.}p_{.j} + p_{.j} \sum_{s=1}^{S} \lambda_s a_{is} b_{js} , \tag{15}$$

which gives the approximated values of the probabilities obtained from the rank-S approximation. One way of looking at this formula is to say that the approximated value is based on its value under the model of independence plus the weighted approximate change in prediction probability.

Biplot representations of the two-rank approximation

In order to display a low rank approximation of $\mathbf{\Pi}$ we will use biplots (Gabriel 1971) rather than the more common correspondence analysis plots. In the standard plots in correspondence analysis, we have both an isometric (or *metric preserving*) representation of the rows and an isometric representation of the columns (Gabriel and Odoroff 1990), but we do not have a correct representation of the row-column relations. With metric preserving we mean that the Euclidean distances in the biplot are correctly represented. In the biplot representation, we have a correct representation of the row-column relationship combined with either an isometric representation of the rows, or an isometric representation of the columns. As, in contrast to ordinary correspondence analysis, NSCA aims to portray the way the columns predict the rows rather than representing the symmetric relationship between rows and columns, in NSCA column-isometric biplots should be used.

The appropriate factorisation of $\widehat{\mathbf{\Pi}}^{(2)}$ derived from the GSVD (11) is called the two-dimensional *isometric factorisation of the columns*, and takes the following form,

$$\widehat{\Pi_{ij}}^{(2)} = \sum_{s=1}^{2}(a_{is})(\lambda_s b_{js}) = \sum_{s=1}^{2} y_{is} z_{js} \, . \tag{16}$$

In Greenacre's (1993, pp. 60, 61) terminology, the rows have *standard coordinates* - y_{is}, and the columns have *principal coordinates* - z_{js}. In the biplot, the row points and the column points are defined by the coordinates

$$\mathbf{Y}_i = (a_{i1}, a_{i2}) \text{ and } \mathbf{Z}_j = (\lambda_s b_{j1}, \lambda_s b_{j2}),$$

respectively.

Transition formulae and barycentric properties. The biplot can be given a *barycentric* interpretation using the so-called *transition formulae*, as is commonly done in ordinary correspondence analysis, especially in the original French literature (see e.g. Benzécri, 1973). In particular, the transition formula for the column coordinates as function of those of the rows is

$$z_{js} = \lambda_s b_{js} \quad = \quad \sum_i \Pi_{ij} a_{is} = \sum_i \left(\frac{p_{ij}}{p_{.j}} - p_{i.} \right) a_{is} \tag{17}$$

The barycentric (or weighted mean or centroid) interpretation rests on the fact that the coordinates of a column j, $z_{js} = \lambda_s b_{js}$, are at the centroids of the coordinates of the rows, $y_{is} = a_{is}$, with weights equal to the $(i,j)^{th}$ element of the difference profile. In nonsymmetric correspondence analysis, the barycentric property for the row coordinates $\mathbf{a}_s$ is not very useful because the associated biplots only portray the changes in prediction probabilities for the rows given the columns and not vice versa.

The barycentric property is such that when two column points $\mathbf{Z}_j$ and $\mathbf{Z}_{j'}$ are located in different parts of the biplot the weights of the $\mathbf{Y}_i$ must be different. However, it is not necessarily true that when two column points are close together in a plane the weight distributions of the rows are comparable, because a centroid can arise from different distributions of weights for the row points $\mathbf{Y}_i$. The uncertainty about the weight distributions introduces an undesirable uncertainty in the interpretation. It seems that the *standard biplot interpretation* based on inner products, is more convenient and more precise. In this interpretation, the size of the difference of a row i in column j, Π_{ij}, is approximated by the inner product between $\mathbf{y}_s$ and $\mathbf{z}_s$ (see equation (16)). If $||\vec{OZ}||$ indicates the length of the vector $\vec{OZ}$ from the origin to a point $\mathbf{Z}$, then the approximated value of Π_{ij}, $\widehat{\Pi_{ij}}$, is equal to $||\vec{OZ_j}||||\vec{OY_i''}||$, where $\mathbf{Y}_i''$ is the orthogonal projection of $\mathbf{Y}_i$ onto the axis $(\vec{OZ_j})$. $\widehat{\Pi_{ij}}$ is thus proportional to the length of $\vec{OY_i''}$, and it is positive if the angle $(\vec{OY_i''}, \vec{OZ_j})$ is acute (the vectors have the same direction), negative in the case of an obtuse angle, and null if the angle is orthogonal. A positive value indicates that a row has a higher value for the column than its marginal value, and a negative value indicates a comparatively smaller value.

Distances

In a geometric context, one of the aims of nonsymmetric correspondence analysis is to portray the difference (or centred) profiles of the columns in a (weighted) Euclidean space, the column space. Of prime interest is thus how the difference between columns j and j', is represented in the column space. The squared distance between the difference profiles of j and j' is defined as

$$\begin{aligned} d_{jj'}^2 &= \sum_i \left[\left(\frac{p_{ij}}{p_{.j}} - p_{i.} \right) - \left(\frac{p_{ij'}}{p_{.j'}} - p_{i.} \right) \right]^2 \\ &= \sum_i \|\Pi_{ij} - \Pi_{ij'}\|^2. \end{aligned}$$

