

## ANALYSIS OF CONTINGENCY TABLES BY IDEAL POINT DISCRIMINANT ANALYSIS

YOSHIO TAKANE  
MCGILL UNIVERSITY

Cross-classified data are frequently encountered in behavioral and social science research. The loglinear model and dual scaling (correspondence analysis) are two representative methods of analyzing such data. An alternative method, based on ideal point discriminant analysis (DA), is proposed for analysis of contingency tables, which in a certain sense encompasses the two existing methods. A variety of interesting structures can be imposed on rows and columns of the tables through manipulations of predictor variables and/or as direct constraints on model parameters. This, along with maximum likelihood estimation of the model parameters, allows interesting model comparisons. This is illustrated by the analysis of several data sets.

Key words: Multidimensional scaling, model evaluation, AIC, the loglinear model, dual scaling (correspondence analysis), canonical analysis, the RC association model.

### 1. Introduction

The data we encounter in behavioral and social sciences are often qualitative. These qualitative data may be summarized in a two-way contingency or cross-classification table whose  $i, j$ -th entry (entry in the  $i$ -th row and the  $j$ -th column) represents the frequency of joint occurrence of the  $i$ -th category of one variable and the  $j$ -th category of another. Each row or column category may, however, be defined as a combination of categories of more than one variable.

Consider, as an example, the data in Table 1. This data set is a hypothetical data set originally contrived by Maxwell (1961) for demonstrating his method of discriminant analysis. The data consist of three criterion groups, schizophrenic (SC), manic-depressive (MD) and anxiety state (AX), and four binary predictor variables each indicating either presence (1) or absence (0) of a certain symptom. The four symptoms are anxiety (A), suspicion (S), schizophrenic type of thought disorders (T), and delusions of guilt (G). The four binary variables were factorially combined to form 16 distinct patterns of symptoms (predictor patterns), and each of these patterns is identified with a row of the table, which contains the cross-classification of 620 patients according to the 16 patterns of symptoms and the three criterion groups.

One may be tempted to apply the loglinear model (e.g., Andersen, 1980; Bishop, Fienberg & Holland, 1975), or dual scaling or correspondence analysis (Greenacre, 1984; Nishisato, 1980) to this data set. If the loglinear analysis is chosen, detailed model evaluation is possible to test which main effects and/or interactions among the predictor variables are statistically significant in discriminating the criterion groups. With this

Presented as the Presidential Address to the Psychometric Society's Annual and European Meetings, June, 1987. Preparation of this paper was supported by grant A6394 from the Natural Sciences and Engineering Research Council of Canada. Thanks are due to Chikio Hayashi of University of the Air in Japan for providing the ISM data, and to Jim Ramsay and Ivo Molenaar for their helpful comments on an earlier draft of this paper.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1, CANADA.

Table 1. Data from Maxwell (1961)

Pattern Number	Predictor Pattern				Observed Frequencies in Groups			
	A	S	T	G	SC	MD	AX	
1	0	0	0	0	38	69	6	
2	0	0	0	1	4	36	0	
3	0	0	1	0	29	0	0	
4	0	0	1	1	9	0	0	
5	0	1	0	0	22	8	1	
6	0	1	0	1	5	9	0	
7	0	1	1	0	35	0	0	
8	0	1	1	1	8	2	0	
9	1	0	0	0	14	80	92	
10	1	0	0	1	3	45	3	
11	1	0	1	0	11	1	0	
12	1	0	1	1	2	2	0	
13	1	1	0	0	9	10	14	
14	1	1	0	1	6	16	1	
15	1	1	1	0	19	0	0	
16	1	1	1	1	10	1	0	
					Total	224	279	117

approach, however, it is relatively difficult to understand the mutual relationship among the predictor patterns and the criterion groups. If, on the other hand, dual scaling is applied, a spatial representation of the 16 patterns and the three criterion groups is obtained, through which one can readily understand their mutual relationship. This approach, however, does not allow detailed model evaluation. Van der Heijden and de Leeuw (1985) consequently recommend complementary use of the two methods; dual scaling to find meaningful structural hypotheses in the data, and the loglinear model to test the hypotheses.

Ideal point discriminant analysis (DA) which we propose to use in this paper, on the other hand, combines the best of the two approaches. It provides a spatial representation of row and column categories, and it allows statistical evaluation of various structural hypotheses about the contingency tables.

In the next section (section 2) we briefly describe ideal point DA. In section 3 we apply ideal point DA to Maxwell's data, and observe what it can do in the analysis of contingency tables. A variety of structural assumptions that may be incorporated will be systematically investigated in section 4. Some interesting cases will be illustrated through examples in the following section (section 5). In section 6 we discuss the relationship between ideal point DA and other methods for analysis of contingency tables. In particular, we focus our attention to Goodman's (1981) RC association model and dual scaling

(correspondence analysis), and compare goodness of fit (GOF) of these models against that of ideal point DA.

## 2. Ideal Point Discriminant Analysis (DA)

Ideal point DA was originally developed for discriminant analysis with mixed measurement level predictor variables (Takane & Shibayama, 1984; Takane, Bozdogan & Shibayama, 1987). It is based on the following three basic assumptions:

1. Each of  $R$  subjects (or other observation units) in DA is represented as a point in a multidimensional euclidean space. Coordinates of the subject points are given by a linear combination of predictor variables (and/or their transformations) characterizing the subject.
2. Each of  $C$  criterion groups has an "ideal point" represented in the same multidimensional space. The ideal point is supposed to represent a prototype of the criterion group.
3. The probability of a particular subject belonging to a particular criterion group is a decreasing function of the distance between the corresponding points. It is also an increasing function of the prior probability (or more generally, the bias) of the criterion group.

Analysis of contingency tables is viewed as a discrimination problem of column categories based on the information about row categories. Thus, the column categories assume the role of criterion groups and the row categories the role of subjects in DA.

Let  $F$  denote an  $R$  by  $C$  contingency table with its  $i, j$ -th element denoted by  $f_{ij}$ , and let  $Y$  be the  $R$  by  $A$  matrix of coordinates of the row categories, where  $A$  is the dimensionality of the representation space. In accordance with assumption 1 above, we assume that  $Y$  is a linear function of "predictor" variables. Assume there are  $p$  predictor variables, which are contained in an  $R$  by  $p$  matrix,  $X$ . Then the linear model states that

$$Y = XB, \quad (1)$$

where  $B$  is a  $p$  by  $A$  matrix of weights to be estimated. We impose different restrictions on relevant portions of  $B$  to reflect different scale types of predictor variables. (These restrictions will be discussed in section 4.1).

The predictor variables carry information about a structure of row categories which we may wish to incorporate. However, when there is no obvious row structure, we may simply set  $X = I$  (identity matrix), and obtain an unconstrained representation in which no special relationships are assumed among the row categories. We assume that discrete predictor variables in  $X$  have been coded into dummy variables, and continuous variables appropriately centered.

The dimensionality of the representation space is at most  $\min(R - 1, C - 1)$ , which implies that there should be at least two rows and two columns in the table to be analyzed. The maximum dimensionality is also bounded above by the number of nonredundant predictor variables,  $p^*$ . This is usually equal to the total number of categories in unordered categorical variables, minus the number of the unordered categorical variables, plus the number of ordered categorical and interval variables.

Let  $M$  denote the  $C$  by  $A$  matrix of coordinates of column categories. We assume, in most cases, that this  $M$  is given by weighted centroids of  $Y$ ; that is,

$$M = D_c^{-1}F'Y = D_c^{-1}F'XB, \quad (2)$$

where  $D_c$  is the diagonal matrix of column totals of  $F$ . This is called the centroid restriction. The same restriction is tacitly imposed in dual scaling. While this restriction is

not absolutely necessary, it reduces the number of parameters to be estimated. It also helps avoid certain degeneracies often encountered in nonmetric multidimensional unfolding analysis (Heiser, 1981). In section 6 we present some empirical evidence indicating that (2) is not at all unrealistic.

