GRAPHICAL REPRESENTATIONS AND ODDS RATIOS IN A DISTANCE-ASSOCIATION MODEL FOR THE ANALYSIS OF CROSS-CLASSIFIED DATA

MARK DE ROOIJ

LEIDEN UNIVERSITY

WILLEM J. HEISER

LEIDEN UNIVERSITY

Although RC(M)-association models have become a generally useful tool for the analysis of cross-classified data, the graphical representation resulting from such an analysis can at times be misleading. The relationships present between row category points and column category points cannot be interpreted by inter point distances but only through projection. In order to avoid incorrect interpretation by a distance rule, joint plots should be made that either represent the row categories or the column categories as vectors. In contrast, the present study proposes models in which the distances between row and column points can be interpreted directly, with a large (small) distance corresponding to a small (large) value for the association. The models provide expressions for the odds ratios in terms of distances, which is a feature that makes the proposed models attractive reparametrizations to the usual RC(M)-parametrization. Comparisons to existing data analysis techniques plus an overview of related models and their connections are also provided.

Key words: Euclidean distances, maximum likelihood estimation, log-linear modelling, multidimensional scaling, multidimensional unfolding.

1. Introduction

Over the last two decades, a wide variety of models for the analysis of contingency tables has been proposed, with particularly important developments in log-linear and RC(M)-association and canonical correlation models (e.g. Goodman, 1972, 1979, 1981, 1985, 1986; Andersen, 1980; Gilula & Haberman, 1986; Becker, 1990; Haberman, 1974, 1978, 1979, 1995). Restrictions on the association terms cause these latter models to give intermediaries between a model of complete independence and one that is saturated in the traditional log-linear analysis approach. The interpretation of these restricted models is often through graphical representations of the row and column sets.

Two types of graphical representations jointly representing both sets of objects can be distinguished: Type I plots where the relationship between the two sets is described by a distance rule, and Type II where this relationship is described by an inner product rule. For Type I plots the categories for both sets should be represented by points in Euclidean space, with the distance between the points describing the relationship between categories of the two sets. For Type II, at least one set should be drawn using vectors, and the points of the other set projected onto these

vectors to represent the relationship. Often, inner product models (Type II) are represented by Type I plots, which are then subject to incorrect interpretations (see, for example, Clogg, Eliason, & Wahl, 1990, who interpret an inner product representation with distances). The authors believe that any plot with categories of both sets represented as points in Euclidean space (i.e., a Type I plot representation) is intuitively interpreted by a distance rule rather than by projection.

The RC(M)-association model and the canonical correlation model produce graphical representations based on inner products, and thus result in Type II plots. In the present paper, we propose a new model in terms of distances that results in a Type I plot and a distance interpretation that is arguably more straightforward than the usual RC(M)-association model parametrization. The topic is related to the difference between Correspondence Analysis (CA; Greenacre, 1984; Nishisato, 1980; Gifi, 1990) and Multidimensional Scaling (MDS; cf. Borg & Groenen, 1997), and the Multidimensional Unfolding generalization of the latter to rectangular matrices (MDU; Heiser, 1981). The CA model uses an inner product rule for the relation between row and column categories, and therefore is Type II. In MDS and MDU, the relationship between row and column points is represented by distances and therefore one should use a Type I plot. In CA, only the distances within the set of row points, or within the set of column points can be interpreted directly if the proper normalization is chosen (see below). The relationship between the row points and the column points can only be assessed by projection (cf. Greenacre, 1984, p. 119). A modification of the CA model proposed by Carroll, Green, and Schaffer (1986, 1987, 1989) to obtain a between row and column set distance representation leads, as both the original authors and Greenacre (1989) note, to a curious definition of distances. Moreover, Greenacre (1989) gives convincing examples of the poor representation of those distances in the resulting graphical representation. In MDS and MDU, the distances between all points can be interpreted directly and where the inner product relationship in CA is replaced by a distance rule in MDU. For a detailed discussion on the relationships among MDS, MDU, and CA, the reader is referred to Heiser and Meulman (1983).

In the next section, we develop our models first for square contingency tables assuming a symmetric association pattern, and then for rectangular tables where the assumption of a symmetric association is dropped. The representation of odds ratios in terms of distances is discussed and a comparison to related models for contingency tables is presented. This section concludes with remarks on indeterminacies in the model plus a way to identify the models. A third section applies the models to some empirical data, followed by a section with some summarizing discussion and an overview of related models and their connections.

2. Distance-Association Models

2.1. Theory

We start with a multiplicative model for the expected frequencies of a $K \times K$ frequency table, with parameters for the margins of α_i and β_j, and parameters for the association of θ_{ij} $(i, j = 1, \ldots, K)$. The most general multiplicative form for a two-way cross-classification can be written as

$$\mu_{ij} = \mu \alpha_i \beta_j \theta_{ij}, \tag{1}$$

where μ_{ij} is the expected frequency of row i and column j, and μ is constant.

Goodman (1979) proposed decomposing the association term (θ_{ij}) with a bilinear model, but we would reparametrize the association term into one involving distances. The frequency for cell ij is inversely related to the distance between points in Euclidean space representing categories i and j, so the more often a combination occurs, the more similar the categories are, and the closer the corresponding categories are in psychological space. To specify the relationships between the

frequencies in the table and the distances, a monotone decreasing function, ψ, mapping distances to the association parameters is used:

$$\theta_{ij} = \psi(d_{ij}), \tag{2}$$

where d_{ij} is a distance satisfying the metric axioms: symmetry ($d_{ij} = d_{ji}$, for all i and j), minimality ($d_{ij} \geq d_{ii} = 0$, for all i and j), and the triangle inequality ($d_{ik} + d_{jk} \geq d_{ij}$, for all i, j, and k). In the field of psychophysics, such a transformation was earlier proposed by Shepard (1957; also see Heiser, 1988; Nosofsky, 1985; Takane & Shibayama, 1986).

Two decisions are required: choice of distance and the form of the transformation function ψ. Distances in psychological space are often assumed to be some Minkowski r-metric,

$$d_{ij}(\mathbf{X}) = \left[\sum_{m=1}^{M} \left| x_{im} - x_{jm} \right|^r \right]^{1/r}, \quad r \geq 1, \tag{3}$$

where x_{im} is the coordinate of category i on dimension m ($m = 1, \ldots, M$), collected into the $K \times M$ matrix, $\mathbf{X}$. For $r = 1$, the distance is city-block, and when $r = 2$, the distance is Euclidean. In psychophysics the two transformations that are generally used are the exponential decay function

$$\theta_{ij} = \exp(-d_{ij}(\mathbf{X})), \tag{4}$$

and the Gaussian function

$$\theta_{ij} = \exp(-d_{ij}^2(\mathbf{X})). \tag{5}$$

It can be shown that for exponential decay, the city-block distance is the canonical distance in that it results in additivity over dimensions; for the Gaussian transformation it is the Euclidean. As the Euclidean distance is the more common, we will use it along with a Gaussian transformation function.

2.2. The One-Mode Distance-Association model

Summarizing, our *one-mode distance-association model* for a two-way square table has the form:

$$\mu_{ij} = \mu \alpha_i \beta_j \exp(-d_{ij}^2(\mathbf{X})). \tag{6}$$

After taking the logarithm, (6) can be rewritten as

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(\mathbf{X}), \tag{7}$$

where $\lambda = \log \mu$, $\lambda_i^R = \log \alpha_i$, and $\lambda_j^C = \log \beta_j$. Expanding the distance function

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - \sum_m (x_{im} - x_{jm})^2,$$

$$= \lambda + \lambda_i^R + \lambda_j^C - \mathrm{tr}\mathbf{X}'\mathbf{A}_{ij}\mathbf{X}, \tag{8}$$

where tr denotes the trace of a matrix, $\mathbf{A}_{ij}$ is defined as $\mathbf{A}_{ij} = (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)'$, and $\mathbf{e}_i$ is the i-th column of the identity matrix of order K. In psychophysics, a conditional version of this model was proposed by Nosofsky (1985), written in terms of conditional probabilities as

$$\pi_{j|i} = \frac{\beta_j \exp(-d_{ij}^2(\mathbf{X}))}{\sum_k \beta_k \exp(-d_{ik}^2(\mathbf{X}))}. \tag{9}$$

For a more general discussion on these kind of models for stimulus recognition data we refer to Takane and Shibayama (1992). If $\mu\alpha_i = 1/\sum_k \beta_k \exp(-d_{ik}^2(\mathbf{X}))$ in (9), the one-mode distance-association model is generated. Once the model is specified, a likelihood function can be optimized under an independent Poisson, Multinomial, or Product Multinomial sampling distribution—details are given in the appendices.

2.3. The Two-Mode Distance-Association model

The one-mode distance-association model can be applied only to square tables. For rectangular $I \times J$ tables, two sets of coordinates are needed, one for the row and one for the column points. The general idea followed comes from the theory of multidimensional unfolding, where given a rectangular table of proximities between two sets of objects, the objective of MDU is to find distances in Euclidean space that approximate the proximities as closely as possible (Coombs, 1964). For finding a least squares solution to an unfolding problem, a multidimensional scaling model can be used in which the input matrix has a special partitioned structure (Heiser, 1981, 1987).