By using the full-rank decomposition of Π_{ij} from equation (11) we get

$$d_{jj'}^2 = \sum_{ss'} \lambda_s \lambda_{s'} \left(b_{js} - b_{j's} \right) \left(b_{js'} - b_{j's'} \right) \sum_i a_{is} a_{is'}, \tag{18}$$

and due to the orthonormality of $\mathbf{a}_s$ (see equation (12))

$$d_{jj'}^2 = \sum_s \left(\lambda_s b_{js} - \lambda_s b_{j's} \right)^2 = \sum_s \left(z_{js} - z_{j's} \right)^2 \tag{19}$$

This expression is the squared Euclidean distance between the columns j and j' in the column space. Thus the differences between the difference profiles are represented in the full dimensional column space as distances between the points which represent these columns. The squared distance of a column j to the origin of the column space follows from

$$d_{j,0}^2 = \sum_i (p_{ij}/p_{.j} - p_{i.})^2 = \sum_s \left(\lambda_s b_{js} \right)^2 = \sum_s z_{js}^2 \,. \tag{20}$$

In a lower dimensional representation the distances between the points are smaller than the real distances, and they are only approximations to the true distances in the full dimensional space (see also Greenacre 1993, p. 42, 43)

Even though one could derive expressions for the distance between two rows i and i' based upon differences in the Euclidean space of the rows, they are not very relevant in the present context, because in the biplot used to display the results of a nonsymmetric correspondence analysis, the Euclidean space of the rows is not represented (see also Greenacre 1993, Module 9).

Measures of fit

In the section Predictability, we discussed the partitioning of τ into the contributions to τ of each column, τ_j. Similarly, one can partition only the numerator of τ, which leads to the same proportional contributions of the columns.

$$N_\tau = \sum_i \sum_j p_{.j} (\Pi_{ij})^2 = \sum_j \left[p_{.j} \sum_i \Pi_{ij}^2 \right] = \sum_j N_{\tau_j}. \tag{21}$$

We can also derive the contribution of each column to the s^{th}-coordinate axis, using equation (11).

$$
\begin{aligned}
N_\tau &= \sum_i \sum_j p_{.j}(\Pi_{ij})^2 = \sum_i \sum_j p_{.j}(\sum_s \lambda_s a_{is} b_{js})^2 \\
&= \sum_i \sum_j p_{.j} \sum_s \sum_{s'} \lambda_s \lambda_{s'} a_{is} a_{is'} b_{js} b_{js'} \,.
\end{aligned} \tag{22}
$$

Using first the orthonormality of $\mathbf{a}_s$ (see equation (12)), and then the weighted orthonormality of $\mathbf{b}_s$ (see equation (13)) we get

$$
N_\tau = \sum_j p_{.j} \sum_s (\lambda_s b_{js})^2 = \sum_j \sum_s p_{.j} z_{js}^2 = \sum_s \left(\sum_j p_{.j} z_{js}^2 \right) = \sum_s \lambda_s^2. \tag{23}
$$

From equation (23) we see that

$$
(p_{.j} z_{js}^2)/\lambda_s^2 \tag{24}
$$

is the proportional contribution of the j^{th}-column to the s^{th} coordinate axis, or more general to the increase in predictability. One can also derive from equations (21) and (22) how much the axes in an approximation contribute to the N_{τ_j} for each column

$$
(\sum_s p_{.j} z_{js}^2)/N_{\tau_j}. \tag{25}
$$

If the contribution is close to one, no further axes are necessary to represent the profile of column j. Columns for whom the contributions are close to zero, are not well represented by the axes.

Similar expressions could be derived for the rows, but for the rows the contribution of the axes to the increase in predictability is a much more interesting quantity. This can also be derived from equation (22), but by starting with the weighted orthonormality of the $\mathbf{b}_s$.

$$
N_\tau = \sum_i \sum_s (\lambda_s a_{is})^2 = \sum_i \sum_s (\lambda_s y_{is})^2 = \sum_i \left[\sum_s (\lambda_s y_{is})^2 \right]. \tag{26}
$$

From equation (26) we see that the proportional contribution of the s^{th}-coordinate axis to the increase in predictability of a row i is equal to

$$
\lambda_s y_{is}^2 / \sum_s \lambda_s y_{is}^2. \tag{27}
$$

One could sum these quantities over the axes actually present in the approximation to establish how well the overall increase in predictability for a row, i.e. $\sum_s (\lambda_s y_{is})^2$, is approximated by the reduced number of axes.

Categorical multiple regression

Most really interesting theoretical prediction questions involve not just a single but several independent variables. When all variables are continuous such problems are typically tackled with multiple regression and the purpose of this section is to show that nonsymmetric correspondence analysis has a similar extension to more categorical variables. We will demonstrate this for two predictors.

The parallelism between ordinary and categorical multiple regression, follows most easily from the formulas for building the R^2 in a hierarchical regression. Suppose that I is the dependent variable and J and K the predictors, then the hierarchical formula is

$$r^2_{Mul} = r^2_{I|J\&K} = r^2_{IJ} + r^2_{IK|J}(1 - r^2_{IJ}) \tag{28}$$

where $r^2_{IK|J}$ is partial correlation between I and K given J. This partial correlation indicates which proportion of the variance of I which has not yet been explained by J is explained by K. Thus the second term on the right of equation (28) expresses the increase in predictability by knowing K given that one already knows how much J explains.