The squared euclidean distance ( $d_{ij}^2$ ) between the  $i$ -th row category and the  $j$ -th column category is now stated as

$$d_{ij}^2 = (\mathbf{y}_i - \mathbf{m}_j)'(\mathbf{y}_i - \mathbf{m}_j), \quad (3)$$

where  $\mathbf{y}_i$  and  $\mathbf{m}_j$  are column vectors of the  $i$ -th row of  $\mathbf{Y}$  and the  $j$ -th row of  $\mathbf{M}$ , respectively. The simple euclidean model (3) tacitly assumes that the distance is measured in the same way across all rows and columns. However, it may be generalized into

$$d_{ij}^2 = (\mathbf{y}_i - \mathbf{m}_j)' \mathbf{H} (\mathbf{y}_i - \mathbf{m}_j), \quad (4)$$

where  $\mathbf{H}$  is a metric matrix, which may depend on rows or columns. The metric matrix controls how we measure the distance between two points. Being analogous to the inverse of covariance matrices, different  $\mathbf{H}$ 's reflect different sensitivities to coordinate differences. In DA a group dependent  $\mathbf{H}$  has an obvious application in quadratic discrimination. It also has an important implication in attempts to incorporate dispersion parameters in the exponential family of distributions (e.g., Efron, 1986; Jorgensen, in press). This differential sensitivity plays a particularly important role in psychology. In pair comparison modeling it is known as differential comparability (Takane, 1980), in test theory it represents differential item discriminability, and in multidimensional scaling it gives rise to an individual differences metric of Carroll and Chang (1970).

Let  $p_{j|i}$  denote the conditional probability of the  $j$ -th column given the  $i$ -th row. We assume that this conditional probability is given by

$$p_{j|i} = \frac{w_j \exp(-d_{ij}^2)}{\sum_k w_k \exp(-d_{ik}^2)}, \quad (5)$$

where  $w_j$  is the bias parameter (to be estimated) for the  $j$ -th column category. The bias parameter may be interpreted as the marginal effect of a certain column, and it represents an overall likelihood of the column (independent of specific rows). We require  $\sum_j w_j = 1$  in order to remove scale indeterminacy in  $w_j$ . Note that  $p_{j|i}$  is proportional to  $w_j \exp(-d_{ij}^2)$  for each  $i$ , so that it is indeed an increasing function of  $w_j$  and a decreasing function of  $d_{ij}$  as required in (c) of the basic assumptions of ideal point DA given earlier. Some justifications for the specific form of the model have been given in Takane, et al. (1987). An important consideration is that (5) is justified under the general exponential family of distributions on the predictor variables. (See Takane, et al. for more detail.)

Finally, we define the (conditional) likelihood of the total set of observations as

$$L = \prod_{i,j} (p_{j|i})^{f_{ij}}, \quad (6)$$

where  $f_{ij}$ ,  $i, j$ -th element of  $\mathbf{F}$ , is the observed frequency of the  $i$ -th row and the  $j$ -th column categories. The log of the conditional likelihood is maximized with respect to  $\mathbf{B}$  and  $w_j$ . We use Fisher's scoring algorithm for the maximization.

Note that we use the conditional probability in defining the likelihood, (6). The marginal probability,  $p_i$ , of row  $i$  is conditioned out, partly because in DA our primary interest is to predict  $j$  based on  $i$ , so that the marginal probability of row  $i$ , being common to all columns, falls out. This conditional estimation is analogous to taking odds rather than working directly on observed frequencies in loglinear analysis of contingency tables.

Some general properties of conditional maximum likelihood estimation are given in Andersen (1980).

Note that we could have employed

$$p_{ij} = uv_i w_j \exp(-d_{ij}^2), \quad (7)$$

for joint probability,  $p_{ij}$ , and the full (joint) likelihood,

$$L = \prod_{i,j} (p_{ij})^{f_{ij}}, \quad (8)$$

where  $v_i$  ( $i = 1, \dots, C$ ) in (7) is the row effect parameter and  $u$  the overall scale parameter. However, this leads to more parameters to be estimated (due to  $v_i$ ), and an increased accuracy in parameter estimates is expected to be negligible. (The latter may be seen by noting the fact that so much of the information contained in  $f_i$ , the marginal frequency of row  $i$ , is "absorbed" by  $v_i$ , and little is left for the estimation of parameters in  $p_{j|i}$ .) We have, therefore, decided to maintain the conditional maximum likelihood estimation described earlier.

Once the maximum (conditional) likelihood,  $L^*$ , is obtained, AIC (Akaike, 1974) is calculated according to

$$\text{AIC} = -2 \ln L^* + 2n_p, \quad (9)$$

and is used for identifying the best fitting model. The effective number of parameters,  $n_p$ , in the above formula is given by

$$n_p = p^*A - \frac{A(A-1)}{2} + (C-1),$$

where  $p^*$  is the number of nonredundant predictor variables. The AIC defined above plays an important role in the evaluation of various structural hypotheses about the contingency tables. A variety of structural hypotheses or constraints (to be discussed in Section 4) may be incorporated in the representation of row and column categories, and their empirical validity may be tested through the minimum AIC procedure.

### 3. An Illustrative Example: Maxwell's Data

Within the basic framework of the method presented in the previous section, a variety of interesting analyses are possible. We explore some of the possibilities here, using Maxwell's data as an example.

Let us first look at Figure 1. This is the unconstrained solution obtained by setting  $\mathbf{X} = \mathbf{I}$ . In this case the configuration was derived entirely on the basis of joint frequencies of row and column categories. In Figure 1 the 16 predictor patterns are indicated by numbered dots and the three criterion groups by asterisks.

Figure 2 depicts a constrained two-dimensional solution obtained by specifying the main effects of the four symptom variables as the predictor variables. Coordinates of the 16 predictor patterns were constrained to be a linear function of the predictor variables. We see some regularity or parallelism in the configuration of the 16 patterns. This is because additivity of the four predictor variables is now assumed, so that, for example, the location of pattern 10 with the predictor pattern of (1001), presence of the first and the last symptoms and absence of the second and the third symptoms, is obtained by the sum of four vectors designated as 1-1 (presence of the first symptom), 2-0 (absence of the second symptom), 3-0 and 4-1. Note that two vectors corresponding to presence and absence of a symptom (e.g., 1-1 & 1-0) are pointing in opposite directions, their lengths