In the one-mode model a single coordinate matrix $\mathbf{X}$ was defined for both ways of the frequency table. However, when the variables differ for the two ways, or when we assume an asymmetric association pattern two coordinate matrices should be used, one for the first way ($\mathbf{X}$), and one for the second way ($\mathbf{Y}$). The distance is then defined by

$$d_{ij}(\mathbf{X}; \mathbf{Y}) = \left[\sum_{m=1}^{M} (x_{im} - y_{jm})^2 \right]^{1/2}. \tag{10}$$

Distances between points of one set are not related to observations, but they can be interpreted like the distances between scores of the RC(M)-association models (see below). The model can be written in log-linear terms as follows

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(\mathbf{X}; \mathbf{Y}),$$
$$= \lambda + \lambda_i^R + \lambda_j^C - \sum_m (x_{im} - y_{jm})^2. \tag{11}$$

We will call this the *two-mode distance-association model.* Takane (1987) proposed the following conditional probability model

$$\pi_{j|i} = \frac{\beta_j \exp(-d_{ij}^2(\mathbf{X}; \mathbf{Y}))}{\sum_k \beta_k \exp(-d_{ik}^2(\mathbf{X}; \mathbf{Y}))}, \tag{12}$$

called the ideal point discriminant analysis (IPDA). This model will be discussed in more detail in Section 2.5.2.

2.4. The Odds Ratio

As the odds ratio describes the association independently of the marginal proportions, it is an important tool in the analysis of cross-classified data. In this section, the odds ratio is discussed under the distance-association model and expressions are given for the odds ratio in terms of distances. Beginning with the two-mode distance-association model, the relationship $\mu_{ij} = N\pi_{ij}$, where N is the sample size and π_{ij} is the probability, the odds ratio for rows i and i' and columns j and j' can be written as

$$\frac{\pi_{ij} \times \pi_{i'j'}}{\pi_{ij'} \times \pi_{i'j}} = \exp\left(-d_{ij}^2 - d_{i'j'}^2 + d_{ij'}^2 + d_{i'j}^2\right). \tag{13}$$

The log of the odds ratio under the two-mode distance model is a difference between sums of squared distances. For a square contingency table, the log-odds of staying in the same category instead of making a transition to another is equal to the sum of squared distances between categories minus the sum of squared distances within categories. Inserting the definition of the Euclidean distance, (13) can be written as

$$\frac{\pi_{ij} \times \pi_{i'j'}}{\pi_{ij'} \times \pi_{i'j}} = \exp\left(2\sum_{m=1}^{M}(x_{im} - x_{i'm})(y_{jm} - y_{j'm})\right), \tag{14}$$

and is the form found by Goodman (1979). So, the same odds ratio interpretation of the RC(M)-association model is kept, which for $M = 1$ was called "attractive" by Gilula & Haberman (1986, p. 782) because the odds ratio in (14) is a product of "distances". Note that in (13), dimensionality does not play a role and the log-odds ratio can be expressed in terms of genuine distances regardless of the dimensionality, whereas in (14) the terms multiplied are signed differences.

For a square matrix and the one-mode distance-association model, where $x_i = y_i$, the odds ratio is even simpler. For the odds ratios along the diagonal,

$$\frac{\pi_{ii} \times \pi_{jj}}{\pi_{ij} \times \pi_{ji}} = \exp\left(2d_{ij}^2\right). \tag{15}$$

This expression shows the odds of staying or giving the same response (a diagonal cell) in one category versus making a change or giving another response (an off-diagonal cell). Under the one-mode model this odds ratio is always larger than one because of the non-negativity of distances, implying that given the marginal proportions, the probability of staying in a category or giving the same response is always larger than the probability of moving or giving another response. The probability of staying in a category for the one-mode distance-association model, corresponds to a zero distance and the largest probability. This appears to be a common property of square contingency matrices.

To further clarify the expressions for the log-odds ratio, suppose one variable is measured twice, and the two categories of the variable are denoted as a and b. The log-odds of staying in either of the categories versus making a transition is $d_{a_1b_2}^2 + d_{b_1a_2}^2 - d_{a_1a_2}^2 - d_{b_1b_2}^2$, where $d_{a_1b_2}^2$ denotes the squared distance between the point representing category a at the first time point, and the point representing category b at the second. So the log-odds of staying versus moving is equal to the sum of squared intercategory distances ($d_{a_1b_2}^2 + d_{b_1a_2}^2$) minus the sum of squared intracategory distances ($d_{a_1a_2}^2 + d_{b_1b_2}^2$). In accordance with our definition of the model, the larger the distances between categories a and b, the higher the chance of staying; the larger the distances between the categories of the same variable ($d_{a_1a_2}, d_{b_1b_2}$) at the first and second time point, the lower the chance. In the one-mode model, the intracategory distances are (by definition) zero, and the intercategory distances are equal for a to b or conversely. The log-odds in that case are simply $2d_{ab}^2$, or twice the squared distance between category a and b.

2.5. Comparison to Related Models for Contingency Tables

In this section we will discuss the relationships of our distance-association models with three existing models (a unidimensional distance model, IPDA, and the RC(M)-association model) from a theoretical perspective. For a theoretical and empirical comparison between the RC(M)-association model, IPDA, CA (both least squares and maximum likelihood), and the Latent Budget Model we refer to Van der Heijden, Mooijaart, and Takane (1994).

2.5.1. *Unidimensional Distance Models*

It is necessary at this point to develop a comparison with distance models for contingency tables proposed by Goodman (1972) and Haberman (1974, Chapter 6); for consistency the log-linear notation presented by Haberman will be used. The distance model for a square contingency table proposed by these authors is written as

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - |\eta_i - \eta_j| + \delta_{ij}\epsilon_i, \tag{16}$$

where δ_{ij} is the Kronecker delta, ϵ_i is a so-called inheritance term to make the diagonal fit the data, and η_i is a scale position for category i. In this model, the difference between the two scale positions is the distance between the two categories on a one-dimensional scale. It should be noted that in a single dimension all Minkowski distances are equal, so the city-block distance equals the Euclidean as well as higher-order Minkowski distances. To enhance comparison, (16) can be rewritten

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - d_{ij}(\mathbf{x}) + \delta_{ij}\epsilon_i, \tag{17}$$

where $\mathbf{x}$ is now a vector representing the unidimensional scale values. For the estimation of this model, the order of the scale values has to be specified in advance.

In contrast to the model of Goodman (1972) and Haberman (1974) (G–H model), the current one-mode distance-association model (OMDA model) can fit multidimensional solutions where the G–H model can fit only a single dimension; also, the G–H model employs the exponential decay function and the OMDA model employs the Gaussian transformation, which in the case of a multidimensional solution makes optimization much simpler; finally, the G–H model has parameters for the diagonal, and requires a pre-specification of category order while the OMDA model does not require such an ordering. When it is not possible to specify an order in advance, it is well known that finding the optimal unidimensional scale is a combinatorial problem (De Leeuw & Heiser, 1977; Defays, 1978; Hubert & Arabie, 1986). The multidimensional solution can be extremely helpful when variables with many categories or multivariate transition data are analyzed. Under these circumstances, it is very hard to fit the data well in only one pre-specified dimension; the models developed here can be very useful in finding structure.

The special treatment of the diagonal in square contingency tables proposed by Goodman and Haberman can also be incorporated into the distance association models, with K extra parameters and the expected frequencies for the diagonal cells fitted to their observed values. The overall fit measured by the traditional chi-square distributed statistics will in general improve a lot.

2.5.2. *Ideal Point Discriminant Analysis*

Takane (1987, 1998) proposed the following conditional probability model

$$\pi_{j|i} = \frac{\beta_j \exp(-d_{ij}^2(\mathbf{X};\mathbf{Y}))}{\sum_k \beta_k \exp(-d_{ik}^2(\mathbf{X};\mathbf{Y}))}, \tag{18}$$

and called it an ideal point discriminant analysis (IPDA). Takane does not estimate the $\mathbf{Y}$ matrix, but simply takes $\mathbf{Y}$ to be the weighted mean of the $\mathbf{X}$. This simplification does reduce the number of parameters, but is unnecessarily restrictive. The distance-association model will be expected to find a somewhat better fit, although the differences might sometimes be negligible. Using the barycentric location of $\mathbf{Y}$, distances within the rows will be better approximated than distances within the columns (Greenacre, 1989), consistent with the conditional approach taken. The distance formulation in IPDA is especially suited for the situation where we have I multinomial

samples of a variable with J categories, with fixed row margins, and not so much for the situation of cross-classified data.

Takane (1998) discussed the interpretation of the graphical display, especially the interpretation of the between row and column distances, and concludes they are "rather intricate" and that "care should be exercised when they are interpreted in probabilistic terms". To summarize his findings: (1) $\pi_{i|j}$ is inversely monotonic with d_{ij} within column j, so that $d_{ij} > d_{i'j} \Leftrightarrow \pi_{i|j} < \pi_{i'|j}$; (2) π_{ij} is not necessarily inversely monotonic with d_{ij} unless β_j is constant across j; (3) $\pi_{j|i}$ is inversely monotonic with d_{ij} within i for different columns (j) only if β_j is constant across j (Takane, 1998, p. 448).