As indicated by D'Ambra and Lauro (1989, 1992; see also Gray & Williams, 1975, much later published as Gray & Williams, 1981) the parallel expression can be developed for τ:

$$\tau_{Mul} = \tau_{I|J\&K} = \tau_{IJ} + \tau_{IK|J}(1 - \tau_{IJ}) \tag{29}$$

with the same type of interpretation. It is instructive to look at the numerators for the three quantities in equation(29). For the prediction of I by J we get the equation

$$N_{\tau_{IJ}} = \sum_j \left[\sum_i p_{.j.} \left(\frac{p_{ij.}}{p_{.j.}} - p_{i..} \right)^2 \right] . \tag{30}$$

Thus the τ_{IJ} is based on the $I \times J$ margin of the $I \times JK$ table, and therefore we might call τ_{IJ} the marginal τ. The partial τ to determine the increase in predictability if we add another predictor K was given by Gray and Williams (1975; 1981, see also D'Ambra and Lauro, 1992) has as its numerator

$$N_{\tau_{IK|J}} = \sum_j \left[\sum_i \sum_k p_{.jk} \left(\frac{p_{ijk}}{p_{.jk}} - \frac{p_{ij.}}{p_{.j.}} \right)^2 \right] , \tag{31}$$

thus in the partial τ each column jk of the $I \times JK$ table is compared to the centroid of stratum j, or the margin of the conditional marginal distribution given j. Finally the multiple τ has as its numerator

$$N_{\tau_{I|J\&K}} = \sum_i \sum_j \sum_k p_{.jk} \left(\frac{p_{ijk}}{p_{.jk}} - p_{i..} \right)^2 , \tag{32}$$

so that each column is compared with the overall margin independent of the structure defined on the columns. In other words, the multiple τ is exactly the same as the τ for the unstructured table, but of course only knowing the multiple τ does not help in assessing the relative importance of the two predictors.

Extensions

The basic techniques presented here have been extended in many ways by Lauro, D'Ambra and their coworkers by considering other variants such as multiple NSCA (Lauro and D'Ambra, 1984; Siciliano, Lauro, Mooijaart, 1990), NSCA for three-way tables (D'Ambra and Lauro 1989; Lombardo, Carlier, & D'Ambra, 1996), and partial NSCA (D'Ambra and Lauro, 1992). Statistical aspects such as asymptotic properties have been studied by Siciliano (1990). Relationships with loglinear and other methods for categorical data have been investigated by D'Ambra and Lauro (1992) and Lauro and Siciliano (1989). For a recent review see Balbi (in press).

Software

Unfortunately, nonsymmetric correspondence analysis is not yet included in any major software package. However, the basic algorithm is fairly similar to that of correspondence analysis, and not too difficult to implement in a matrix-based language like Matlab, and for such a program the S-plus (Becker, Chambers, and Wilks 1988) source available from the second author may serve as a model. Alternatively, one may download an executable FORTRAN program ASYMTAB available at the Website of The Three-Mode Company, *http://www.fsw.leidenuniv.nl/~kroonenb/genprogs/programs.htm* [1], from which a postscript version of this paper can be downloaded, as well. In the original papers by Lauro and D'Ambra (1984) and D'Ambra and Lauro (1989) detailed computational information can be found.

Application: Mother-Infant Interactions over Time

In this section data collected by Van den Boom (1988) will be analysed to illustrate some of the basic properties of nonsymmetric correspondence analysis. An earlier analysis of these same data was presented in Carlier and Kroonenberg (1996) illustrating three-mode correspondence analysis based on the Lancaster additive decomposition of interaction without using a dependence structure. The example is a three-mode one with one response and two predictor variables, illustrating the multiple regression variant of nonsymmetric correspondence analysis. A much simpler two-mode example is treated in the Appendix in much more numerical detail.

[1] This program also can homogenise the margins of a contingency table (see e.g. Fienberg, 1971) and computes the symmetric and skew-symmetric part if the table is square (e.g. Constantine & Gower, 1978)

Data

In her study of (Dutch) irritable infants, Van den Boom (1988; Van den Boom and Hoeksma, 1994) collected data of 30 infant-mother pairs during the first six months of life (for a discussion of irritability, see Van den Boom, 1988, p. 70ff.). Each month, each mother-infant pair was observed at home in two sessions of forty minutes which were video-taped. The video tapes were coded by trained observers, and each six seconds the most salient behaviour of both the infant and the mother was coded, for instance, infant *cried* and mother *soothed*. The original 14 categories for infant behaviour were reduced for this analysis to 7 categories and those of the mother to 6 categories. For each month and each mother-infant pair a 7 by 6 co-occurrence matrix was constructed from the categorical longitudinal sequences. Subsequently, the matrices were aggregated over mother-infant pairs, so that statements could be made about mother-infant interaction irrespective of the individual pairs. In order to avoid confusion with statements about the mother, the infant will always be referred to as 'he'.

The seven infant categories were crying, exploring, sucking, smiling and similar positive social behaviour, inactivity, i.e the infant does not do anything in particular, looking at the mother, and vocalising. The six mother behaviours were soothing, looking at the infant, stimulating, offering, contact seeking or maintaining with the infant, and other, i.e. behaviour not directed at the infant.

Thus the data set under consideration forms a 6 (mother behaviours) by 7 (infant behaviours) by 6 (months) three-way contingency table, which is reshaped for the present analysis as a 6 by 7×6 two-way table. The underlying structure for this table is such that there is one response variable (Mother behaviour) and two predictor variables (Infant Behaviour and Time).