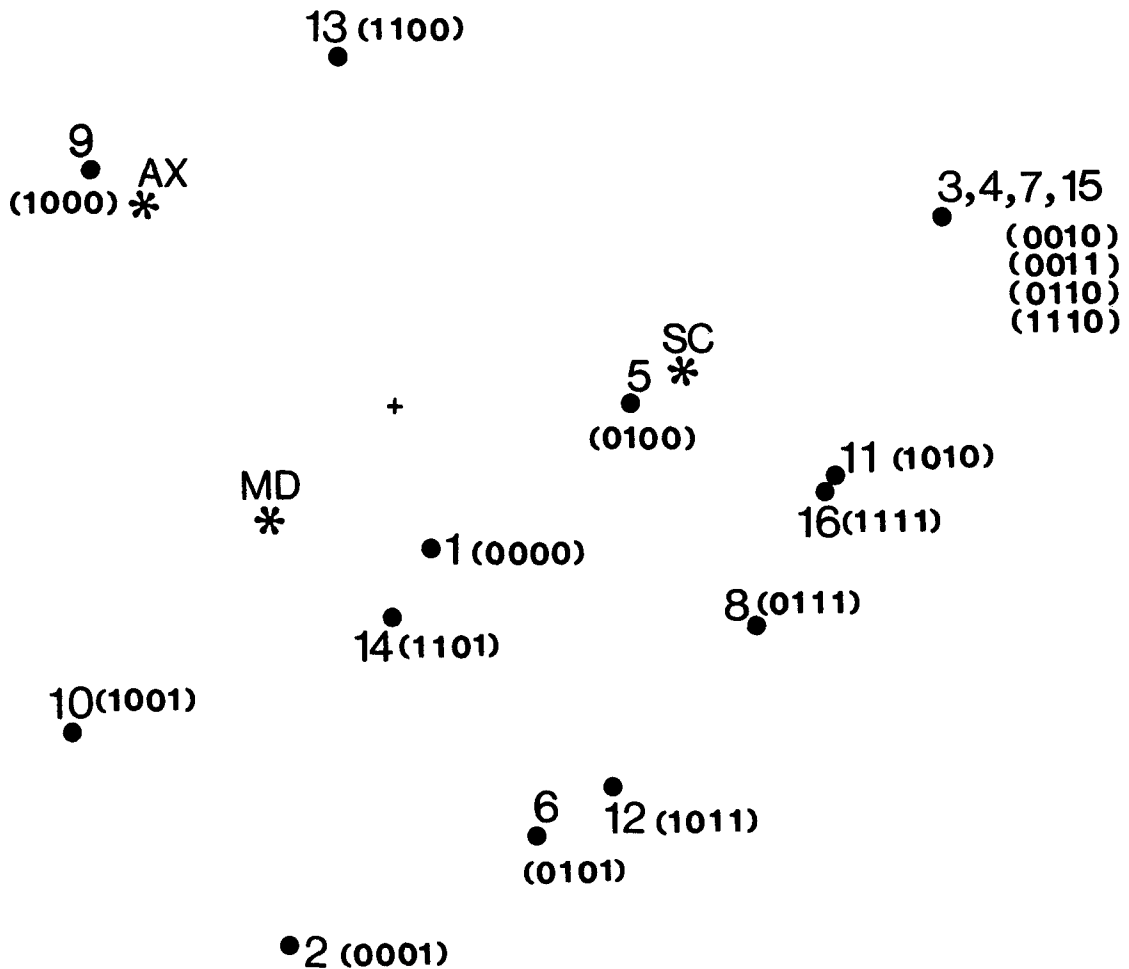


FIGURE 1.

being proportional to the reciprocal of their marginal frequencies. This is due to the binary nature of the predictor variables.

An important consequence of a specific row structure like additivity is that a point can be located in the configuration (*albeit* less reliably), even if information is missing for some of the symptoms. For example, the location of pattern (0 0 0 ?) marked by a star, the fourth symptom information missing, is obtained by adding only three vectors corresponding to 1-0, 2-0 and 3-0.

The constrained solution in Figure 2 uses only nine parameters, while the unconstrained solution in Figure 1 uses 31. Table 2 gives the AIC values corresponding to the two solutions, which are 841 and 869 for the constrained and the unconstrained solutions, respectively and the constrained solution turns out to be a much better solution than its unconstrained counterpart. Comparing the two configurations, we notice that relative locations of several points, namely 4, 8, 11, 12 and 15, differ quite radically from one configuration to the other. This means that these points could change their locations in Figure 1 without much effect on the value of the overall likelihood. That is, they are rather poorly estimated in the unconstrained solution. This was confirmed by other evidence; I tried to draw 95% asymptotic confidence regions (Ramsay, 1978) for points in

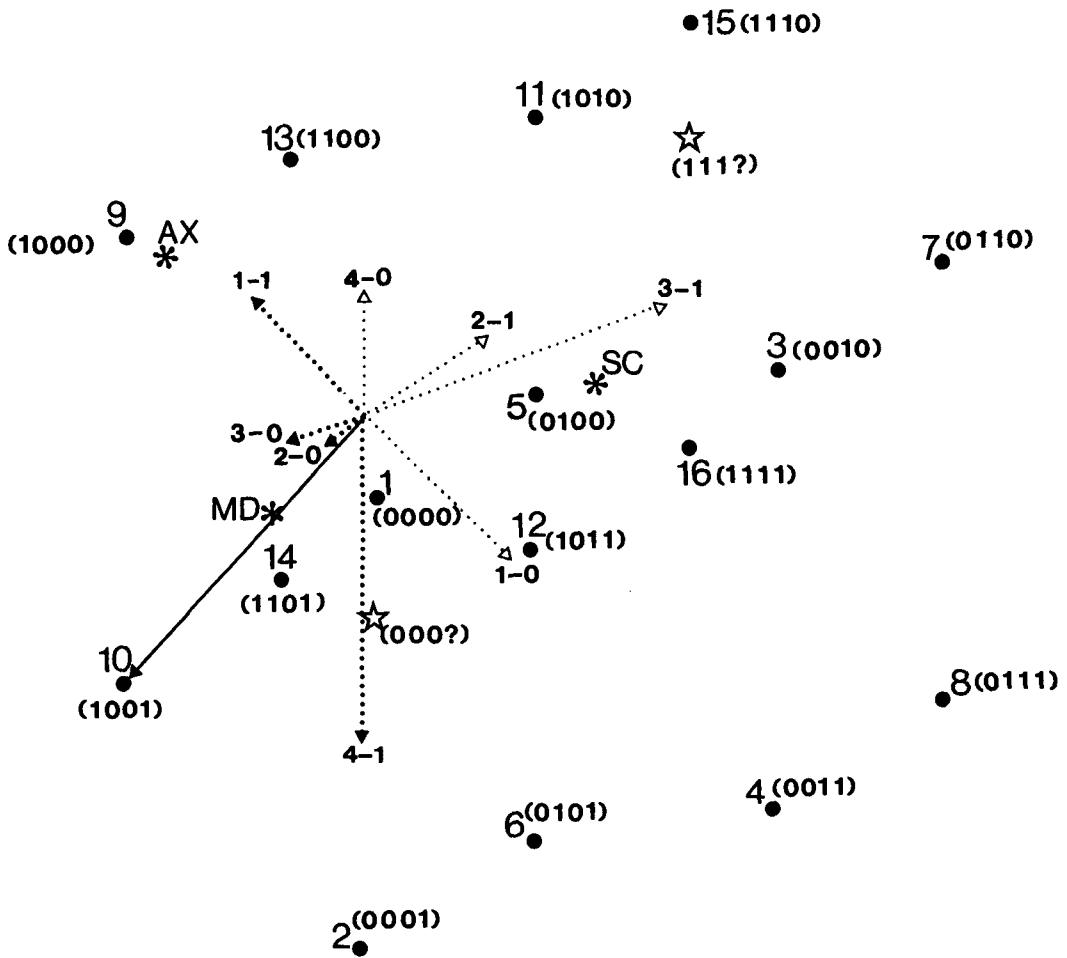


FIGURE 2.

the unconstrained solution, but they turned out to be much too big to fit in the figure. However, the points are much more tightly estimated in the constrained solution. This suggests an importance of having proper constraints for reliable estimation of model parameters, particularly when the data are weak in the sense that there are many low frequency cells.

We may ask other questions about Maxwell's data. For example, in the above analyses the dimensionality of the representation space was assumed to be two, but are the two dimensions really necessary? The main effects of the four symptom variables were used as the predictor variables, but do they provide the best structure for the row categories? Are all of them significant? Aren't there any significant interactions among them? Also, the three criterion groups were assumed to be all distinct, but are they really? If not, which groups are distinct and/or not distinct from which groups?