Two facets are generally of importance in the analysis of contingency tables: the odds ratio and the joint probability. The representation of the odds ratio with the IPDA model is the same as with our model, and in the distance association model, the distances are inversely monotonic related to $\mu_{ij}/\mu\alpha_i\beta_j$. It is the ratio of the frequencies and the expected frequencies under the independence model that are inversely monotonically related to the distances between points in Euclidean space. The distance is not necessarily monotonically related to the joint probability unless α_i and β_j are constant across i and j. However, as is shown by De Rooij (2001) and De Rooij and Heiser (2002) the main effects can be included as unique dimensions in the distance representation. The distance in the resulting graphical representation is then directly related to the estimated joint probability. Therefore we define a vector $\mathbf{r} = [r_1, r_2, \ldots, r_I, \mathbf{0}'_J]'$ where $r_i = \sqrt{|\lambda_i^R - \max_i(\lambda_i^R)|}$, and a vector $\mathbf{c} = [\mathbf{0}'_I, c_1, c_2, \ldots, c_J]'$ where $c_j = \sqrt{|\lambda_j^C - \max_j(\lambda_j^C)|}$. Then (11) can be rewritten as

$$\log(\mu_{ij}) = \lambda^* - (r_i - r_j)^2 - (c_i - c_j)^2 - \sum_m (x_{im} - y_{jm})^2$$

$$= \lambda^* - d_{ij}^2(\mathbf{X}; \mathbf{Y}; \mathbf{r}; \mathbf{c}), \tag{19}$$

where $\lambda^* = \lambda + \max_i(\lambda_i^R) + \max_j(\lambda_j^C)$ relates to the maximum expected frequency in the model. Thus, the extended distances are monotonically related to the joint probabilities, and the complete graphical representation is extremely easy to interpret. A graphical representation of this model will be shown in the second example of Section 3.

2.5.3. The RC(M)-Association Model

The RC(M)-association model is often used to analyze cross-classified data, with results represented graphically by a Type II plot. The RC(M)-association model is defined as

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \sum_m \phi_m \vartheta_{im} \nu_{jm}. \tag{20}$$

Graphical displays are commonly used to interpret the RC(M)-association model, with two displays given for the row and column scores. As only indirect relations between the rows and indirect relations between the columns are displayed, there is no way to really interpret the association. Only joint graphical displays of the rows and the columns can show how any one category of the row variable is associated with some category of the column variable.

A number of joint graphical displays can be used that are all of Type II, but as Greenacre (1984, p. 65) noted:

> "There are advantages and disadvantages of the simultaneous display. Clearly an advantage is the very concise graphical display expressing a number of different features of the data in a single picture. The display of each set of points indicates the nature

*of similarities and dispersion within the set. Notice, however, that we should avoid the
danger of interpreting the distances between the points of different sets, since no such
distances have been explicitly defined."*

In general, all plots are mathematically correct with row coordinates $\vartheta_{im}^* = \phi_m^\tau \vartheta_{im}$ and
column coordinates $v_{jm}^* = \phi_m^\kappa v_{jm}$, where $\tau + \kappa = 1$. In practice, one of the following graphical
displays is often found (using the terminology of a major statistical package Meulman, Heiser, &
SPSS Inc., 1999):

Row principal normalization: the row categories are plotted as points with coordinates $\vartheta_{im}' = \phi_m \vartheta_{im}$, and the column categories as vectors with coordinates v_{jm}. In row principal normalization, the Euclidean distances between the row points approximate (possibly weighted) Euclidean
distances between the row entries of the contingency table that are logarithmically transformed
and corrected for the main effects. The column vectors have a direction and a length. The association with the row categories is reconstructed by projection, and the length indicates how well a
column fits the chosen dimensionality.

Column principal normalization: the row categories are plotted as vectors with coordinates
ϑ_{im} and the column categories as points with coordinates $v_{jm}' = \phi_m v_{jm}$. The interpretation of
this display is analogous to (1), with the role of rows and columns reversed.

Symmetric normalization: the row categories are plotted as vectors with coordinates $\vartheta_{im}'' = \phi_m^{1/2} \vartheta_{im}$, and the column categories as vectors with coordinates $v_{jm}'' = \phi_m^{1/2} v_{jm}$. This normalization spreads the intrinsic association terms symmetrically over the rows and columns. Note that
neither the distances between the row points or between the column points are approximations
to data-related distances. This plot can only be interpreted by projecting the row (column) points
onto the direction indicated by any column (row) point.

Principal normalization: the row categories are plotted as points with coordinates $\phi_m \vartheta_{im}$ and
the column categories as points with coordinates $\phi_m v_{jm}$. Here, the intrinsic association terms are
spread twice in the solution, once over the row scores and once over the column scores. This is
basically an incorrect graphical display since $\tau + \kappa \neq 1$. This method of normalization can only
be used for making separate plots of row categories and column categories, respectively.

Comparing the mathematical structure of the RC(M)-association model with that of the
two-mode distance-association model, they produce the same expected frequencies. The RC(M)-
association model is defined in (20). Denoting $u_{im} = \sqrt{(1/2)}\phi_m^\tau \vartheta_{im}$, and $v_{jm} = \sqrt{(1/2)}\phi_m^\kappa v_{jm}$,
with $\tau + \kappa = 1$, we obtain

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \sum_m 2u_{im}v_{jm}, \tag{21}$$

and defining $\lambda_i^{R*} = \lambda_i^R + \sum_m u_{im}^2$, and $\lambda_j^{C*} = \lambda_j^{C*} + \sum_m v_{jm}^2$, (21) becomes

$$\log(\mu_{ij}) = \lambda + \lambda_i^{R*} - \sum_m u_{im}^2 + \lambda_j^{C*} - \sum_m v_{jm}^2 + \sum_m 2u_{im}v_{jm}. \tag{22}$$

By reordering the expressions, we obtain

$$\log(\mu_{ij}) = \lambda + \lambda_i^{R*} + \lambda_j^{C*} - \sum_m (u_{im}^2 + v_{jm}^2 - 2u_{im}v_{jm})$$

$$= \lambda + \lambda_i^{R*} + \lambda_j^{C*} - \sum_m (u_{im} - v_{jm})^2, \tag{23}$$

which is the two-mode distance-association model in (11). Note that the row and column scores
only changed by a factor $\sqrt{1/2}$, so a distance interpretation of the RC(M)-association model is
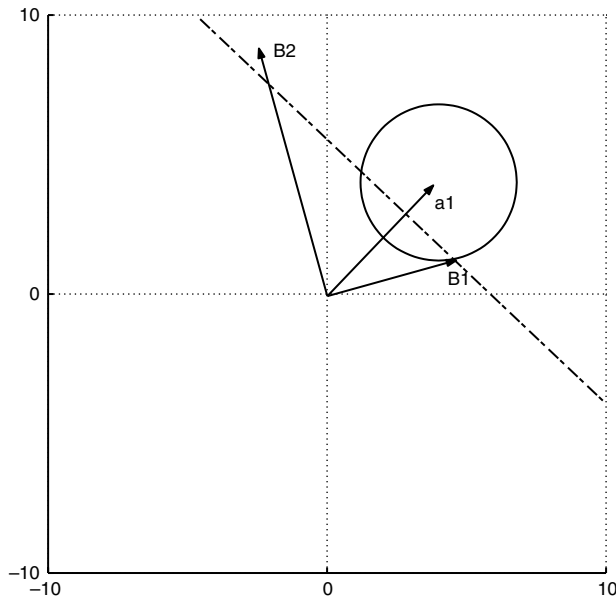
FIGURE 1.
Comparison of RC(M)-association model with distance-association model, 1.

provided. The most important changes are in the main effect parameters, which are often named unique components in distance models (Winsberg & Carroll, 1989).

To make a further comparison between the RC(M)-association model and the distance-association model in terms of the spatial representation and the interpretation, Figure 1 shows a plot in which one row category, $a1$, and two column categories, $B1$ and $B2$ are shown. Suppose the plot is a Type II plot (RC(M)-association model), where the symmetric normalization is chosen for the joint plot. The association of $a1$ with $B1$ is given by the length of vector $B1$ times the length of vector $a1$, times the cosine of the angle between the two vectors. All points on the dotted line have the same value for the association as between $a1$ and $B1$. All points on the lower left side of this line have a smaller association, and all points on the higher right side have a larger value. Point $B2$ is on the higher right side of the dotted line, so the association of $a1$ with $B2$ is larger.

Suppose the plot is a Type I (distance-association model), and all points on the circle have the same value for the association with $a1$ as $B1$ has. Points outside the circle have a smaller value for the association; within the circle they have a higher value than $a1$ has with $B1$. In the distance-association model the association between $a1$ and $B2$ is smaller than between $a1$ and $B1$. The conclusions derived from both spatial models are contradictory; those from the distance-association model are intuitively clearer.

A second comparison given in Figure 2 has two row points $a1$ and $a2$ and one column point $B1$. In a Type II plot (RC(M)-association model), the projection of $B1$ on the vector of $a1$ or $a2$ is the same, and the association between the row and column points is now determined by the length of the vector. As the length of vector $a2$ is larger than the length of vector $a1$, the association of $B1$ with $a2$ is larger. In a Type I interpretation, the distance from $a1$ to $B1$ is smaller indicating a larger association. Whether a plot is of Type I or II makes a significant difference, and therefore it is important to clearly distinguish between the two. A Type I plot can and should be made using only points in multidimensional space, whereas a Type II plot should be drawn making use of vectors for at least one set.