Basic results

The multiple τ for predicting the mother behaviour from the monthly infant behaviours is .062, but as argued above the overall size of τ is difficult to interpret. Its significant asymptotic chi square is 44344 with 205 degrees of freedom. The (again significant) marginal τs are .055 and .008 for infant behaviour and time as predictors, respectively, while the parallel partial τs are also .055 and .008, respectively. The equality indicates that the predictors have virtually independent contributions towards the predictability of the mother behaviour, and that one could examine the two marginal tables separately. However, in this particular case we will not do this, and prefer to look directly at the complete picture. Looking at the partial τs, it can be seen that the τ for predicting mother behaviour from infant behaviour is much larger than that for predicting the mother behaviour from the age of the infant. This indicates that the stimulus-response situation is fairly clear cut, in the sense that certain infant behaviour calls forth particular responses from the mother, and that there are general changes in the mother behaviour (certain behaviours become relatively more frequent and other less frequent) but that these general time trends are not as strong.

The infant categories (i.e. columns) which contribute most towards the increase in predictability of the mother behaviour are crying of the infant in each of the six months (8%, 11%, 6%, 7%, 5%, 5%, respectively), in other words if we know that the infant is crying in the first month of life, we can predict the mother behaviour 8% better than we could from the marginal distribution alone. Thus when an infant cries he calls forth a maternal response pattern deviating from the average one. Other types of infant behaviour which show increases in predictability are smiling in most months, exploring in the last two months, and sucking in the last month.

The overall or marginal distribution of mother behaviour is as follows. During the time the mother is with her infant she spends 37% looking at her infant, 20% offering objects, 18% doing not-infant-related things, 11% contact maintaining or seeking, 9% stimulating the infant, and 5% soothing. It is with this distribution that the conditional distributions (given an infant behaviour) are compared.

Nonsymmetric correspondence analysis – NSCA

Given .hat there are 6 rows and 42 columns, there are maximum of 5 dimensions in the NSCA. The first three dimensions explain 47%, 26%, and 21% of the within-column variability (see equation (3)), leaving only 6% for the remaining two dimensions. Thus we can describe what is going on between infant and mother with three dimensions to a good degree of approximation. The nature of the interaction could be presented with tables of the coordinates but it is far more insightful to use a (three-dimensional) biplot for this. The first graph (Fig. 1) shows the first dimension against the third, and the second graph (Fig. 2) shows the second dimension against the third. These particular graphs were chosen because they make the inspection and interpretation the easiest.

The main conclusion from Figure 1 (concentrating on the first dimension) is that if an infant cries, whether he is one month or six months old, the mother soothes. To be more precise, over and above a general low tendency to sooth as is evident from the marginal distribution, the mother soothes intensively when her infant cries, as is to be expected. Other mother behaviours are less likely than average as is shown by their being on the opposite side of the first dimension compared to crying. With respect to the time trend, there seems to be a tendency to sooth less in the last two months, but the effect is fairly small certainly looking at the month-by-month variability. The third dimension is will be discussed together with the second.

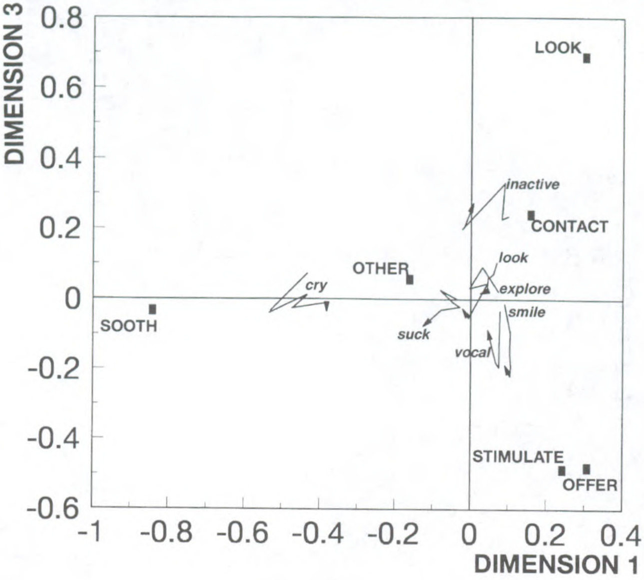


Figure 1 Van den Boom data: Nonsymmetric Correspondence Analysis Biplot (Dimension 1 versus Dimension 3).
Legend: Infant behaviours (*italics*) in standard coordinates; Mother behaviours (CAPITALS) in principal coordinates. The trajectories start at month 1 and end at month 6.

Figure 2 is far more complex. First of all, it illustrates the increases in predictability for the non-soothing behaviours given the infant is doing something else then crying. The general patterns are that if the infant is inactive the mother tends to look more and seek more contact than average. When the infant smiles or vocalises the mother stimulates more and offers more objects, while when the infant looks or explores the mother reacts more or less in an average way, i.e. primarily looking, offering and doing non-infant-related things. This figure also shows what the general time trend is. As the infant gets older the mother tends to do more non-infant-related things, nearly independent of what the infant does. This follows from two patterns: (1) all time arrows for infant behaviours end up pointing more towards Other mother behaviour and (2) they all point away from the mother behaviours they are most associated with. One may interpret this that as the infant gets older the mother leaves the infant more and more to his own devices even if she is in his neighbourhood.