These questions may be answered by model evaluation capabilities of ideal point DA. The best dimensionality of the representation space is determined by obtaining solutions in different dimensionalities and by comparing their fits to the data through AIC. The optimal set of predictor variables may be determined by subset selection of the predictor variables. That is, by deleting variables and/or adding interactions, and by observing how

Table 2. Summary statistics for Maxwell's

	K++	Dim	AIC (# of para.)	Lower Bound for AIC
Saturated Model	3		855 (32)	
Ideal Point DA				
[1] 1,2,3				
Four main effects	3	2	841* (9)	
Variable 1 deleted	3	2	945 (7)	
Variable 2 deleted	3	2	871 (7)	
Variable 3 deleted	3	2	1013 (7)	
Variable 4 deleted	3	2	915 (7)	
2-way interactions added	3	2	853 (21)	
3-way interactions added	3	2	866 (29)	
4-way interactions added	3	2	869 (31)	
(Equivalent to the Unconstrained Solution)	3	1	918 (17)	888 (2)+
[2] (1,2),3	2	1	1121 (17)	1091 (2)+
[3] 1,(2,3)	2	1	970 (6)	940 (2)+
[4] (1,3),2	2	1	1098 (17)	1068 (2)+
[5] (1,2,3) (Equivalent to the Null Model)	1	0	1296 (2)	1296 (2)+

++ K: the number of distinct criterion groups assumed in each clustering alternative

+ The cases in which the lower bound for the AIC value that can be possibly attained by subset selection is larger than the minimum AIC for Case [1], dim=2.

\* Minimum AIC

the goodness of fit of the model changes as a result. Multisample cluster analysis (Takane et al., 1987) may be used to identify distinct and nondistinct subsets of criterion groups. This is done by constraining the coordinates of some of the criterion groups (presumed to be nondistinct) to be equal, and by comparing the constrained fit against that of the original model.

We have applied the above procedures to Maxwell's data rather systematically and the results are reported in Table 2. With three criterion groups there are five possible clustering alternatives labeled by bracketed numbers in the table. For each case the numbers enclosed in parentheses indicate indices of the criterion groups presumed to be nondistinct.

In Case 1 all the three criterion groups are assumed distinct. In this case both one- and two-dimensional solutions can be obtained, and within each dimensionality an optimal subset of predictor variables may be identified. We first obtained the unconstrained one-dimensional solution. The AIC value of 918 associated with this solution is found to be much larger than that of the constrained two-dimensional solution discussed earlier. We then calculated the lower bound for the AIC value that can be possibly attained by



subset selection of predictor variables in the one-dimensional solution. This value is obtained by assuming that the log likelihood attained by the unconstrained one-dimensional solution is maintained, even if all the predictor variables were eliminated. As it has turned out, this lower bound for the AIC value also exceeds the AIC value of the constrained two-dimensional solution. This implies that no one-dimensional solutions can possibly fit the data better than the constrained two-dimensional solution. That is, the best dimensionality of the representation space is two.

The above fact implies that there are three distinct groups. To see this is indeed the case, we obtained the unconstrained solutions for Cases 2, 3, and 4, in which two of the three criterion groups were in turn assumed nondistinct. Again the lower bounds for the AIC value (that could be achieved by subset selection) were calculated for these cases, but they all exceed the AIC value of the constrained two-dimensional solution in Case 1, which is so far the best solution. The AIC value for Case 5 also exceeds that of the constrained two-dimensional solution. This confirms the distinctiveness of the three criterion groups in Maxwell's data.

Given that there are three distinct criterion groups to be represented in a two-dimensional euclidean space, we may now apply a subset selection procedure to identify the optimal set of predictor variables that provide the best row structure for Maxwell's data. The constrained solution in Figure 2 used the main effects of the four symptom variables as the predictor variables. We eliminated each of the predictor variables in turn from this model, and fitted the reduced models. None of the reduced models fit the data better than the original model according to the minimum AIC criterion. This implies that all the four main effects are significant for discrimination.

We then added interaction terms among the symptom variables to the main effects model. There are 6 two-way interactions, 4 three-way interactions, and 1 four-way interaction that can be defined among the four variables. In order to reduce the number of possible models to be fitted we added all the interactions of a same order simultaneously and incrementally from the lowest order interactions. Again none of the incremented models were found to provide a better fit than the main effects model.

The constrained two-dimensional solution depicted in Figure 2 is indeed the best solution obtained. The four main effects of the original predictor variables provide the best row structure for Maxwell's data. This roughly corresponds with the condition under which Maxwell's data were generated. The above analysis thus confirms the ability of ideal point DA and the minimum AIC criterion to detect and recover the original "true" structure.

#### 4. A Variety of Model Specifications

A variety of specialized models may be specified and subjected to model comparisons. We have already seen some of them in the previous section. In this section we conduct a more systematic survey of possible model specifications. A few particularly interesting ones will be highlighted through examples in the next section. The specifications are divided into three groups: (a) direct constraints on model parameters, (b) row structures, and (c) column structures.

There are two particularly important benchmark models worth mentioning. They serve as an upper bound (the saturated model) and a lower bound (the null model) for all other models. In the saturated model no structural assumptions are made, so that  $p_{j|i}$  is estimated by  $f_{ij}/f_i$ . The null model, on the other hand, assumes independence of rows and columns (rows have no predictive power on columns), which amounts to assuming that  $p_{j|i} = p_j$  for all  $i$ . The goodness of fit (GOF) of specialized models are to be compared against that of the benchmark models.

#### 4.1. Direct Constraints on Model Parameters

There are two kinds of model parameters, weight matrix  $\mathbf{B}$  and bias parameters  $w_j$ ,  $j = 1, \dots, C$ . Constraints on  $w_j$  are relatively simple; the bias parameters are either left free, or fixed at  $w_j = 1/C$  for all  $j$  or at some other externally supplied values. In some cases we may allow different sets of  $w_j$  to apply different subsets of rows. This is useful, for example, when we manipulate  $w_j$  without changing  $d_{ij}$ , as in a signal detection experiment. Constraints on  $\mathbf{B}$  are more varied and need some explanation. They may be classified into several groups: (1) the scale type restriction, (2) the dimensionality restriction, (3) the zero restriction, (4) the equality restriction and (5) the multiplicative decomposition of weights for unordered categorical variables.

##### (1) The Scale Type Restriction

As alluded to earlier, different types of constraints reflecting different scale levels of the predictor variables are imposed on relevant portions of  $\mathbf{B}$ . We summarize these restrictions as follows.

(a) *Unordered categorical variable.* If variable  $k$  is unordered categorical, we find multidimensional quantifications (weights) of categories subject to the centering restriction,

$$\sum_s f_{k(s)} b_{k(s)a} = 0 \quad \text{for all } a, \quad (10)$$

where  $f_{k(s)}$  is the marginal frequency of category  $s$  in variable  $k$ , and  $b_{k(s)a}$  is the weight to be applied to the category on dimension  $a$ . The summation extends over all categories in variable  $k$ . The same kind of restriction is also implicit in dual scaling.

(b) *Ordered categorical variable.* It is not possible to find two orthogonal sets of quantifications satisfying the same prescribed order. We therefore seek to find only one set of quantifications subject to an order restriction and a centering restriction similar to (10). These unidimensional quantifications are multidimensionally weighted to obtain multidimensional quantifications.

(c) *Interval variable.* In this case we assume that category quantifications are already given, and obtain only dimensional weights.

##### (2) The Dimensionality Restriction

Each analysis by ideal point DA presupposes a specification of dimensionality  $A$  in the representation space. This  $A$  usually corresponds to the number of columns in matrix  $\mathbf{B}$ . (More precisely it corresponds to the rank of  $\mathbf{B}$ .) As has been demonstrated in the previous section the best dimensionality of the representation space can be identified by comparing the GOF of solutions obtained in different dimensionalities.