We showed that RC(M)-association models also have a distance interpretation. The graphical display of the RC(M)-association model can be interpreted both by a inner product rule and by
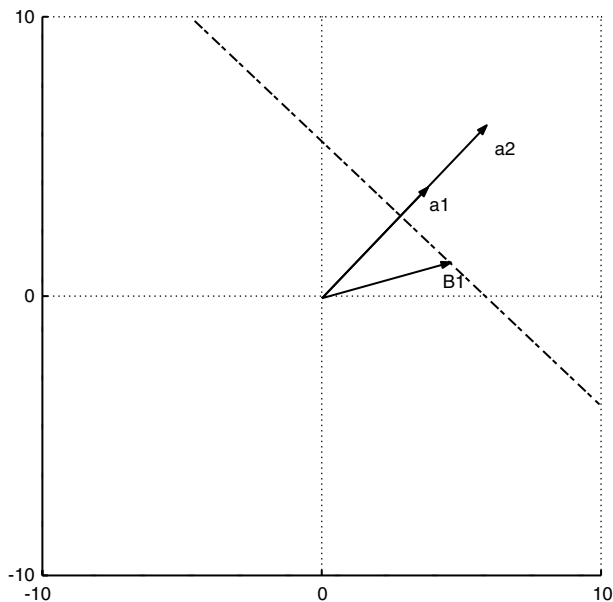
FIGURE 2.
Comparison of RC(M)-association model with distance-association model, 2.

a distance rule. Where the inner product parametrization can best be seen in terms of main and interaction effects, the distance parametrization can best be seen in terms of common and unique contributions, as discussed in Section 2.5.2. The two types of plot are integrated and the user can choose whichever type he prefers.

2.6. Indeterminacies in Model Values, in Distances, and the Degrees of Freedom

The distance-association model has different sets of parameters: i.e. a constant, main effects and coordinates for the row and the column categories. Together they form the expected frequencies. This section studies the number of indeterminacies in this model. Three levels of indeterminacies are considered: (1) The level of expected frequencies; (2) The level of main effects plus interaction effects; (3) The level of interaction effects.

2.6.1. Indeterminacies in Model Values (Expected Frequencies)

The distance formulation shown in (11) is not the most simple formulation to assess the indeterminacies in the model. Therefore, the model is rewritten to get more insight into the indeterminacies

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(\mathbf{X}; \mathbf{Y})$$

$$= \lambda + \lambda_i^{R*} + \lambda_j^{C*} + 2\sum_m x_{im} y_{jm}$$

where $\lambda_i^{R*} = \lambda_i^R - \sum_m x_{im}^2$ and $\lambda_j^{C*} = \lambda_j^C - \sum_m y_{jm}^2$.

This form is well known and the following indeterminacies can be identified: The usual centering indeterminacy in the main effects; in $\mathbf{X}$ and $\mathbf{Y}$ there are indeterminacies for the location, which amounts to $2M$ indeterminacies; and both $\mathbf{X}$ and $\mathbf{Y}$ might be linearly transformed, which amounts to M^2 indeterminacies. Altogether these are $M(M+2)$ indeterminacies in the

association, and for each main effect one. These can be solved in several ways (see below) but these do not leave the distances invariant although the model values are invariant and all solutions are equally valid.

2.6.2. Indeterminacies in Extended Distances

As shown in Section 2.5.2, the model can also be written as a extended distance model with a unique dimension for the row categories and a unique dimension for the column categories

$$\log(\mu_{ij}) = \lambda - d_{ij}^2(\mathbf{X}; \mathbf{Y}; \mathbf{r}; \mathbf{c}).$$

The number of dimensions in the extended distance model is $M + 2$, so the number of parameters is $(I + J) \times (M + 2)$ of which $(I + J)$ are constrained to be equal to zero. This solution can be rotated and translated without changing the distances. The dimensionality is $M + 2$, so the number of indeterminacies is $[(M + 2 + 1) \times (M + 2)]/2 = (M^2 + 5M + 6)/2$. At this level, we keep the λ constant and the distances are invariant, but the coordinates are not unique. Also notice that by rotation of the solution the unique row and column coordinates do not have the same pattern of zeros anymore.

2.6.3. Indeterminacies in Association Distances

Now indeterminacies in the association distances (defined by $\mathbf{X}$ and $\mathbf{Y}$) are studied, for fixed position of λ, $\mathbf{r}$, and $\mathbf{c}$. These distances remain invariant under translation, rotation, and reflection of a solution, which amounts to $M \times (M + 1)/2$ indeterminacies.

2.6.4. Identification and Degrees of Freedom

Once the algorithm has found a solution we have $\hat{\mu}_{ij}$, λ, λ_i^R, λ_j^C, $\mathbf{X}$, $\mathbf{Y}$, and possibly ϵ_i. As shown in Section 2.6.1, indeterminacies are present in the parameter estimates. An identified solution might be obtained by writing all parameters as a function of singular values and singular vectors as follows. First determine $G_{ij} = \log(\hat{\mu}_{ij})$ for $i \neq j$ and $G_{ii} = \log(\hat{\mu}_{ii}) - \epsilon_i$ for the diagonal of square matrices.[1] To obtain identified parameters first define $\tilde{\lambda} = \frac{1}{I \times J} \sum_i \sum_j G_{ij}$, $\tilde{\lambda}_i^R = \frac{1}{J} \sum_j G_{ij} - \tilde{\lambda}$, $\tilde{\lambda}_j^C = \frac{1}{I} \sum_i G_{ij} - \tilde{\lambda}$, and $\delta_{ij} = G_{ij} - \tilde{\lambda} - \tilde{\lambda}_i^R - \tilde{\lambda}_j^C$. Define the matrix $\mathbf{\Delta}$ with elements $\mathbf{\Delta} = \{\delta_{ij}\}$. Let the singular value decomposition of $\mathbf{\Delta}$ be given by $\mathbf{USV'}$. Identified parameters are obtained by

$$\mathbf{X} = \frac{1}{\sqrt{2}} \mathbf{US}^\tau$$

$$\mathbf{Y} = \frac{1}{\sqrt{2}} \mathbf{VS}^\kappa,$$

with $\tau + \kappa = 1$ and

$$d_{x,i} = \sum_m x_{im}^2$$

$$d_{y,j} = \sum_m y_{jm}^2$$

[1] Note that in case no parameters for the diagonal are fitted $\epsilon_i = 0$, $\forall i$.

$$\dot{\lambda}_i^R = \tilde{\lambda}_i^R + d_{x,i}$$

$$\dot{\lambda}_j^C = \tilde{\lambda}_j^C + d_{y,j}$$

$$\lambda = \tilde{\lambda} + \frac{1}{I}\sum_i \dot{\lambda}_i^R + \frac{1}{J}\sum_j \dot{\lambda}_j^C$$

$$\lambda_i^R = \dot{\lambda}_i^R - \frac{1}{I}\sum_i \dot{\lambda}_i^R$$

$$\lambda_j^C = \dot{\lambda}_j^C - \frac{1}{J}\sum_j \dot{\lambda}_j^C.$$

Since the singular value decomposition is unique, and is characterized by $M(M+2)$ constraints, we have resolved the indeterminacies of the model. Through this set of constraints the difference between the graphical displays of the RC(M) and distance-association model is just a scaling factor. The plot can thus be interpreted by both a distance rule and an inner product rule. The degrees of freedom are equal to $(I \times J) - 1 - (I-1) - (J-1) - (I+J)M + M(M+2) = (I-1-M)(J-1-M)$.

For the one-mode distance-association model things are less complicated. For every dimension K parameters need estimation, with one parameter for each dimension saved from centering the solution, and $M(M-1)/2$ parameters saved for rotational freedom. For a M-dimensional model, the degrees of freedom are reduced with $(K-1) \times M - [M(M-1)/2]$ compared to the independence model, with the degrees of freedom then equaling to $(K-1)^2 - (K-1)M + [M(M-1)/2]$. The maximum number of dimensions that can be fitted is equal to $K-1$, providing the model closest to the quasi-symmetry model of Caussinus (1965) assuming that the association parameters satisfy the metric axioms and are Euclidean. Under these conditions, K points can be fit perfectly in $K-1$ dimensions.

3. Data Analysis

The distance-association models will be applied to the data described in Tables 1, 2, and 3. Table 1 concerns reproductive behavior of male bitterlings, studied by Wiepkema (1961), with data derived from 13 sequences using a moving time-window of size two. The behaviors are jerking (jk), turning beats (tu), head butting (hb), chasing (chs), fleeing (fl), quivering (qu), leading (le), head down posture (hdp), skimming (sk), snapping (sn), chafing (chf), and finflickering (ffl). These data were analyzed by Van der Heijden (1987) using correspondence analysis. This data will be used to show both the one- and two-mode distance-association models for square contingency tables.

Table 2 is from Srole, Langner, Michael, Opler, and Rennie (1962) and gives a cross-classification of subjects according to their mental health status and parents' socio-economic status. Mental health has four categories: (a) well; (b) mild symptom formation; (c) moderate symptom formation; and (d) impaired. There are six categories of socio-economic status. These data have been analyzed by Goodman (1979) using the RC(1)-association model. This data will be used to show the extended distance model, as discussed in Section 2.5.2.

Table 3 gives a cross-classification of eye and hair color for 5,387 children from Caithness, Scotland, and has been analyzed previously by Goodman (1981) and Becker and Clogg (1989). Eye color has four categories: blue, light, medium, and dark; hair color has five categories: fair, red, medium, dark, and black. This data set will be used to discuss the representation of odds ratios by distances as discussed in Section 2.4. For all analysis τ and κ will be set to 1/2.