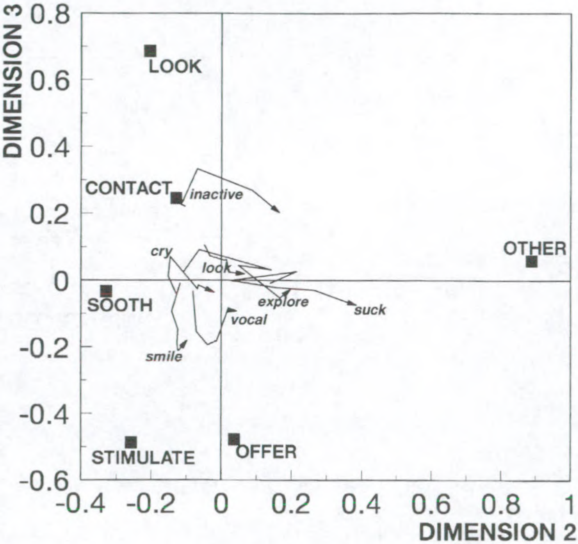


Figure 2 Van den Boom data: Nonsymmetric Correspondence Analysis Biplot (Dimension 2 versus Dimension 3).
Legend: Infant behaviours (*italics*) in standard coordinates; Mother behaviours (CAPITALS) in principal coordinates. The trajectories start at month 1 and end at month 6.

Discussion and Conclusions

In this paper we have presented the basic theory of nonsymmetric correspondence analysis (NSCA) using relatively simple mathematics. The technique can fulfill several functions depending on the way one desires to present the information on the difference profiles, i.e. the differences between the conditional and marginal distributions. Even in those cases where one does not want to make use of the full graphical capabilities of the technique via biplots, nonsymmetric correspondence analysis can still serve to find meaningful patterns in prediction. One may also use a feature not shown in this paper, i.e. arranging the columns and rows of the table with difference profiles into an optimal order using the coordinates from the singular value decomposition. In which way in any particular situation nonsymmetric correspondence analysis can best be used depends on the detail with which one wants to inspect the data. For concentrating on the larger scale patterns the biplots seem ideal, however if one wants to inspect each and every detail the full difference profiles are obviously better, but even in that case the analysis can help to organise the inspection via rearranging the table.

One of the attractive aspects of NSCA is that its results can be presented in a format which can be easily understood by even statistically unsophisticated persons, both via rearranged tables and graphical displays. For the biplots, probably all that needs to be explained is the concept of a projection (dropping a perpendicular line from a point onto a line), and the idea that the increase in predictability for rows in a particular column can be read from the relative positions of their projections on the column vectors. The most difficult aspect of interpreting the biplots is that a row point located relatively far away from a column vector with a large projection on the column vector, still has a strong increase in predictability for that column. Another problem may occur when a two-dimensional representation does not account for a large part of the deviations from the overall or marginal distribution. Three-dimensional graphs can be constructed, but they are always more difficult to read than a two-dimensional one, unless the points nicely align along the axes (as was the case for the Van den Boom data). The number of categories in a table can be fairly large without making the biplot unreadable, especially when there are clear-cut patterns. However, there is clearly a limitation to the number of categories that can be displayed simultaneously. When one wants to look at situations with more than two predictors each of which has a sizeable number of categories the displays may become rather cluttered and one might have to take special measures to maintain readability of the graph.

Nonsymmetric correspondence analysis is different from regular correspondence analysis and the two techniques can lead to different results. This in itself is not surprising as the data design in both cases is different. The dependence structure is directly modelled in NSCA as it is a categorical equivalent of regression analysis for continuous data. On the other hand, regular correspondence analysis is concerned with the interdependence of row and column variables. Notwithstanding this fundamental difference in some cases due to the specific data structure, the displays of regular correspondence analysis and those of nonsymmetric correspondence analysis may be comparable, but this still does not make the techniques interchangeable. One (side) effect of the different deviations analysed, is that NSCA does not suffer from the undue influence of rows with small marginal row totals, as correspondence analysis does.

Correspondence analysis has been applied to arbitrary rectangular data matrices with non-negative entries. Technically, NSCA could be used in a similar fashion, but its interpretation leans heavily on the possibility of interpreting profiles of proportions and distributions of differences between proportions. It depends entirely on the kind of (nonnegative) numbers at hand whether a sensible interpretation can be found if one does not start with frequencies or proportions.

APPENDIX
Worked mini-example: Transfer of Attachment from Mother to Infant

A.1 Data and Proportions

The data are derived from Van IJzendoorn (1995) and refer to the relationship between the attachment classification of the infant as assessed with the Strange Situation and that of the mother as assessed with the Adult Attachment Interview. The substantive details can be found in the original publication and its references. No really substantive interpretations will be given in this Appendix. It is intended to show the numeric information supplied by a nonsymmetric correspondence analysis, and its statistical interpretation.

Table A.1 Input frequency table with marginal proportions

Classif. Infant	Attachment Classification Mother 1 Ds	2 F	3 E	4 U	Row Margin	p(i.)
1 A	62	29	14	11	116	.212
2 B	24	210	14	39	287	.524
3 C	3	9	10	6	28	.051
4 D	19	26	10	62	117	.214
ColMargin	108	274	48	118	548	
p(.j)	.197	.500	.088	.215		1.000

The row marginal proportions indicate the relative frequency with which infants are classified into one of the four attachment categories and together they form the overall distribution with which the conditional distributions (i.e. given one knows the mother category) are compared. The column marginal proportions indicate the relative frequency with which mothers are classified in each of the categories. The near equality of the row and column margins is of considerable theoretical interest and is not an artifact of the data collection.

