MULTIWAY DATA ANALYSIS
R. Coppi and S. Bolasco (Editors)

IDEAL POINT DISCRIMINANT ANALYSIS: IMPLICATIONS FOR MULTIWAY DATA ANALYSIS*

YOSHIO TAKANE

McGill University, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, Quebec, H3A 1B1 Canada

Many contingency tables obtained in social survey research are multiway. Ideal point discriminant analysis (IPDA) can be applied to multiway contingency tables by rearranging them into two-way tables. Multiway structures of rows and/or columns of the contingency tables may be incorporated as constraints in their representations, and the best fitting structures may be identified through model comparisons by AIC. Two examples are given; one is a 2 X 2 X 2 table commonly encountered in social science research, and the other 2 X 2 X L tables arising from stimulus recognition (identification) and signal detection experiments where L is the number of manipulated experimental conditions.

1. INTRODUCTION

Ideal point discriminant analysis (IPDA) was proposed by Takane, Bozdogan and Shibayama (1987) for discriminant analysis with mixed measurement level predictor variables. The method has recently been extended to the analysis of contingency tables (Takane, 1987). In a manner similar to correspondence analysis (Greenacre, 1984; Nishisato, 1980) IPDA provides joint spatial representations of row and column categories of the contingency tables. At the same time it allows statistical evaluation of various structural hypotheses on the representations. Some related work has been done by Van der Heijden & de Leeuw (1985), Heiser (1987), ter Braak (1987), Escoufier (1987), and Ihm & van Groenewoud (1984).

In this paper we explore possible applications of IPDA to multiway data analysis. Two specific examples are given. One was obvious from the initial inception of the model, and calls for incorporation in the predictor set of interactions among the original predictor variables and their subset selection. The other was somewhat unexpected. It relates to psychophysical experiments in detection, discrimination and identification of stimuli, where stimulus presentation and other conditions are systematically

* The work reported in this paper was supported by grant A6394 from Natural Sciences and Engineering Research Council of Canada.

manipulated. All the examples to be discussed in this paper require only unidimensional representations, and are not particularly suitable to demonstrate the advantage of spatial representations by IPDA. However, we will see a lot of instances of the second advantage, namely detailed model evaluation feature of IPDA.

2. THE BASIC DATA

The basic data for IPDA are two-way contingency tables. Rows of the tables represent categories of predictor variables, while columns represent categories of criterion variables. Multiway contingency tables have to be collapsed into two-way contingency tables before IPDA is applied. The most important consideration is the distinction between the predictor variables and the criterion variables. A decision has to be made regarding which variables go to the predictor side and which variables to the criterion side. Once this decision is made, the remaining problem is how to combine categories when there are more than a single predictor (or criterion) variable. IPDA allows us to specify various structural hypotheses on rows and/or columns of contingency tables and is capable of choosing the best structure through model comparisons by AIC (Akaike, 1974). Input data should be arranged so as to allow maximum flexibility in the specification of those structural hypotheses. For example, factorial combinations of categories of the predictor variables may be taken initially, from which the best prediction model is arrived at by eliminating unnecessary predictor variables and their interaction effects. See Table 1 for an example of a multiway table rearranged into a two-way table by factorially combining categories of two predictor variables.

3. THE MODEL

In IPDA, contingency table analysis is viewed as a discrimination problem of column categories (criterion groups) based on the information about row categories (predictors). Both row and column categories are represented as points in a multidimensional euclidean space. Their coordinates are assumed to be simple (usually, linear) functions of external variables reflecting structural relationships among the categories. The probability of an observation falling in column j given row i is stated as a decreasing function of the distance between the respective points.

Let f_{ij} denote the joint frequency of the i-th row category and the j-th column category. Let F denote the R by C matrix of f_{ij}. Let X denote an R by p matrix of predictor variables, where discrete variables are represented in dummy-variable form, and continuous variables centered. Interactions among the original predictor variables may be included in X. Let Y be the R by A matrix of coordinates of row categories, where A is the prescribed dimensionality of the representation space. The maximum dimensionality is min (R-1, C-1). It is assumed that

(1) $$Y=XB$$

where B is a matrix of coefficients (weights).