TABLE 1.
Cross-classification of reproductive behavior of the male bitterling of $t-1$ versus t

follow preceed	jk	tu	hb	chs	fl	qu	le	hdp	sk	sn	chf	ffl
jk	654	128	172	56	27	25	1	28	0	46	14	18
tu	101	132	62	27	5	1	1	11	0	8	5	9
hb	171	62	197	130	0	25	0	50	14	18	14	12
chs	60	22	152	135	0	8	0	43	16	15	12	4
fl	19	2	0	0	419	19	0	2	0	17	5	11
qu	36	1	18	5	12	789	119	295	26	70	1	14
le	4	0	0	0	0	57	167	73	0	8	0	0
hdp	22	9	40	37	5	245	7	171	287	53	8	13
sk	3	2	7	38	0	120	8	134	19	28	4	0
sn	42	2	17	16	20	70	11	67	9	225	12	12
chf	18	3	10	13	6	5	0	8	0	24	97	9
ffl	27	3	6	5	10	13	0	18	0	10	8	29

Note: jk = jerking, tu = turning beats, hb = head butting, chs = chasing, fl = fleeing, qu = quivering, le = leading, hdp = head down posture, sk = skimming, sn = snapping, chf = chafing, ffl = finflickering.

TABLE 2.
Cross-classification of subjects according to their mental health and their parents' socioeconomic status

	1	2	3	4	5	6
Well (a)	64	57	57	72	36	21
Mild (b)	94	94	105	141	97	71
Moderate (c)	58	54	65	77	54	54
Impaired (d)	46	40	60	94	78	71

TABLE 3.
Cross-classification of eye color by hair color

	Fair	Red	Medium	Dark	Black
Blue	326	38	241	110	3
Light	688	116	584	188	4
Medium	343	84	909	412	26
Dark	98	48	403	681	85

Comparisons of models will be made through both the chi-square statistic

$$X^2 = \sum_{ij} \frac{(f_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}},$$ (24)

where $\hat{\mu}_{ij}$ denotes the maximum likelihood estimate of the expected frequency, and the Likelihood Ratio statistic

$$\text{LR} = 2 \sum_{ij} f_{ij} \log \frac{f_{ij}}{\hat{\mu}_{ij}}.$$ (25)

The Likelihood Ratio statistic can be used to compare two nested models, and under the independence model gives a measure for the total amount of association in a table. Given a distance model in M-dimensions ($D(M)$), define the percentage association accounted for (%AAF) as (Goodman, 1971)

TABLE 4.
Results of the analysis of bitterling data (Table 2)

Model	Measure	1D	2D	3D
Symmetric	X^2	8771	7110	1627
	LR	4528	2075	975
	df	110	100	91
Asymmetric	X^2	39608	3510	487
	LR	3290	1061	410
	df	100	81	64
Symmetric +	X^2	1169	463	349
inheritance	LR	1114	465	373
	df	98	88	79
Asymmetric +	X^2	1112	294	109
inheritance	LR	1101	252	117
	df	88	69	52

$$\%\text{AAF} = 100 \times \frac{\text{LR}_I - \text{LR}_{D(M)}}{\text{LR}_I}, \tag{26}$$

where LR_I is the Likelihood Ratio statistic under the Independence Model, and $\text{LR}_{D(M)}$ under a distance model in M dimensions. In the formula above, a distance model is compared to the independence model, but we could use the same formula to compare any two nested models (e.g., one model in two dimensions and another in three dimensions to obtain a measure of association accounted for by the third dimension). The %AAF is especially useful in large contingency tables where the traditional chi-squared distributed statistics tend to dismiss all models except the one that is saturated because of the sample size. The %AAF then gives a relative measure of how much information is explained by a model compared to another.

3.1. Analysis of Table 1

The independence model strongly deviates from the data ($X^2 = 16985$, LR $= 10485$, df $= 121$); the quasi-independence model (i.e., the independence model plus parameters to fit the diagonal of a square table) fits the data much better as 77% of the association is accounted for, but still fits badly ($X^2 = 4156$, LR $= 3825$, df $= 109$). Table 4 gives goodness of fit statistics for the distance-association models in 1D to 3D.

3.1.1. The One-Mode Distance-Association Model With and Without Inheritance Terms

Although none of the models fits the data using the chi-square statistics, all give a high value for the %AAF. Compared to the independence model, the one-dimensional model accounts for 73% of the association, the two-dimensional 88%, and the three-dimensional 94%.

The one-mode distance-association models with inheritance terms account for 89, 97, and 98% of the association in one to three dimensions, respectively. Compared to quasi-independence, 71% of the association is accounted for by the one-dimensional model. The increase in %AAF for the two-dimensional model is reasonably large, and in two dimensions the one-mode distance-association model accounts for 88% of the association not accounted for by the quasi-independence model. The %AAF for the three-dimensional model compared to quasi-independence is 90%, not much better than two dimensions.

In Figure 3, fleeing is on the boundary of the solution, indicating not many transitions occur to or from this category. Two close groups can be distinguished, with one of chasing, head butting,
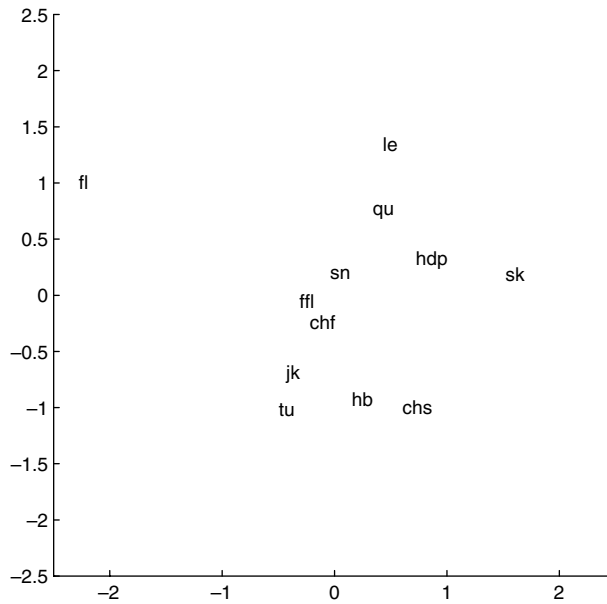
FIGURE 3.

The configuration for the two-dimensional one-mode distance-association model with inheritance terms for the bitterling data. For abbreviations, see the text.

jerking, turning beats, and finflickering; and another group of skimming, snapping, head down posture, quivering and leading. Many transitions occur within these two groups. The horizontal dimension contrasts fleeing from the other categories. The vertical dimension might be interpreted as aggressive versus sexual behavior: turning beats, jerking, head butting and chasing represent aggressive behavior, whereas leading, quivering, skimming and head down posture represent sexual behavior. Wiepkema (1961) used factor analysis on Spearman rank correlations for the analyses (see Van der Heijden, 1987), and found three factors of sexual, aggressive, and non-reproductive behavior (consisting of the categories fleeing, finflickering, chafing, and snapping). Using correspondence analyses Van der Heijden (1987) found four dimensions, of which the first two lead to conclusions very similar to ours.

3.1.2. The Two-Mode Distance-Association Model With and Without Inheritance Terms

The two-mode distance-association models without parameters for the diagonal cells account for 81, 94, and 98% of the association in the data in one to three dimensions, respectively. With inheritance terms, these values go up to 94, 99 and 99%. Compared to quasi-independence, they account for 71% (one dimension), 93% (two dimensions), and 97% (three dimensions) of the association not accounted for by the quasi-independence model.

Although the two-mode models in two and three dimensions plus inheritance terms do not fit the data in terms of the model test, they account for 99% of the association in the table. Compare Figure 4, where the observed frequencies (horizontally) are plotted against the fitted frequencies (vertical dimension) for the two-mode distance-association model in two dimensions plus inheritance terms. Overall, the points are close to a straight line (left-hand-side plot) representing perfect fit, although in the smaller frequencies there are reasonable discrepancies between observed and fitted frequencies (right-hand-side plot). For tables with large frequencies the chi-square distributed statistics tend to dismiss all models except one that is saturated; in contrast to the use of chi-square statistics, the %AAF is to be preferred as a measure of fit.
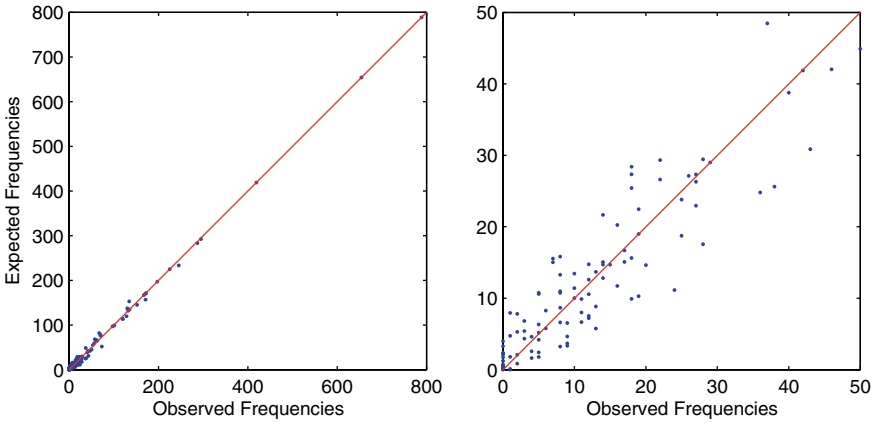
FIGURE 4.
Observed versus expected frequencies for two-dimensional two-mode distance-association model with inheritance terms.
The left-hand-side plot shows all frequencies; The right-hand-side plot shows a detailed look at the smaller frequencies.