##### (3) The Zero Restrictions

The zero restriction, that is, to set certain elements of  $\mathbf{B}$  to zero, plays two important roles in ideal point DA. First, we may use it to test the significance of a contribution of some predictor variable on certain dimensions. Secondly, we may force certain dimensions of the representation space to coincide with some specific predictor variables. This is useful to make ideal point DA completely equivalent to the loglinear model.

##### (4) The Equality Restriction

The equality restriction may be imposed to test the equality of contributions of two or more predictor variables on certain dimensions. Both the zero restriction and the

equality restriction imposed directly on  $\mathbf{B}$  are more general than analogous restrictions imposed through manipulations of predictor variables (to be explained in section 4.2) in that they can be imposed separately for each dimension.

#### (5) *Multiplicative Decomposition*

Multidimensional quantifications of categories usually obtained for unordered categorical variables may be decomposed into the product of a single set of category quantifications and dimensional weights. This is similar to the scale type restriction for ordered categorical variables, but without the order restriction. Again the empirical validity of this assumption may be tested through model comparisons.

### 4.2 *Row Structures*

This class of restrictions are perhaps the most important ones in ideal point DA. The restrictions representing certain row structures are usually incorporated via predictor variables. It is thus crucial to be able to define appropriate predictor variables for specific purposes of analysis.

We have already mentioned that the matrix of predictor variables may be set to an identity matrix if row categories are to be left unconstrained. A particular row structure model will be tested against the unconstrained case, whose  $\mathbf{X}$  is set to an identity matrix, by specifying that there is only one unordered categorical variable with as many categories as there are rows in a contingency table. Row structures most often take the form of an additive function of several predictor variables. This however, does not exclude the possibility of having a single predictor variable. Interactions among the original predictor variables may be defined and included in the predictor set. Additivity (or no interaction hypothesis) of the predictor variables may be tested against specific interaction models. Takane, et al., (1987) describe how the interactions are to be defined in ideal point DA.

Other specialized comparisons or contrasts may be specified, including polynomials or trends of different degrees. This is important, for example, in a psychophysical study where row categories represent some kind of physical stimuli, and we would like to find out how well physical attributes of the stimuli can explain subjects' responses.

Other designs may be meaningful. For example, in pair comparison data we provide a design matrix which takes the difference between scale values of compared stimuli. The design matrix for pair comparisons may be combined with other stimulus information to form a new design matrix.

Certain effects may be partialled out from the predictor set. This is done in a manner similar to that in the usual linear regression. Suppose we would like to eliminate the effects of variables  $\mathbf{Z}$  from  $\mathbf{X}$ . Then we define  $\mathbf{X}_z = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{X}$  and use it as the predictor variables. This matrix represents the portion of  $\mathbf{X}$  that cannot be accounted for by  $\mathbf{Z}$ . If we set  $\mathbf{X} = \mathbf{I}$ , we obtain  $\mathbf{I}_z = \mathbf{X}_z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ . This represents the variation accounted for by the unconstrained model, but left unaccounted for by the model in which we have  $\mathbf{Z}$  as the predictor variables. In this particular case  $\mathbf{I}_z$  is singular. A minimal set of linearly independent predictor variables may be obtained by the set of eigenvectors of  $\mathbf{I}_z$  corresponding to unit eigenvalues. To be more consistent with correspondence analysis we may introduce a metric matrix,  $\mathbf{D}_z$ , which is a diagonal matrix of row marginals, and redefine  $\mathbf{I}_z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{D}_z\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}_z$ . Attempts have been made to partial out certain effects in correspondence analysis (Nishisato, 1982; ter Braak, in press; van der Heijden & de Leeuw, 1985; Yanai, 1987), and our development is analogous to these developments.

The effect of a particular variable may be tested by deleting that variable from the predictor set. This tests the overall contribution of the variable, rather than its dimension-

al contribution. The test of equality of certain rows may be considered as a special case of the test of the overall contribution of a predictor variable. The overall equality of contributions of two or more variables may be tested by creating a new variable by summing the variables whose effects are presumed to be equal.

In all the above cases subset selection of predictor variables plays an important role. The number of parameters, and consequently the number of predictor variables, should be kept as small as possible to ensure reliable estimation. The best predictor set is supposed to provide the best structure for row categories.

### 4.3 Column Structures

The centroid restriction, (2), is usually imposed on columns. Other structures similar to those on rows may also be imposed on the columns. For example, the column categories may have a factorial structure, and then additivity analysis may be performed on the column categories. In section 5.1 we will discuss an example of this and explain how this type of restriction can be incorporated.

In certain cases coordinates of column categories may be fixed at prescribed values. This is useful, for example, for analysis of pair comparison data. For certain square matrices in which rows and columns have certain correspondence, for example, stimulus recognition data,  $\mathbf{M}$  may be set equal to  $\mathbf{Y}$ . This will make ideal point DA a model of quasisymmetry (Caussinus, 1965).

Equality restrictions may be imposed on certain columns, allowing the test of equality among the columns. Statistically nondistinct columns may be combined into one. Equality restrictions may also be used to impose a prescribed order among columns. Those columns in an order violation are simply combined into one. When the test of equality among column categories is systematically applied, it is called multisample cluster analysis (CA). This allows multiple comparisons of the column categories in the context of ideal point DA. Takane, et al. (1987) discuss how multisample CA may be performed in ideal point DA. It involves prior combinations of column categories (assumed nondistinct) and a post adjustment of the value of the maximum log likelihood. Equality restrictions may also be directly imposed on  $\mathbf{M}$  in a manner similar to the way an additivity hypothesis is incorporated.

## 5. More Examples

In this section we present a few more examples to demonstrate some of the interesting model specifications discussed in the previous section.

### 5.1 *The ISM Data on the Traditional Versus Modern Views*

This data set was taken from part of a large scale survey on Japanese nationality conducted at the Institute of Statistical Mathematics in Tokyo. The current data set pertains to the survey conducted in 1973. The following two questions on traditional versus modern views were selected, and six categories were formed by factorially combining three answer alternatives in the first question and two in the second question. The six categories constitute the criterion groups.

Q1. If you have no children, do you think it is necessary to adopt a child in order to continue the family line?

- A1. Would adopt
- A2. Would not adopt
- A3. Depends on circumstances

Q2. Some people say that if we get good political leaders, the best way to improve the country is for the people to leave everything to them.

- B1. Agree  
B2. Disagree

In addition three background variables about the subjects were provided: sex (1.male 2. female), age groups (1. 20–30 2. 30–40 4. 40–50 5. 50–60 6. >60) and levels of education (1. elementary and junior high, 2. senior high and 3. college). These background variables were used as predictor variables. Over 2,000 subjects were cross-classified according to the six response categories and the three background variables. We were interested in testing the additivity of the background variables (for a row structure) as well as the additivity of columns implied by the factorial nature of the response categories.

The additivity hypothesis on columns may be incorporated as follows. Define the design matrix,  $G$ , for the columns. This matrix may look like

$$G = \begin{bmatrix} 1 & 0 & 0 & \vdots & 1 & 0 \\ 0 & 1 & 0 & \vdots & 1 & 0 \\ 0 & 0 & 1 & \vdots & 1 & 0 \\ 1 & 0 & 0 & \vdots & 0 & 1 \\ 0 & 1 & 0 & \vdots & 0 & 1 \\ 0 & 0 & 1 & \vdots & 0 & 1 \end{bmatrix}$$

Rows of  $G$  correspond to the six response categories, while columns to the five answer alternatives in the original questions (Q1 & Q2). Matrix  $G$  indicates which response alternatives in the original questions constitute the derived response categories. Let

$$F^* = FG.$$

Then under the additivity hypothesis  $M$  is defined as

$$M = GD_c^{*-1}F^*Y$$

where  $D_c^*$  is the column total of  $F^*$ .