Let M be the C by A matrix of coordinates of column categories. In the simplest case this M is assumed to be given by centroids of Y. Let D_c denote the diagonal matrix of column totals of F. Then

(2) $$M=D_c^{-1}F'Y$$

Other meaningful structures may also be incorporated in a manner similar to (1). The Y in (1) and M in (2) provide a joint spatial representation of the row and the column categories.

Let d_{ij} denote the euclidean distance between row i and column j. Then

(3) $$d_{ij}^2 = (\underline{y}_i - \underline{m}_j)'H(\underline{y}_i - \underline{m}_j)$$

where $\underline{y}_i$ and $\underline{m}_j$ are the i-th and the j-th row of Y and M, respectively. An A by A metric matrix, H, may depend on (subsets of) rows and/or columns. When H does not depend on rows or columns, it may be set to an identity matrix.

Let $p_{j/i}$ denote the conditional probability of column j given row i. It is assumed that this conditional probability is given by

(4) $$p_{j/i} = \frac{w_j \exp(-d_{ij}^2)}{\Sigma w_k \exp(-d_{ik}^2)}$$

where w_j is the bias parameter for column j. To remove scale indeterminacy in w_j it is convenient to require $w_k=1$. This bias parameters may depend on subsets of rows. The conditional likelihood for the entire data set is defined as a product multinomial using $p_{j/i}$, and model parameters, B and w_j (and possibly H), are estimated so as to maximize the conditional likelihood. Once the maximum likelihood is obtained, the AIC statistic can be evaluated in a straightforward manner, and used for identifying the best fitting model.

4. EXAMPLES

Within the basic framework of IPDA presented in the previous section several interesting multiway data analyses are possible. There are three possible places in the model where multidimensional structures of the data might be captured:

(A) The matrix of predictor variables, X, may be manipulated to reflect multiway structures of row categories.

(B) The metric matrix, H, in the distance model may be used to represent differences in the structural relationship between rows and columns of subtables. (Throughout this paper, however, we assume the identity metric.)

(C) The bias parameters $w_j (j=1,\ldots,C)$ may be allowed to vary over different subsets of rows. This is interesting, when the data are taken under the conditions that systematically bias responses.

Two examples are given in this section to illustrate (A) and (C) above.

4.1. An Obvious Application

The first example pertains to (A) above. A multiway contingency table is defined by several discrete variables, one of which is taken as the criterion variable (its categories constituting columns of the table) and the others taken as the predictor variables, whose categories are factorially combined to form row categories of the table. By incorporating interactions among the original predictor variables IPDA allows an analysis similar to loglinear discriminant analysis (Andersen, 1980) or logit analysis of multiway contingency tables. Subset selection of the predictor variables (and interactions among them) allows us to identify the best row structure of the tables.

Table 1 gives frequencies of death and nondeath penalties in murder cases in Georgia as functions of race of defendant and that of victim. The rate of death sentence varies dramatically across rows defined by the latter two variables. An important question to ask is what is the best (the most parsimonious) way to account for the difference. With this simple example what is going on is fairly obvious without even applying any statistical methods. However, an obvious example like this one helps confirm the validity of analysis results obtained by IPDA.

We first obtained the unconstrained solution; that is, no special relationships are assumed among the rows. The unconstrained solution can be obtained by setting $X = I$. Four row points are located in the order of (2), (4), (3), (1) from left to right in a unidimensional space. This coincides with descending order of the death penalty rate. That is, the more to the left a point is located, the higher is the probability of death sentence. A closer scrutiny of these four points reveals two distinct groups, (2) and (4) on one hand, and (3) and (1) on the other. These two groups contrast white victim and black victim. Within each of these groups a point located more to the left corresponds with the case in which the race of defendant and that of victim disagree. This suggests that victim's race and the interaction

Table 1 Frequency of death sentence as a function of race of defendent and that of victim

	Race of Defendent	Race of Victim	Death Sentence Yes	Death Sentence No	Rate
(1)	Black	Black	18	1420	1.2%
(2)		White	50	178	21.9%
(3)	White	Black	2	62	3.1%
(4)		White	58	678	7.8%

(From the Baldus and Woodworth study cited in the first issue of Chance, 1988, p. 7).

between defendant's race and victim's race (whether they agree or not) will provide an excellent account of the rate of death penalty.