Table A.2 Column profiles $p_{ij}/p_{.j}$

Classif. Infant	Attachment Classification Mother 1 Ds	2 F	3 E	4 U	p(i.)
1 A	.574	.106	.292	.093	.212
2 B	.222	.766	.292	.331	.524
3 C	.028	.033	.208	.051	.051
4 D	.176	.095	.208	.525	.214
ColMargin	108	274	48	118	1.000

The *column profiles* are the conditional distributions of the infant categories given the classification of the mother. Comparing the conditional distributions with the marginal one shows that there is considerable predictive value in knowing the mother category, as all distributions deviate strongly from the marginal distribution.

Table A.3 Centred column profiles $(p_{ij}/p_{.j}) - p_{i.}$

Classif. Infant	Attachment Classification Mother 1 Ds	2 F	3 E	4 U	p(i.)
1 A	.362	-.106	.080	-.118	.212
2 B	-.302	.243	-.232	-.193	.524
3 C	-.023	-.018	.157	.000	.051
4 D	-.038	-.119	-.005	.312	.214
ColMargin	108	274	48	118	1.000

The *centred column profiles* are the differences between the column profiles and the marginal profile, and they are at the heart of nonsymmetric correspondence analysis. They indicate the extent to which cells in a profile have a higher (lower) proportion than the marginal proportion, and as such they indicate the increase (decrease) in the gain in predictability given the column category. Thus once we know a mother has a Ds classification this increasing the probability that she has an A infant with .362, and it decreases the probability of a B infant with -.302. The proportions of C and D infants are the same as those in the marginal distribution, i.e. equal to the case where we did not know the mother classification. This table clearly shows that for each mother category the probability of one particular infant category is greatly enhanced (Ds $\rightarrow$ A; F $\rightarrow$ B; (E $\rightarrow$ C); U $\rightarrow$ D) be it that E-C link is not very strong. In fact, there is more clearly an indication of a decrease in predictability: if the mother has an E classification then a B infant is less likely than one would have guessed from the marginal distribution.

A.2 Analysis of τ

```
tau (Rows dependent) = .199.
Asymptotic approximate chi square = 326.514 with df = 9
```

The τ value is not in itself very interpretable, as explained in the text, but the asymptotic chi square is very large compared to the number of degrees of freedom, and therefore clearly significant. In other words, there is 'permission' to interpret the results.

Table A.4 Proportional contribution to τ of each column

	Ds	F	E	U
1	.351	.336	.059	.254

The τ indicates the increase in predictability. Looking at the conditional distributions we see from the proportional contributions that Ds, F, and U have more or less similar contributions. However, knowing E does not increase the predictability much. There are two reasons for this: (1) the conditional distribution does not deviate as much from the marginal distributions as the other conditional distributions and (2) there are relatively few E classifications compared to the other classifications.

Table A.5 Proportional contribution to τ of each row

	A	B	C	D
1	.279	.478	.019	.225

These quantities can be used to assess to what extent it helps knowing the mother classification in improving the prediction of the infant classification. Knowing the mother classification clearly improves the prediction for the B category, but not much for the C category. Both the smaller number of C children and the small values in the centred profiles contribute the latter effect.

Table A.6 Analysis of Variance (Light & Margolin)

	MS(Total)	MS(Between)	MS(Within)
Row dependent	.3164	.0629	.2534

The partitioning of the total variance in a between and within part shows again a rather low value for the between variability, but here the same is true as with τ that it is difficult to make a statement about the absolute values. Note that $\tau = \text{MS(Between)}/\text{MS(Total)} = .0629/.3164 = .199$. Giving a clear interpretation of τ as a proportion explained variance, and stressing the analogue with analysis of variance.

A.3 Results from Non-symmetric correspondence analysis

As mentioned above the centred profiles are at the core of nonsymmetric correspondence analysis. The idea behind the technique is that we want to portray similarity between the centred profiles for column categories by normal (Euclidean) distances in a graph. Moreover, we would like to compare the centred profiles with the marginal profile, and we would like to assess which row categories have gained in predictability by knowing the column category. All this can be derived from a biplot of the rows and columns of the centred profiles. The required coordinates follow from a (generalised) singular value decomposition of centred profiles.

Table A.7 Singular values and Principal Inertias (Eigenvalues)

No.	Singular Value	Principal Inertia	Proportion	Cumulative Proportion
1	.29562	.08739	.694	.694
2	.18847	.03552	.282	.976
3	.05460	.00298	.024	1.000

```
Total inertia          .12589
            [= 2 * MS(Between) = Sum squared singular values]
```

As the minimum of the number of rows and the number of columns is 4, there are at most 3 dimensions or axes. The first two dimension take already 98% of the variability, and thus a two-dimensional graph shows virtually all there is to see. Later we will indicate what is contained in the remaining dimension. Note that the inertia is equal to the square of the singular value (e.g. $.29562^2 = .08739$)

Table A.8 Standard row coordinates

		1	2
1	A	-.464	.671
2	B	.826	.094
3	C	-.046	-.029
4	D	-.316	-.735

Standard coordinates (mathematically: left singular values) have lengths equal to 1, and thus squaring the entries in a column and adding them gives a value equal to 1. To make it a proper length the square root of this value should be taken, but that value is, of course, also 1. One may also calculate the standard coordinates for the columns (right singular vectors), but these are only used for calculating the principal coordinates.