We analyzed the data under the row additivity hypothesis only, and then under both the row and column additivity hypotheses. Circular dots in Figure 3 represent the column points obtained under the row additivity hypothesis only. Points associated with A1, A3, and A2 were connected by a solid line separately for B1 and B2. The two lines are not parallel. Squares in Figure 3 represent the column points obtained under the row as well as the column additivity hypothesis, and are similarly connected by dotted lines, which are now parallel, because under the column additivity hypothesis the point corresponding to (A1, B2), for example, is obtained by the sum of two vectors designated as A1 and B2. This situation is analogous to how row points are obtained under a row additivity hypothesis.

The GOF of various models fitted are reported in Table 3. The model with only the row additivity hypothesis has turned out to be the best fitting model.

## 5.2 Guilford's Data by the Constant Method

Ideal point DA has interesting relationships to some of the traditional psychometric methods. In this section we apply ideal point DA to two data sets collected by the constant method. In this method a single stimulus is presented to the subject in each trial, and the subject is usually required to make a two-category judgment about the stimulus

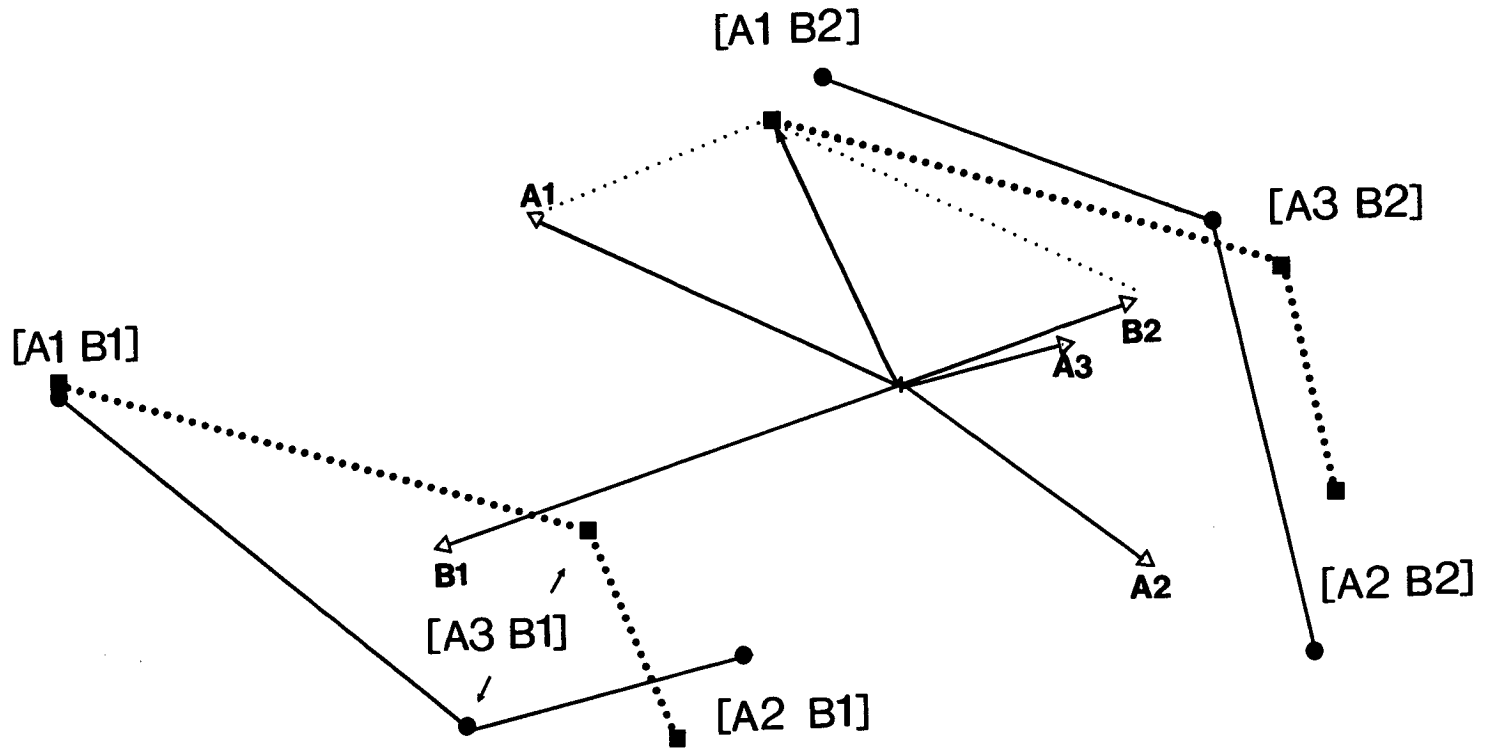


FIGURE 3.

Table 3. Summary statistics for the ISM data on traditional vs. modern views

	AIC (# of para.)
Saturated model	6511 (140)
No structures on rows	6399 (58)
Rows constrained by sex, age and education	6364* (18)
Rows constrained as above plus additivity in column	6381 (18)
Null model	6728 (5)

\* Minimum AIC

according to some criterion (See data set (A) below). However, in some cases judgments are made in more than two response categories (See data set (B) below).

The two data sets were taken from Guilford (1954; p. 150). In data set (A), each of five stimuli (two pointed objects applied to the skin and separated by a distance varying from 18 mm to 34 mm in steps of 4 mm) was judged whether it was a "one object" or "two objects" one hundred times. In data set (B), each of seven comparison stimuli (weights ranging from 88 g to 106 g in steps of 3 g) was compared with a standard weight of 100 g and a judgment was made as to whether the comparison stimulus was heavier than, doubtful or lighter than the standard stimulus. Again this was done one hundred times. In both cases we are interested in how well the physical attribute of the stimuli can account for the data. This was investigated by incorporating the physical measurement of the stimuli as the predictor variable.

Table 4 summarizes the GOF of various models fitted to the two data sets. For data set (A) we can obtain only one-dimensional solutions. Notice that with two response categories the unconstrained solution is identical to the saturated model, and this is generally true. A squared physical separation term was also included in one model, but the model in which the stimulus representation was constrained by a linear function of the distance is found to be the best fitting model.

Results for data set (B) are similar. A major difference is that for this data set (having three response categories) we can obtain two-dimensional as well as one-dimensional solutions. As it turned out, the constrained one-dimensional solution with a linear function of the physical attribute as the predictor variable is again found to be the best fitting model. The three response categories are arranged in the expected order in the derived unidimensional space although no order restrictions were explicitly imposed on the column categories.

When there are only two response categories, it can be shown that ideal point DA has a simple relationship with the logit analysis of the constant method. In this particular case  $A = 1$  and

$$d_{ij}^2 = (y_i - m_j)^2, \quad j = 1, 2,$$

Table 4. Summary statistics for Guilford's data

Model	Dimension	(A)	Data	(B)
Saturated model		474 (5)		1097 (14)
No structures on rows	2			1098 (13)
	1	474 (5)		1091 (8)
Linear in physical attribute	1	472*(2)		1083* (3)
Linear and quadratic in physical attribute	2			1084 (5)
	1	473 (3)		1084 (4)

\*Minimum AIC

(where  $y_i$  may be further constrained by  $x_i'b$ ), so that

$$p_{1|i} = [1 + \exp(-q_i)]^{-1},$$

where

$$\begin{aligned} q_i &= 2(m_1 - m_2)y_i - (m_1 - m_2)(m_1 + m_2) - \ln\left(\frac{w_1}{w_2}\right) \\ &= ay_i + c, \end{aligned}$$

where

$$a = 2(m_1 - m_2),$$

and

$$c = -(m_1 - m_2)(m_1 + m_2) - \ln\left(\frac{w_1}{w_2}\right).$$

Note that  $q_i$  is linear in  $y_i$ , and is also linear in  $x_i$ . Coefficient  $a$  is a sensitivity parameter (analogous to  $d'$  in signal detection theory) and  $c$  a bias (or threshold) parameter. Parameter  $c$  may also be interpreted as an order effect parameter. In the simplest case we may set  $a = 1$  ( $m_2 = -m_1 = 1/4$ ) and  $c = 0$  ( $w_1 = w_2$ ) in  $q_i$ .