We actually tried all possible models. There are eight of them altogether including nonhierarchical ones, formed from possible subsets of three predictor variables: 1. Race of Defendant, 2. Race of Victim, 3. Interaction between 1 and 2. Of the eight possible models two have special status. The one with all the three predictor variables is equivalent to the unconstrained model. It is also equivalent to the saturated model. The one with no predictor variables is equivalent to the independence model between rows and columns.

Table 2. Summary statistics for the data in Table 1.

	Predictor Variables: Race of Defendant	Race of Victims	Interaction+++	AIC	number of parameters
1+	Yes	Yes	Yes	866.6	4
2	Yes	Yes	No	869.3	3
3	Yes	No	Yes	919.3	3
4	No	Yes	Yes	864.8*	3
5	Yes	No	No	999.8	2
6	No	Yes	No	895.1	2
7	No	No	Yes	937.5	2
8++	No	No	No	1009.5	1

* The minimum AIC solution
\+ Equivalent to the saturated model
++ Equivalent to the independence model
+++ The race of defendant and that of victim agree or don't agree

Table 3. Estimated coefficients and their standard errors.

Variable	Category	Coefficient	Standard Error
Victim	Black	.473	.041
	White	-.730	.063
Races of Defendant and Victim	Agree	.087	.041
	Disagree	-.653	.105
		Coordinates of Four Cases	
(1) Defendant / Victim	Black / Black	.560	.031
(2) Defendant / Victim	Black / White	-1.383	.072
(3) Defendant / Victim	White / Black	-.180	.139
(4) Defendant / Victim	White / White	-.643	.074
		Coordinates of Ideal Point	
Death Penalty	Yes	-.755	.030
	No	.041	.002

Results are summarized in Table 2. Indeed, as we expected, the model with Race of Victim and Interaction is found to be the best fitting model according to the minimum AIC criterion. Table 3 indicates estimated coefficients for categories of the two best predictor variables. The probability of death sentence is higher when the victim is white, and when the race of victim and that of defendant disagree. This gives the highest probability of death sentence when the victim is white and the defendant is black, while the lowest probability when both victim and defendant are black. Coordinates of the four row points and of the column points given in Table 3 (corresponding to the best fitting solution) are remarkably similar to those corresponding to the unconstrained solution, confirming the validity of the minimum AIC procedure. Notice that Case (4) is located closer to "Yes to Death Penalty" than is Case (2), and yet the latter is associated with a higher rate of death sentence. (21.9% as opposed to 7.8%). This is because what counts in the conditional probability is the difference in the squared euclidean distances of a row point to the column points. Indeed the difference between the squared distance from (2) to "Yes (to death penalty)" and that from (2) to "No (to death penalty)" is larger than the analogous difference for row point (4), which explains the higher rate of death penalty for Case (2). In general the IPDA model stipulates a monotonic probability function along the direction connecting two column points.

Similar analyses as above could have been performed by the loglinear model. We have attempted to compare the results from IPDA and those from loglinear analysis. Unfortunately the best fitting model identified by IPDA could not be fitted by an available computer program for the loglinear analysis because of the nonhierarchical nature of the model.

4.2. An Unexpected Application

The second set of examples pertains to (C) above, and represents a somewhat unexpected application of IPDA. In many psychology experiments both stimulus and response sets (categories) are fixed, while other experimental conditions are systematically manipulated. Examples of such manipulations are stimulus exposure duration, response deadline, prior probabilities, pay-offs, etc.

Two types of experimental manipulations should be distinguished. One type of manipulations such as exposure duration and response deadline uniformly enhance or deteriorate subject's performance. This type of manipulations may be captured in the metric matrix H. The other type of manipulations such as prior probabilities and pay-offs, on the other hand, systematically bias responses. This type of effects is more adequately represented by different sets of bias parameters associated with different experimental conditions. It is this latter type of situations we are concerned with in this paper. Absolute identification (stimulus recognition) experiments and signal detection experiments provide excellent examples of this kind of situations. Although in neither types of experiments the number of stimuli and the number of response categories are restricted to two, we focus our attention to 2 by 2 cases in this paper.