The two-dimensional two-mode distance-association model plus inheritance term seems to fit the data very well. The solution is given by Figure 5, where small letters correspond to points for the first time (rows) and capitals to points for the second time (columns). The configuration is very much the same as the one-mode configuration in Figure 5 except that some important asymmetries are now apparent. For example, the distance from leading (le) at the first time point to skimming (SK) at the second is much smaller than the distance from skimming at the first time (sk) to leading at the second (LE). Other similar conclusions can be drawn from this representation (compare head butting/ chasing or snapping/ fleeing).
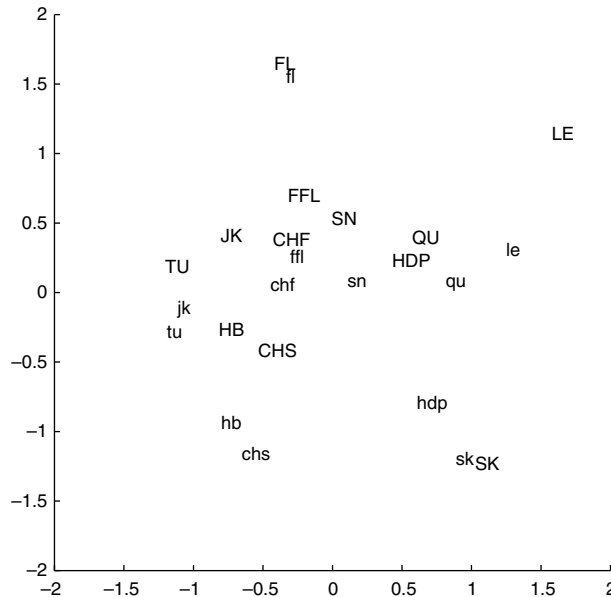

FIGURE 5.
The configuration for the two-dimensional two-mode distance-association model plus inheritance terms. Small letters correspond to the categories at $t - 1$; Capitals correspond to the categories at t. For abbreviations, see the text.
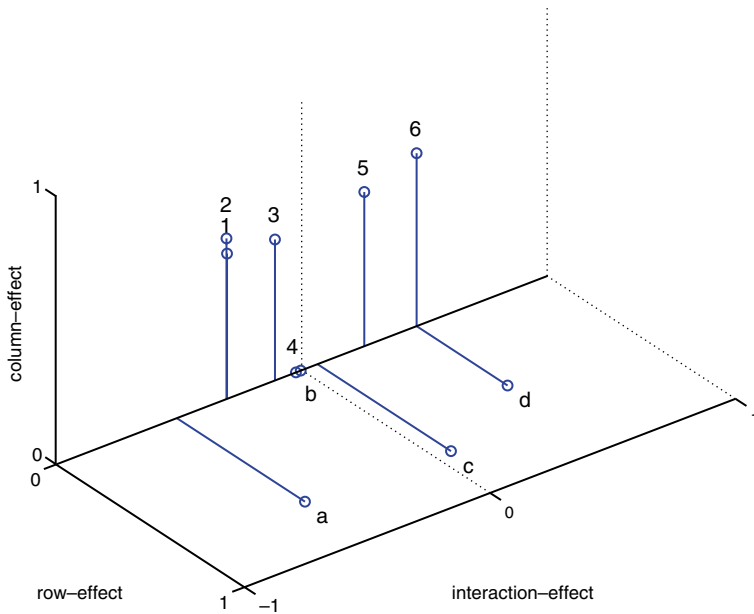
FIGURE 6.
Distance-association model with main effects represented as a row and column dimension.

3.2. Analysis of Table 2

The Srole data have been analyzed by Goodman (1979) and Haberman (1974, 1979). The one-dimensional distance-association model fits the data well ($X^2 = 3.57$ and LR = 3.57 with df = 8). Figure 6 presents the extended distance model representation as discussed in Section 2.5.2. One dimension pertains to the row points and one to the column points, both representing the unique contributions of the categories, the third dimension is a common dimension pertaining to the association. The squared distances in three-dimensional space approximate the frequencies in a monotone decreasing manner, where the larger the distance, the smaller the frequency, and the distance represents the departure from the maximum frequency (for categories b and 4). The marginal frequencies are represented in a row and a column dimension, where the closer the point to the origin, the higher the marginal frequency. Through the use of squared Euclidean distances the unique and common dimensions are additive: The squared distance from a row point, say a, to a column point, say 6, equals the squared distance on the row dimension from zero to point a plus the squared distance on the column dimension from zero to point 6 plus the squared distance on the common dimension from a to 6. This representation is only possible for the one-dimensional distance-association model, otherwise two dimensions are needed for the association and the total configuration will be four-dimensional.

3.3. Analysis of Table 3

This cross-classification was analyzed with the RC(M)-association model in two dimensions by Becker and Clogg (1989), and here we show the graphical representation of the distance-association model. The model fits the data well, with $X^2 = 5.29$ and LR = 5.41 on two degrees of freedom. Figure 7 displays the fitted points in a joint plot, where capital letters denote hair color, and standard letters represent eye color.
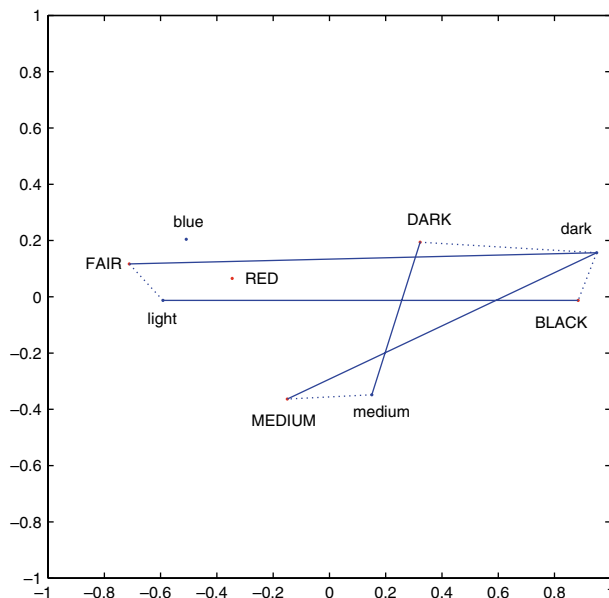
FIGURE 7.
The graphical representation obtained with distance-association model for the eye-hair color cross-classification. Included is the representation of log-odds ratios by sums of positive (solid lines) and negative (dotted lines) coded distances.

Figure 7 is used to discuss the odds ratio under the model. As Section 2.4 indicated the log-odds ratio is expressed as a sum of positive and negative squared distances. Beginning with the odds ratio of the four cells with light versus dark eye color, and Fair versus Black hair color, the log-odds ratio is equal to

$$d^2(\text{light, Black}) + d^2(\text{dark, Fair}) - d^2(\text{light, Fair}) - d^2(\text{dark, Black}), \qquad (27)$$

where the first two terms are represented as solid lines in Figure 7 and the latter two as dotted lines. The log-odds ratio is clearly positive, so people with light eyes tend to have fair hair more often, compared to people with dark eyes, who have black hair more often. We do not immediately get a value From the graphical representation, but since the solid lines are much longer than the dotted, this odds ratio will be large, and larger than the odds ratio for the cells corresponding to medium versus dark eye color and medium versus dark hair color. This log-odds ratio is also represented in Figure 7 again by solid (coded positive in the odds ratio) and dotted lines (coded negative in the odds ratio) that have to be added. The positive parts are still larger compared to the negative parts so the log-odds ratio will be positive, but the relationship is not as strong compared to the odds ratio of light and dark eyes, and fair and black hair color.

4. Discussion

Distance-association models were developed in this paper for the analysis of contingency tables that can fit multidimensional solutions; the distance models proposed by Goodman (1972) and Haberman (1974) can only handle unidimensional distances with some a priori known ordering. Our models have an advantage when the variables are nominal and no natural order of the categories exists, even though finding an optimal unidimensional scale is a hard combinatorial problem. Our multidimensional solutions avoid this problem. Another advantage lies in the analysis of multivariate transition data where unidimensional solutions are unlikely to fit. Takane
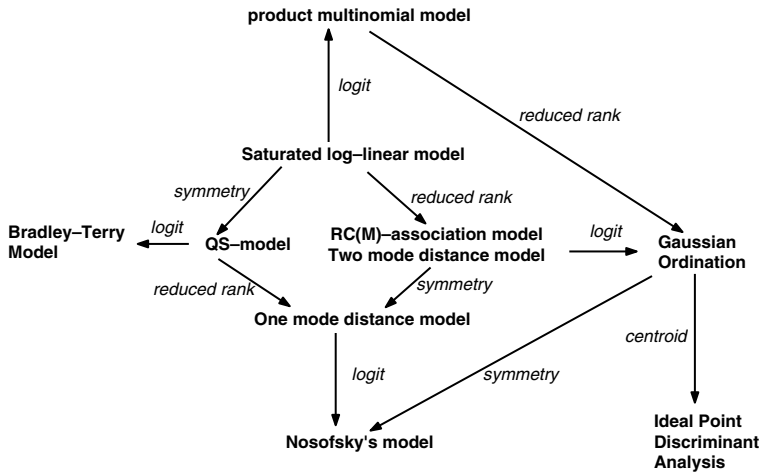
FIGURE 8.
Relationships between different models through symmetry, rank, and logit restrictions on the association term.