Ideal point DA has interesting relationships to other psychometric methods such as pair comparison models (Bradley & Terry, 1952; Colonius, 1981; Luce, 1959; Strauss, 1981), successive categories scaling (e.g., Torgerson, 1958), signal detection theory (e.g., Green & Swets, 1966), information theory (e.g., Krippendorff, 1986; McGill, 1954), models of stimulus recognition data (e.g., Takane & Shibayama, 1986), the method of similar reactions (Thurstone, 1929; Torgerson, pp. 395-402), and so forth. However, this topic requires a separate treatment.



## 6. Relations to Other Methods

Ideal point DA may be viewed as a special form of unfolding analysis (Coombs, 1964). In ecology a similar method called Gaussian ordination (Ihm & van Groenewoud, 1975) has been in use for quite some time. Some attempts have also been made to establish the relationship between unfolding analysis, Gaussian ordination, and correspondence analysis (Escoufier, 1987; Heiser, 1981, 1987; Ihm & van Groenewoud, 1984; ter Braak, 1986). Ideal point DA has many other relatives in statistics. Its relation to loglinear DA (e.g., Andersen, 1980), logistic discrimination (e.g., Anderson, 1982) and the generalized linear models (McCullagh & Nelder, 1983) has already been discussed in Takane, et al. (1987). In this section we discuss two more methods bearing interesting relationships to ideal point DA, namely Goodman's RC association model and dual scaling (correspondence analysis).

We may rewrite (8) as

$$\begin{aligned} p_{ij} &= uv_i w_j \exp(-d_{ij}^2) \\ &= u^* v_i^* w_j^* \exp(\mathbf{y}_i^{*'} \mathbf{m}_j^*) \end{aligned} \quad (11)$$

by adjusting scales of  $u^*$ ,  $v_i^*$ ,  $w_j^*$ ,  $\mathbf{y}_i^*$  and  $\mathbf{m}_j^*$ . Squared terms in the distance model are absorbed in row and column effect parameters,  $v_i^*$  and  $w_j^*$ . If no centroid restriction is imposed on columns, (11) becomes Goodman's (1981) RC association model. The difference between the two models is rather critical. The distance interpretation no longer holds for the RC association model. Later in this section we demonstrate the importance of the centroid restriction.

Note that (11) leads to

$$p_{j|i} = \frac{w_j^* \exp(\mathbf{y}_i^{*'} \mathbf{m}_j^*)}{\sum_k w_k^* \exp(\mathbf{y}_i^{*'} \mathbf{m}_k^*)},$$

which may be called the conditional probability representation of the RC association model. This form of the RC association model can be also derived from (5), the conditional probability form of ideal point DA.

We may also rewrite (8) as

$$\ln L = fu^* + \sum_i f_i v_i^* + \sum_j f_j w_j^* - \sum_{i,j} f_{ij} d_{ij}^2, \quad (12)$$

where  $u^*$ ,  $v_i^*$  and  $w_j^*$  are log transformed overall scale, row effect and column effect parameters, and  $f$ ,  $f_i$  and  $f_j$  are total, row marginal and column marginal frequencies, respectively. (A similar line of argument follows if (6), instead of (8), is used.) Obviously the last term in (12) is the most important term for the estimation of parameters in the distance model. Minimizing  $\sum f_{ij} d_{ij}^2$  under appropriate ortho-normalization restrictions leads to dual scaling (correspondence analysis), as has been shown by Heiser (1981). Consequently, ideal point DA (in the unconstrained case) is expected to give similar results to correspondence analysis. With constraints on row categories ideal point DA gives similar results to canonical correspondence analysis (ter Braak, 1986; see also Carroll, 1973; Fisher, 1948; Hayashi, 1952; Johnson, 1950; and Maxwell, 1961), which is a special form of correspondence analysis with constraints. Correspondence analysis and canonical correspondence analysis are, respectively, known as the third kind and the second kind of quantification methods in Japan (Hayashi, 1952). The latter is mathematically equivalent to canonical DA applied to dummy coded discrete predictor variables. Correspondence analysis or canonical correspondence analysis (with special provisions for

Table 5. Comparison with maximum likelihood canonical analysis and the RC association model

Data	Ideal Point DA		Canonical analysis		Association model	
	$G^2(df)$	AIC	$G^2(df)$	AIC	$G^2(df)$	AIC
Srole, et al.	3.6(12)	-20.4*	2.8(8)	-13.2	3.6(8)	-12.4
Fisher	9.4 (6)	- 2.6*	4.7(2)	0.7	8.7(5)	- 1.3
Snee	2.0 (4)	- 6.0*			3.7(4)	- 4.3
Cramer	2.8 (6)	- 9.2*			0.8(3)	- 5.2
Rasch-Andersen	3.5 (9)	-14.5*			3.5(6)	- 8.5
Caussinus	0.8 (5)	- 9.2*			0.2(2)	- 3.8

$G^2$  is the value of the log likelihood ratio chi-square from the saturated model.

\* Minimum AIC's

ordinal and interval predictor variables), as the case may be, is used to obtain initial estimates of parameters in ideal point DA. Gilula and Haberman (1986) recently proposed what they call "canonical analysis" which is a sort of maximum likelihood correspondence analysis. The relationship between correspondence analysis and the association model has been discussed in Goodman (1985), and Gilula and Haberman.

The GOF of ideal point DA was compared with that of canonical analysis and that of the RC association model. The following six data sets were used for this purpose. The first two were fitted by Gilula and Haberman (1986), and the remaining four analyzed by Goodman (1984, 1986).

1. Srole, Michael, Opler and Rennie (1962): 1,660 midtown Manhattan subjects cross-classified according to six parental socio-economic status strata and four categories of mental health status (well, mild symptom formation, moderate symptom formation, and impaired).

2. Fisher (1940): 5,387 school children in Caithness cross-classified according to four levels of eye color (blue, light, medium, and dark) and five levels of hair color (fair, red, medium, dark, and black).

3. Snee (1974): 592 subjects cross-classified according to their eye color (four levels: brown, blue, hazel, and green) and hair color (four levels: black, brunette, red, and blond).

4. Cramer (1946): 25,263 Swedish families cross-classified according to their yearly income (four levels: 0-1, 1-2, 2-3, and >3 thousand Kronor) and number of children (five levels: 0, 1, 2, 3, and  $\geq 4$ ).

5. Rasch-Andersen (Goodman, 1986): 7,025 Danish men cross-classified by year (four levels: 1955-1958) and age (five levels: 15 through 19 years) for whom initial charges have been dropped.

6. Caussinus (1986): 2,730 patients cross-classified according to five age groups (<50, 50-60, 60-70, 70-80, and >80) and four cancer types.