Table A.9 Principal coordinates for columns

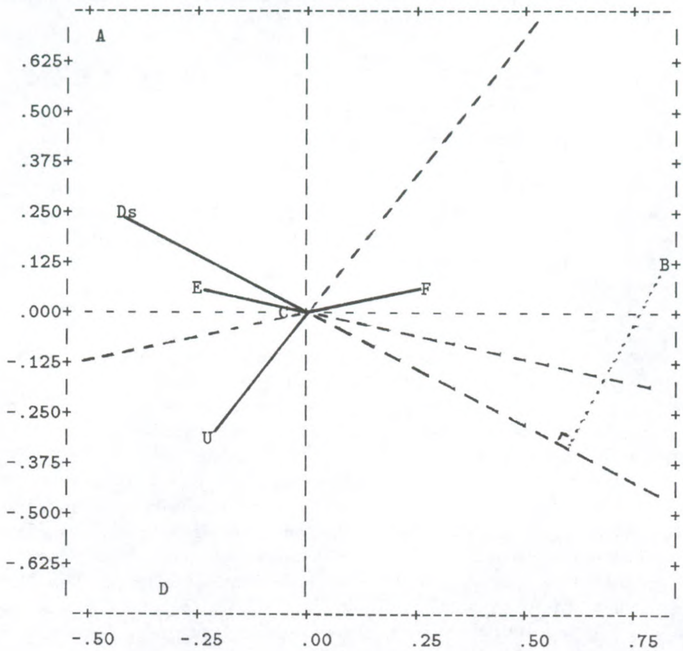
		1	2	Length	Marg.Prop.
1	Ds	-.404	.243	.472	.197
2	F	.288	.040	.291	.500
3	E	-.234	.031	.236	.088
4	U	-.203	-.327	.385	.215
Length		.087	.036		

Principal coordinates have lengths equal to the singular values. Thus squaring the entries (say, of the first column), weighting them with their marginal proportion and adding them gives the inertia ($.197 * -.404^2 + .500 * .288^2 + .088 * -.234^2 + .215 * -.203^2 = .087$). The lengths of the vectors for the mother classifications in the biplot are given under the column headings length, e.g. for the Ds category the length is computed as $\sqrt{(-.404^2 + .243^2)} = .472$.

A.4 Biplot

Figure A.1 Biplot of the Child and Mother Categories
(Child categories in standard coordinates; Mother categories in principal coordinates)

Dimension 1 (Horizontal) Versus Dimension 2 (Vertical)



As the biplot is based on two of the three dimension there is a bit of distortion but not very much as it contains 98% of the variability. The origin of the plot represent the marginal distribution and the (Euclidean) distance of column with respect to the origin indicates the extent to which a column deviates from the marginal distribution. similarly, distances between the column points indicate the extent to which their conditional distributions are similar.

The patterns we discerned in the table of the centred profiles are faithfully represented here in the plot. Thus the Ds classification of the mother increases the proportion of A infants in the conditional distribution, F does the same for B and U for D. The way to see this is by projecting the row point onto the vector connecting the origin with the column point. If the *inner product*, i.e.

product of the length of column vector times the length of the projection of the row vector is large and positive, there is a large increase in prediction for the row category; this is the case for all pairs mentioned above. If it is large and negative there is a large decrease in the prediction for that row category, e.g. for B if the mother has a Ds classification. The mother E and infant C classification is the smallest in accordance with what we have said above.

A.5 Supplementary information

Most analyses are not complete without an analysis of the residuals to assess how faithful the results reflect the original data and to assess whether there are any problems with specific aspects of the analysis.

Table A.10 Implied centred column profiles and residuals

	Implied Centred Profiles				Residuals			
	1 Ds	2 F	3 E	4 U	1 Ds	2 F	3 E	4 U
1 A	.351	-.107	.130	-.125	.011	.001	-.050	.007
2 B	-.311	.242	-.191	-.199	.009	.001	-.041	.006
3 C	.011	-.014	.010	.019	-.034	-.004	.147	-.019
4 D	-.051	-.120	.051	.305	.013	.001	-.056	.007

Using the two dimensions the centred profiles have been recalculated from the coordinates. One might call these the fitted or implied values based on the model used. When they are subtracted from the original centred profiles to form the residuals we can investigate the discrepancies between the two-dimensional solution and the full three-dimensional one.

In this case we see that the residuals are small as was to be expected given that the two dimensions took care of 98% of the variability. The largest discrepancies can be found for the distribution of mothers with an E classification. The positive value .147 indicates that the predictability of C is higher for these mothers than one could read from the two-dimensional graph. The third dimension will thus serve to lift the vector of E and the point C out of the plane. Apart from this relationship there are no further serious discrepancies to be seen.

Table A.11 Proportional contribution of columns to axes (absolute contributions sum to eigenvalues)

	1	2
1 Ds	.369	.328
2 F	.474	.022
3 E	.055	.002
4 U	.102	.648

The above proportional contributions show which columns, thus which mother classifications, contribute most to which axes. The Ds and F classification do so for the first and U and Ds for the second, while E does not contribute at all. In a two-dimensional graph this information can easily be seen, for higher dimensional graphs this can be more problematic.