(1987) also proposed a multidimensional model, but adopted a conditional approach. Comparing the conditional probability model with our distance-association model, the differences are that IPDA models the conditional probability where we model the joint probability or frequency. In IPDA, the coordinates of the second set are not estimated but are fixed at the weighted centroid of the row coordinates. A feature included in the IPDA model is that design matrices can be incorporated to fit the row coordinates, which is especially useful when there are multiple predictors and no interactions are assumed between these predictors. Such design matrices for the row or column coordinates can be included in our models also. Thus, we can conclude that the IPDA model is a constrained version of our model.

In Section 2.5.3, relationships were noted between our models and other reduced rank models for contingency tables. An often-used device for square contingency tables is the quasi-symmetry model (QS-model; Caussinus, 1965), and it can be shown that the one-mode distance-association model is a reduced rank version of the quasi-symmetry model. In full dimensionality the models are equal. Figure 8 shows relationships between the saturated log-linear, the RC(M)-association, the quasi-symmetry, and the models defined in the present paper. The QS-model is equal to the saturated log-linear model with symmetry restrictions on the association parameters. The RC(M)-association model and the two-mode distance-association model impose rank restrictions on the association term. By imposing metric restrictions on the symmetric association terms of the QS-model, the one-mode distance-association model is obtained; by imposing symmetry restrictions on the RC(M)-association model and the two-mode distance-association model the one-mode distance-association model is generated. These symmetry restrictions can only be imposed in the case of square matrices. All these models can also be put in logit formulation. The logit formulation of the log-linear model is the product multinomial model. The logit formulation of the two-mode distance-association model is known as Gaussian ordination (Ihm & Van Groenewoud, 1975, 1984). Attempts to approximate the Gaussian ordination model with correspondence analysis can be found in Ter Braak (1985). By imposing symmetry restrictions on the Gaussian ordination model, Nosofsky's model is obtained. This model is a logit formulation of the one-mode distance-association model. Fienberg and Larntz (1976) showed that the logit formulation of the QS-model is the Bradley–Terry model (Bradley & Terry, 1952). By imposing the centroid condition on the Gaussian ordination model, IPDA is obtained.

By building a distance model for two-way contingency tables, we arrived at a reparametrization of the well-known RC(M)-association model. We believe that the distance-association

models provide a useful tool for the analysis of association, since a distance model has interpretational advantages over inner product models. Particularly, when the RC(M)-association model with one component does not fit the data, our distance-association models have a clear advantage in representing the association and log-odds ratios in terms of distances. A simple interpretation of the odds ratio or log-odds ratio exists in terms of the distances, as does a graphical representation. Until now such a graphical representation has not been available. The present representation is a new helpful tool in understanding the association in cross-classified data. Meulman and Heiser (1998) studied the representation of odds ratios in homogeneity analysis (multiple correspondence analysis), and in full dimensional space found that the odds are represented as a ratio of distances. Note, however, that in homogeneity analysis, the full space is projected on a lower dimensionality and distances are generally not preserved in projection. Here, expressions were provided for the odds ratios in terms of distances in the low-dimensional space.

Two types of joint plots were distinguished with Type I having an interpretation in terms of distances and Type II having an interpretation through inner products. Type II plots should always be made so the inner product relationship is clear, and thus at least one set should be represented by vectors. Often, inner product models are represented by Type I plots, and we strongly urge the use of correct plotting procedures, otherwise incorrect interpretations seem unavoidable: when only points are used to represent the categories of the two sets, interpretation will be based on distances.

A question that has been kept implicit in the paper but which is basic to the problem under study is to explain what interaction is: In statistical models, interaction is often seen as the residual from the independence or main effects model; and on the residual quantities some kind of multiplicative model is fitted, called interaction between variables. In psychometrics, in contrast to the statistical point of view, a distance model is often the starting point, and unique contributions are fitted to the residuals of the distance model. So it seems that in traditional statistical methods the emphasis is on main effects, whereas in psychometric models the emphasis is on the association part. Recent statistical developments like regression trees (Breiman, Friedman, Olshen, & Stone, 1984) also place emphasis on the association or interaction part.

A. Appendix A: Maximum Likelihood Estimation

We will give an algorithm for obtaining coordinates in Euclidean space, first for the one-mode distance-association model and afterwards for the two-mode distance-association model.

Appendix A.1: ML-Estimation for the One-Mode Distance-Association Model

For the one-mode distance-association model specified in Section 2.2, estimates of the parameters can be found by maximizing the likelihood under one of the usual sampling schemes (i.e., Poisson sampling, Multinomial sampling or Product Multinomial sampling). The log-likelihood function under independent Poisson sampling can be written

$$L = \sum_{ij} f_{ij} \log \mu_{ij} - \sum_{ij} \mu_{ij}, \tag{28}$$

where f_{ij} are the observed frequencies, and μ_{ij} the expected frequencies. Expanding

$$L = f_{++}\lambda + \sum_i f_{i+}\lambda_i^R + \sum_j f_{+j}\lambda_j^C - \text{tr}\mathbf{X}'\mathbf{G}\mathbf{X}$$

$$- \sum_{ij} \exp[\lambda + \lambda_i^R + \lambda_j^C - \text{tr}\mathbf{X}'\mathbf{A}_{ij}\mathbf{X}]. \tag{29}$$

The matrix $\mathbf{G}$ has elements $g_{ij} = -(f_{ij} + f_{ji})$ if $i \neq j$, and $g_{ii} = -\sum_l g_{il}$ else. We will call this the *Sym*-operation.

An iterative method based on the Newton–Raphson theory can be given to maximize this function. Standard methods show that the Maximum Likelihood Estimates must satisfy the following system of equations

$$f_{i+} - \hat{\mu}_{i+} = 0;$$

$$f_{+j} - \hat{\mu}_{+j} = 0;$$

$$\left(\hat{\boldsymbol{\Gamma}} - \mathbf{G} \right) \mathbf{x}_s = \mathbf{0}, \tag{30}$$

where the matrix $\hat{\boldsymbol{\Gamma}}$ is defined in the same way as the matrix $\mathbf{G}$ but now on the estimated expected frequencies, and $\mathbf{x}_s$ is the s-th column of $\mathbf{X}$. Showing the derivation of the third equation, and taking the derivative of the likelihood function (29) with respect to $\mathbf{x}_s$, we get the following

$$\frac{\partial L}{\partial \mathbf{x}_s} = -2\mathbf{G}\mathbf{x}_s + 2 \sum_{ij} \hat{\mu}_{ij} \mathbf{A}_{ij} \mathbf{x}_s$$

$$= 2\hat{\boldsymbol{\Gamma}}\mathbf{x}_s - 2\mathbf{G}\mathbf{x}_s. \tag{31}$$

Now a fitting scheme can be derived. For the row and column parameter, the updates are standard. The Hessian for the configuration matrix is a $M \times M$ block matrix with each block equal to

$$\frac{\partial^2 L}{\partial \mathbf{x}_s \partial \mathbf{x}_t'} = 2\delta^{st}(\hat{\boldsymbol{\Gamma}} - \mathbf{G}) - 4 \sum_{ij} \hat{\mu}_{ij} \mathbf{A}_{ij} \mathbf{x}_s \mathbf{x}_t' \mathbf{A}_{ij}, \tag{32}$$

where δ^{st} is equal to one if $s = t$, zero otherwise. Each block is of size $K \times K$. The updates in the $t + 1$st cycle, according to the Newton–Raphson theory are given by

$$\lambda_i^{R(t+1)} = \lambda_i^{R(t)} + \frac{(f_{i+} - \tilde{\mu}_{i+})}{\tilde{\mu}_{i+}}; \tag{33}$$

$$\lambda_j^{C(t+1)} = \lambda_j^{C(t)} + \frac{(f_{+j} - \tilde{\mu}_{+j})}{\tilde{\mu}_{+j}};$$

$$\text{Vec}(\mathbf{X}^{(t+1)}) = \text{Vec}(\mathbf{X}^{(t)}) - \tau \mathbf{H}^- \left(\mathbf{I}_M \otimes [\tilde{\boldsymbol{\Gamma}} - \mathbf{G}] \right) \text{Vec}(\mathbf{X}^{(t)}),$$

where $\tilde{\mu}$ is computed using the parameter estimates of the previous cycle. The matrix $\mathbf{H}^- = (\mathbf{H} + \mathbf{nn}'/\mathbf{n}'\mathbf{n})^{-1} - \mathbf{nn}'/\mathbf{n}'\mathbf{n}$, with $\mathbf{n}$ the null-space of $\mathbf{H}$, is a generalized inverse that satisfies the Moore–Penrose conditions of $\mathbf{H}$, the Hessian, defined by (32), and τ denotes a step-size determined by a line-search procedure. $\text{Vec}(\cdot)$ denotes the operation in which the columns of a matrix are stacked vertically. The Kronecker product ($\otimes$) produces a block-diagonal matrix with the matrix $\tilde{\boldsymbol{\Gamma}} - \mathbf{G}$ on the diagonal.

Experience shows that the likelihood function has many local maxima and at convergence one should always check the relations described by (30). Furthermore, one should check the Hessian because it must be negative semi-definite with $M + M(M - 1)/2$ eigenvalues equal to zero (Groenen, De Leeuw, & Mathar, 1996).