As an example let's look at Table 4, which displays five 2 by 2 frequency tables stacked on top of one another obtained from a two-stimulus absolute identification (stimulus recognition experiment (Laming, 1968). Either stimulus a or b was presented in each trial, and the subject's task was to identify the stimulus. Two possible responses are designated by corresponding capital letters (A for a and B for b). Under a specific stimulus presentation condition we obtain a single 2 by 2 table. Prior probabilities of presenting two stimuli were systematically varied from .25 to .75 in steps of .125. Altogether there were five conditions. Each condition was based on 4,800 observations (aggregated from 200 observations each of 24 subjects.) The bias parameter is expected to vary systematically over the different prior probability conditions, but an interesting question is whether the separation between the two stimuli is constant across the conditions. The separation measured in terms of $d = y_a - y_b$ supposedly reflects subjects' sensitivity, which should be constant according to the theory of signal detectability (Green & Swets, 1966). The question may be answered by obtaining two solutions, one solution obtained under the assumption that d is constant, and the other obtained under no such assumption, and by comparing their goodness of fit (GOF). The latter model may in effect be fitted by applying IPDA to each 2 by 2 table separately. IPDA in this case is analogous to choice model analysis (Luce, 1963) of signal detection data. Like Luce's choice model, IPDA as applied to a single 2 by 2 table is not restrictive. It is in fact equivalent to the saturated model. The likelihood for the entire data set is obtained by aggregating individual likelihoods across conditions. Estimates of parameters and GOF of this model are given in columns labelled (1) in Table 4. The estimates of d do not vary very much across conditions. Corresponding estimates and GOF of the restricted model are given in columns labelled (2). According to the minimum AIC criterion this restricted model is the better fitting model, indicating that the equality of d is supported in this particular example. This is really amazing because there are 24,000 observations in total which are adequately described by just six parameters.

The next two examples come from Yes-No signal detection experiments, in which two stimuli are "noise" and "signal", and two response categories are "Yes, there is a signal" and "No, there is no signal". In the first example each of four subjects was examined under 12 to 13 different pay-off conditions. Unlike the previous example, each subject's data were analyzed separately. Each pay-off condition is based on 400 observations, and the prior probability of each stimulus was set to .5. Results are reported in Table 5. In all the four cases $d=y_s-y_n$ is found to vary significantly across different pay-off conditions. The perceptual process, as reflected in d, is

Table 4. Analysis of Laming's (1968, p. 55) stimulus identification data in which prior probabilities of two stimuli are manipulated. (Each condition is based on 200 observations from 24 subjects.)

			Response		(1)		(2)	
Condition	p(a)	Stimulus	A	B	d+	w	d++	w
1.	.250	a	1121	79	1.965	.261	1.950	.271
		b	33	3567				
2.	.375	a	1726	74	1.955	.381	1.950	.384
		b	49	2951				
3.	.500	a	2330	70	1.948	.482	1.950	.478
		b	59	2341				
4.	.625	a	2942	58	1.923	.615	1.950	.616
		b	83	1717				
5.	.750	a	3572	28	1.977	.756	1.950	.745
		b	8	1117				
				AIC	5458.9		5454.6*	
				#(para.)	10		6	

* Minimum AIC

\+ Obtained under the assumption that d varies across different p(a). Equivalent to the saturated model.

++ Obtained under the assumption that d is constant across different p(a).

not invariant under the different pay-off conditions. There may also be a confounding sessional (within-conditions) effect, although in this particular example there is no way to tear apart between-conditions and within-conditions effects, since there are no replicated sessions run under identical conditions.

Table 5. Summary of analysis of Swets, Tanner & Birdsall's (1961) visual signal detection data in which pay-offs are manipulated (12 ~ 13 sessions & pay-off conditions, equal prior probabilities of noise and signal, each session based on 400 observations).

	d, w vary across conditions+	d constant w varies across conditions
Observer 1	4200.7*	4308.4
	(24)	(13)
Observer 2	4233.0*	4274.9
	(26)	(14)
Observer 3	4034.4*	4074.0
	(24)	(13)
Observer 4	5042.4*	5068.4
	(26)	(14)

* Minimum AIC
Legend: AIC and the number of parameters in parentheses
\+ Equivalent to the saturated model.