Table A.12 Proportional contribution of the two axes to the column contributions to tau

	1	2	3	4
1	.993	1.000	.658	.997

Table A.4 showed how much each column contributed to the overall increase in predictability τ, indicated by τ_j. The numbers above indicate how well the two axes together succeeded in reproducing this contribution. Thus all contributions to the increase in predictability are accounted for by the two axes, except for the E mothers.

Table A.13 Proportional contribution of axes to the increase in predictability of the rows

	1	2	Total
1 A	.537	.455	.993
2 B	.992	.005	.997
3 C	.075	.012	.087
4 D	.309	.680	.988

The final table shows how well we are able to predict the increase of predictability for each row and which dimensions are responsible. The most obvious things to note is that the A, B and D categories are catered for, but the increase in predictability of the C category is not really included in the two-dimensional analysis.

A.6 Final remarks

The message from this small table is really too simple for all the calculations carried out here. The amount of newly calculated numbers far exceeds the original ones. However, the exposé above is primarily meant to illustrate the flow of the analysis and its assessment, entirely in accordance with the function of a tutorial.

Notwithstanding one can say that the proposed approach towards this small table gives insight in the relations in the table. However, for this example the centred profiles in principle already show all there is to tell.

References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Balbi, S. (1997). Graphical displays in Non-Symmetric Correspondence Analysis. In Michael Greenacre and Jörg Blasius (Eds.), *Visualization of categorical data* (pp.##). London: Academic Press.

Becker, R.A., Chambers, J.M., and Wilks, A.R. (1988). *The new S language.* Pacific Grove, CA: Wadsworth & Brooks/Cole.

Benzécri, J.P. (1973). *L'Analyse des données. Vol. II, L'Analyse des correspondances* [Data analysis. Vol. II, correspondence analysis]. Paris: Dunod.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete multivariate analysis.* Cambridge, MA: The MIT Press.

Carlier, A. and Kroonenberg, P.M. (1996). Decompositions and biplots in three-way correspondence analysis. *Psychometrika, 61*, 355-373.

Constantine, A.G. and Gower, J.C. (1978). Graphical representation of asymmetry. *Applied Statistics, 27*, 297-304.

D'Ambra, L. and Lauro, N.C. (1989). Non symmetrical analysis of three-way contingency tables. In R. Coppi and S. Bolasco (Eds.), *Multiway data analysis* (pp. 301-315). Amsterdam: Elsevier.

D'Ambra, L. and Lauro, N.C. (1992). Non symmetrical exploratory data analysis. *Statistica Applicata, 4*, 511-529.

Eckart, C. and Young, G. (1936), The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211-218.

Fienberg, S.E. (1970). A statistical technique for historians: Standardizing tables of counts. *Journal of Interdisciplinary History, 1*, 305-315.

Gabriel, K.R. (1971). The biplot graphic display with application to principal component analysis, *Biometrika, 58*, 453-467.

Gabriel, K.R. and Charles L. Odoroff (1990). Biplots in biomedical research. *Statistics in Medicine, 9*, 469-485.

Gini, C.W. (1912). Variabilità e mutabilità, contributo allo studio delle di distribuzioni e relazioni statistiche. *Studio-Economico-Giuridici della R. Università di Cagliari.*

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

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

Greenacre, M.J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Lauro, N. C. and D'Ambra, L. (1984). L'Analyse non symétrique des correspondances [Nonsymmetric correspondence analysis]. In E. Diday et al. (Eds.), *Data Analysis and Informatics III* (pp. 433-446). Amsterdam: Elsevier.

Lauro, N.C. and Siciliano, R. (1989), Exploratory methods and modelling for contingency tables analysis: An integrated approach. *Statistica Applicata, 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.

Lombardo, R., Carlier, A. & D'Ambra, L.(1996). Decomposition models for dependence analysis in three-way contingency tables. *Methodologica, 4*, 59-80.

Marcotorchino, F. (1984). *Utilisation des Comparaisons par Paires en Statistique des Contingences: Partie I, Partie II* [Use of Paired Comparisons in the Statistical Analysis of Contingency Tables], Unpublished Technical Reports No F 069, No. F 071, Étude du Centre Scientifique, IBM-France.

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.

Siciliano, R. (1990). Asymptotic distribution of eigenvalues and statistical tests in non symmetric correspondence analysis. *Statistica Applicata*, 2-3, 259-276.

Siciliano, R., Lauro, N.C. and Mooijaart, A. (1990). Exploratory approach and maximum likelihood estimation of models for non symmetrical analysis of two-way multiple contingency tables. In K. Momirovic and V. Mildner (Eds.), *COMPSTAT 1990* (pp. 157-162). Heidelberg: Physica Verlag.

Van den Boom, D.C. (1988). *Neonatal irritability and the development of attachment: Observation and intervention.* Unpublished doctoral thesis, Department of Psychology, Leiden University.

Van den Boom, D.C. and Hoeksma, J.B. (1994). The effect of infant irritability on mother-infant interaction: A growth-curve analysis. *Developmental Psychology, 30*, 581-590 (Correction: 1995, *31*, 197).

Van IJzendoorn, M.H. (1995). Adult attachment representations, parental responsiveness, and infant attachment. A meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychological Bulletin, 117*, 387-403.

ontvangen: 15-02-1998
geaccepteerd: 09-04-1998