Table 5 shows the likelihood ratio chi-square ( $G^2$ ) representing the difference between

the saturated model and the model selected as the best fitting model by the respective methods (along with the associated degrees of freedom of the chi-square in parentheses) for the six data sets. The table is not complete, since the results from canonical analysis were obtained by Gilula and Haberman (1986) and those of the RC association model by Goodman (1985, 1986). In all cases the chosen model provides a satisfactory fit in terms of the chi-square GOF tests.

In order to make direct comparisons among the three methods, AIC was calculated by

$$\text{AIC} = G^2 - 2 \text{ df.}$$

The AIC values indicate that in all the six data sets ideal point DA provides a better fit than canonical analysis or the RC association model. The result also indicates that the centroid restriction in ideal point DA is not at all restrictive. On the contrary it has the positive effect of reducing the number of model parameters without impairing the overall GOF of the model.

The above comparison is actually somewhat unfair to ideal point DA, which is based on the conditional likelihood as opposed to the full (joint) likelihood. This is because the conditional likelihood and the marginal likelihood of rows, separately maximized and multiplied, never exceeds the joint maximum likelihood. Thus,  $G^2$  derived from the joint maximum likelihood gives a lower bound for  $G^2$  derived from the conditional maximum likelihood. Ideal point DA with the centroid restriction still fits better than the two methods.

Ideal point DA could be applied to the above data sets with rows and columns interchanged. When this was done, similar results were obtained; the advantage of ideal point DA remained intact in all the six cases.

## 7. Concluding Remarks

We have seen a variety of interesting analyses that can be performed by ideal point DA. Throughout these analyses we have emphasized various advantages of ideal point DA. It allows a spatial representation like the one obtained in dual scaling. At the same time it allows detailed model evaluation that is feasible in loglinear analysis of contingency tables. A variety of constraints may be incorporated at different levels, and can be subjected to interesting model comparisons. As we have seen, proper constraints are also important for reliable estimation of model parameters. Ideal point DA can be widely used, and subsumes a number of interesting models as special cases. Perhaps it is not as widely applicable as correspondence analysis, since it requires certain statistical assumptions of the data. However, it certainly surpasses correspondence analysis in the variety of analyses it can undertake on a same set of data. This is quite important in data analysis in general to avoid accepting just any model which happens to be fitted (de Leeuw, 1984).

There are a number of developments that could make the method even more attractive. The most important one of these is the availability of various diagnostic capabilities, for which many of the techniques (Landwehr, Pregibon & Shoemaker, 1984; Pregibon, 1981; Rubin, 1984) developed for the generalized linear model (McCullagh & Nelder, 1983) will be of great value. In particular the method should allow extensive residual analysis. Diagnostic methods are extremely important in data analysis to obtain insight for further improvements of the model.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

- Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North-Holland.
- Anderson, J. A. (1982). Logistic discrimination. In P. R. Krishnaiah, & L. N. Kanal (Eds.), *Handbook of statistics 2*. Amsterdam: North Holland.
- Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: The MIT Press.
- Bradley, R. A., & Terry, M. E. (1952). The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, *39*, 324–345.
- Carroll, J. D. (1973). Categorical conjoint measurement. In P. E. Green & Y. Wind, (Eds.), *Multiattribute decisions in marketing: A measurement approach*. Hinsdale IL: Dryden Press.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*, 238–319.
- Caussinus, H. (1965). Contribution à l'analyse statistique tableau de corrélation [Contribution to Statistical Analysis of Correlation Table]. *Annals of the Faculty of Science University*, Toulouse, *29*, 77–182.
- Caussinus, H. (1986). Discussion of paper by L. A. Goodman. *International Statistical Review*, *54*, 274–278.
- Colonius, H. (1981). A new interpretation of stochastic test models. *Psychometrika*, *46*, 223–225.
- Coomb, C. H. (1964). *A theory of data*. New York: Wiley.
- Cramer, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.
- de Leeuw, J. (1984). Beyond homogeneity analysis (RR-84-08). Leiden: University of Leiden, Department of Data Theory.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, *81*, 709–721.
- Escoufier, Y. (1987, June). In the neighborhood of correspondence analysis. Paper presented at the first IFCS Conference, Aachen.
- Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics*, *10*, 422–429.
- Fisher, R. A. (1948). *Statistical methods for research workers* (10th ed.). London: Oliver and Boyd.
- 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. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, *76*, 320–334.
- Goodman, L. A. (1985). New methods for the analysis of two-way contingency tables: An alternative to Diaconis and Efron. *The Annals of Statistics*, *13*, 887–893.
- Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243–309.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Greenacre M. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.) New York: McGraw Hill.
- Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, *2*, 69–98.
- Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Unpublished Doctoral Dissertation, University of Leiden.
- 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.
- 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.
- Johnson, P. O. (1950). The quantification of qualitative data in discriminant analysis. *Journal of the American Statistical Association*, *45*, 65–76.
- Jorgensen, B. (in press). Exponential dispersion models. *Journal of the Royal Statistical Society, Series B*.
- Krippendorff, K. (1986). *Information theory: Structural models for qualitative data*. Beverley Hills, CA: Sage Publications. (Sage University Paper Series on Quantitative Applications in the Social Sciences.)
- Landwehr, J., Pregibon, D., & Shoemaker, A. (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, *79*, 61–71. (With discussions)
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.
- Maxwell, A. E. (1961). Canonical variate analysis when the variables are dichotomous. *Educational and Psychological Measurement*, *21*, 259–271.
- McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London: Chapman and Hall.

- McGill, W. J. (1954). Multivariate information transmission. *Psychometrika*, *19*, 97–116.
- Nishisato, S. (1980). *Analysis of categorical data: dual scaling and its applications*. Toronto: University of Toronto Press.
- Nishisato, S. (1982). *Shitsu teki deta no suryoka* [Quantification of qualitative data]. Tokyo: Asakura Shoten. (In Japanese)
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, *9*, 705–724.
- Ramsay, J. O. (1978). Confidence regions for multidimensional scaling analysis. *Psychometrika*, *43*, 145–160.
- Rubin, D. B. (1984). Assessing the fit of logistic regressions using the implied discriminant analysis. (Comment on Landwehr, et al.'s, paper). *Journal of the American Statistical Association*, *79*, 79–80.
- Snee, R. (1974). Graphical display of two-way contingency tables. *American Statistician*, *38*, 9–12.
- Srole, L., Michael, T. S., Opler, S. T., and Rennie, T. A. C. (1962). *Mental health in metropolis: The midtown Manhattan study*. New York: McGraw Hill.
- Strauss, D. (1981). Choice by features: An extension of Luce's model to account for similarities. *British Journal of Mathematical and Statistical Psychology*, *34*, 50–61.
- Takane, Y. (1980). Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment. *Japanese Psychological Research*, *22*, 188–196.
- Takane, Y., Bozdogan, H., & Shibayama, T. (1987). (in press). Ideal point discriminant analysis, *Psychometrika*.
- Takane, Y., & Shibayama, T. (1984). Multiple discriminant analysis for predictor variables measured at various scale levels. *Proceedings of the 12th Annual Meeting of the Behaviormetric Society of Japan*, 99–100.
- Takane, Y., & Shibayama, T. (1986). Comparisons of models for stimulus recognition data. In J. de Leeuw, W. Heiser, J. Meulman, and F. Critchley (Eds.), *Multidimensional data analysis* (pp. 119–148). Leiden: DSWO Press. (With discussion)
- ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, *67*, 1167–1179.
- ter Braak, C. J. F. (in press, June). Partial canonical correspondence analysis. In H. H. Bock (Ed.), *Proceedings of the first IFCS conference*, Aachen.
- Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological Review*, *36*, 222–241.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- van der Heijden, P. G. M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, *50*, 429–447.
- Yanai, H. (1987, June). Partial correspondence analysis. Paper presented at the 1987 Annual Meeting of the North American Classification Society, Montreal.