To start the iteration process, initial estimates have to be obtained. We follow Becker's (1990) procedure but instead of a singular value decomposition we perform an eigenvalue decomposition on the symmetrized matrix. Another option is to do an eigenvalue decomposition on $\mathbf{G}$. To check whether the algorithm derived a global optimum, a number of random starts should always be performed.

Appendix A.2: ML-Estimation for the Two-Mode Distance-Association Model

The likelihood function under Poisson sampling for the asymmetric distance model becomes

$$L = f_{++}\lambda + \sum_i f_{i+}\lambda_i^R + \sum_j f_{+j}\lambda_j^C - \sum_{ij} f_{ij}d_{ij}^2(\mathbf{X}; \mathbf{Y})$$

$$- \sum_{ij} \exp[\lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(\mathbf{X}; \mathbf{Y})]$$

$$= f_{++}\lambda + \sum_i f_{i+}\lambda_i^R + \sum_j f_{+j}\lambda_j^C$$

$$-\mathrm{tr}\mathbf{X}'\mathbf{D}_R\mathbf{X} - \mathrm{tr}\mathbf{Y}'\mathbf{D}_C\mathbf{Y} + 2\mathrm{tr}\mathbf{X}'\mathbf{F}\mathbf{Y}$$

$$- \sum_{ij} \exp[\lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(\mathbf{X}; \mathbf{Y})] \tag{34}$$

where $\mathbf{D}_R = \mathrm{diag}(\sum_j f_{ij})$, $\mathbf{D}_C = \mathrm{diag}(\sum_i f_{ij})$, and $\mathbf{F}$ is the matrix containing the observed frequencies. Differentiating (34) with respect to x_{am} (the coordinate for category a of the row variable on dimension m),

$$\frac{\partial L}{\partial x_{am}} = 2 \sum_j (\hat{\mu}_{aj} - f_{aj})(x_{am} - y_{jm}). \tag{35}$$

It is well known in least squares that for finding a solution to the MDU problem, an MDS algorithm can be used in which the input matrix has a special structure. The matrix has size $(I + J) \times (I + J)$, and in the upper-right block the original contingency table is placed with the other parts equal to zero. If we apply the *Sym*-operation to this partitioned matrix a matrix is obtained with upper-right $(I \times J)$ and lower-left $J \times I$ corners containing matrices with elements $-f_{ij}$ and $-f_{ji}$. In the upper-left block, a diagonal matrix with elements equal to the row sums of our original contingency table is placed; in the lower-left block there is a diagonal matrix with the column sums of our original contingency table. If this matrix $\mathbf{G}$ is multiplied with $\mathbf{S}$ (with $\mathbf{X}$ and $\mathbf{Y}$ concatenated), the correct form, as described in (35) is generated. To obtain starting values we can use Becker's (1990) procedure.

Experience shows that the procedure described here works reasonably well for low-dimensional solutions, but the estimation of the two-mode distance-association model gets troublesome with large tables in three or more dimensions. In that case, it is better not to use the whole Hessian, and when only the diagonal is considered the algorithm works well. The algorithm then reduces to a quasi Newton algorithm in which the Hessian is replaced by a simpler form.

References

Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North-Holland.

Becker, M. P. (1990). Maximum likelihood estimation of the RC(M) association model. *Applied statistics*, *39*, 152–167.

Becker, M. P., & Clogg, C. C. (1989). Analysis of sets of two-way contingency tables using association models. *Journal of the American Statistical Association*, *84*, 142–151.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling; theory and applications*. New York: Springer.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete designs I. the method of paired comparisons. *Biometrika*, *39*, 324–345.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1986). Interpoint distance comparison in correspondence analysis. *Journal of Marketing Research*, *23*, 271–280.

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1987). Comparing interpoint distance in correspondence analysis: A clarification. *Journal of Marketing Research*, *24*, 445–450.

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1989). Reply to Greenacre's commentary on the Carroll-Green-Schaffer scaling of two-way correspondence analysis solutions. *Journal of Marketing Research*, *26*, 366–368.

Caussinus, H. (1965). Contribution á l'analyse statistique des tableaux de corrélation [contributions to the statistical analysis of correlation matrices]. *Annals of the faculty of Science, University of Toulouse*, *29*, 715–720.

Clogg, C. C., Eliason, S. R., & Wahl, R. J. (1990). Labor-market experiences and labor force outcomes. *American Journal of Sociology*, *95*, 1536–1576.

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

De Leeuw, J., & Heiser, W. J. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingoes, E. E. Roskam, & I. Borg (Eds.), *Geometric representations of relational data* (pp. 735–752). Ann Arbor, MI: Mathesis Press.

De Rooij, M. (2001). *Distance models for the analysis of transition frequencies*. Unpublished doctoral dissertation, Leiden University.

De Rooij, M., & Heiser, W. J. (2002). A distance representation of the quasi-symmetry model and related distance models. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments on psychometrics: proceedings of the international meeting of the psychometric society* (pp. 487–494). Tokyo: Springer-Verlag.

Defays, D. (1978). A short note on a method of seriation. *British Journal of Mathematical and Statistical Psychology*, *3*, 49–53.

Fienberg, S. E., & Larntz, K. (1976). Loglinear representation of paired and multiple comparison models. *Biometrika*, *63*, 245–254.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.

Gilula, Z., & Haberman, S. J. (1986). Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association*, *81*, 780–788.

Goodman, L. A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, *13*, 33–61.

Goodman, L. A. (1972). Some multiplicative models for the analysis of cross-classified data. *Sixth Berkely Symposium*, *1*, 649–696.

Goodman, L. A. (1979). Simple models for the analysis of association in cross classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537–552.

Goodman, L. A. (1981). Association models and cannonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, *76*, 320–334.

Goodman, L. A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetric 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 log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243–309.

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

Greenacre, M. J. (1989). The Carroll-Green-Schaffer scaling in correspondence analysis: A theoretical and expirical appraisal. *Journal of Marketing Research*, *26*, 358–365.

Groenen, P. J. F., De Leeuw, J., & Mathar, R. (1996). Least squares multidimensional scaling with transformed distances. In W. Gaul & D. Pfeifer (Eds.), *Studies in classification, data analysis, and knowledge organization* (pp. 177–185). Berlin: Springer.

Haberman, S. J. (1974). *The analysis of frequency data*. Chicago: university of Chicago Press.

Haberman, S. J. (1978). *Analysis of qualitative data, (vol. 1)*. New York: Academic Press.

Haberman, S. J. (1979). *Analysis of qualitative data (vol. 2)*. New York: Academic Press.

Haberman, S. J. (1995). Computation of maximum likelihood estimates in association models. *Journal of the American Statistical Association*, *90*, 1438–1446.

Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Unpublished doctoral dissertation, Leiden University.

Heiser, W. J. (1987). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189–221). Berlin: Springer Verlag.

Heiser, W. J. (1988). Selecting a stimulus set with prescribed structure from empirical confusion frequencies. *British Journal of Mathematical and Statistical Psychology*, *41*, 37–51.

Heiser, W. J., & Meulman, J. (1983). Analyzing rectangular tables by joint and constrained multidimensional scaling. *Journal of Econometrics*, *22*, 139–167.

Hubert, L. J., & Arabie, P. (1986). Unidimensional scaling and combinatorial optimization. In J. De Leeuw, W. J. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 181–196). Leiden: DSWO press.

Ihm, P., & Van Groenewoud, H. (1975). A multivariate ordering of vegetation data based on Gaussian type gradient response curves. *Journal of Ecology*, *63*, 767–778.

Ihm, P., & Van Groenewoud, H. (1984). Correspondence analysis and Gaussian ordination. *COMPSTAT Lectures*, *3*, 5–60.

Meulman, J. J., & Heiser, W. J. (1998). Visual display of interaction in multiway contingency tables by use of homogeneity analysis: the 2x2x2x2 case. In J. Blasius & M. J. Greenacre (Eds.), *Visualization of categorical data* (pp. 277–296). New York: Academic Press.

Meulman, J. J., Heiser, W. J., & SPSS Inc. (1999). *Spss Categories 10.0*. Chicago, IL: SPSS Inc.

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

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, *38*, 415–432.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.

Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., & Rennie, T. A. C. (1962). *Mental health in the metropolis: The midtown Manhattan study*. New York: McGraw-Hill.

Takane, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrika*, *52*, 493–513.

Takane, Y. (1998). Visualization in ideal point discriminant analysis. In J. Blasius & M. J. Greenacre (Eds.), *Visualization of categorical data* (pp. 441–459). New York: Academic Press.

Takane, Y., & Shibayama, T. (1986). Comparison of models for the stimulus recognition data. In J. De Leeuw, W. J. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 119–138). Leiden: DSWO Press.

Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Probabilistic multidimensional models of perception and cognition* (pp. 335–362). Hillsdale, NJ: Erlbaum.

Ter Braak, C. J. F. (1985). Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, *41*, 859–873.

Van der Heijden, P. G. M. (1987). *Correspondence analysis of longitudonal categorical data*. Leiden: DSWO.

Van der Heijden, P. G. M., Mooijaart, A., & Takane, Y. (1994). Correspondence analysis and contingency models. In M. J. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 79–111). New York: Academic Press.

Wiepkema, P. R. (1961). An ethological analysis of the reproductive behavior of the bitterling (rhodeus amarus bloch). *Archives Neerlandais Zoologique*, *14*, 103–199.

Winsberg, S., & Carroll, J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika*, *54*, 217–229.