The last example partly addresses the above question. Two subjects were observed in ten sessions and under five different pay-off conditions. Two sessions each were run under an identical pay-off condition. In this case it may be possible to isolate within-conditions and between-conditions effects. Each session consisted of 300 trials. Again each subject's data were analyzed separately. Four different models were fitted: (1) Both $d=y_s-y_n$ and w are assumed to vary both within and across pay-off conditions. (This is equivalent to the saturated model.) (2) The d is assumed constant, but w assumed variable both within and across conditions. (3) Both d and w are assumed constant within conditions, but assumed variable across conditions. (4) The d is assumed constant both within and across conditions, but w assumed constant within, but variable across, conditions. Results are reported in Table 6. For subject 1 Model (1) is the best fitting model, whereas for subject 2 Model (2) is the best fitting model. There seem to be significant variations both <u>within</u> and <u>across</u> the pay-off conditions. This is the case only for the bias parameter for subject 2, but it is true for both the bias and d for subject 1. After having done these analyses we realized that two important models had been left out. (The current program is not capable of fitting these models). These models are: (5) The d is assumed constant within, but variable across, conditions, but w assumed variable both within and across conditions, and (6) The d is assumed variable both within and across conditions, but w assumed constant within, but variable across, conditions. Undoubtedly fitting these models would have provided more insight into the process.

There is an interesting relationship between IPDA and conventional approaches to estimation problems in the signal detection theory (Abrahamson & Levitt, 1969; Dorfman & Alf, 1968, 1969; Ogilvie & Creelman, 1968). However, due to space limitation, this topic has to be deferred to another paper.

Table 6. Summary of analysis of Tanner, Swets & Green's (1956) auditory signal detection data in which pay-offs are manipulated. (10 sessions, 5 conditions, equal prior probabilities of noise and signal, each session based on 300 observations)

		Observer 1	Observer 2
(1)	d, w vary across 10 sessions+	2972.2* (20)	3286.9 (20)
(2)	d constant w varies across 10 sessions	2976.8 (11)	3279.0* (11)
(3)	d, w vary across 5 pay-off conditions (constant within conditions)	2977.9 (10)	3325.6 (11)
(4)	d constant w varies across 5 pay-off conditions (constant within conditions)	2988.1 (6)	3320.1 (6)

* Minimum AIC's
Legend: AIC and the number of parameters in parentheses
\+ Equivalent to the saturated model.

5. CONCLUDING REMARKS

As we have seen, IPDA offers a number of interesting possibilities for multiway data analysis. The matrix of predictor variables may be manipulated to reflect multiway structures of contingency tables. The bias parameter may be allowed to vary systematically over different subsets of rows. In this paper we have seen just a couple of examples of these. We have not systematically explored other possibilities which may prove fruitful. Among them are time indexed data, such as panel data (transition matrices, mobility tables), cohort data, event history data, and other longitudinal data. This, however, is left for future investigation.

REFERENCES

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

Abrahamson, I.G., & Levitt, H. (1969). Statistical analysis of data from experiments in human signal detection. Journal of Mathematical Psychology, 6, 391-417.

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

Dorfman, D.D., & Alf, E. (1968). Maximum likelihood estimation of parameters of signal detection theory - a direct solution. Psychometrika, 33, 117-124.

Dorfman, D.D., & Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals - rating-method data. Journal of Mathematical Psychology, 6, 487-496.

Escoufier, Y. (1987, June). In the neighborhood of correspondence analysis. Paper presented at the first IFCS conference, Aachen.

Green, D.M., and Swets, J.A. (reprinted 1973). Signal detection theory and psychophysics. Huntington, N.Y.: Krieger, 1966.

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

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

Ihm, P., and van Groenewoud, H. (1984). Correspondence analysis and Gaussian ordination. COMPSTAT lectures 3, 5-60.

Laming, D. (1968). Information theory of choice-reaction time. London: Academic Press.

Luce, R.D. (1963). Detection and recognition. In Luce, R.D. et al. (Eds.) Handbook of mathematical psychology (Vol. 1), New York: Wiley.

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

Ogilvie, J.C., & Creelman, C.D. (1968). Maximum likelihood estimation of receiver operating characteristic curve parameters. Journal of Mathematical Psychology, 5, 377-391.

Swets, J.A., Tanner, W.P., & Birdsall, T.G. (1961). Decision processes in perception. Psychological Review, 68, 301-340.

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

Takane, Y., Bozdogan, H., and Shibayama, T. (1987). Ideal point discriminant analysis, Psychometrika, 52, 371-392.

Tanner, W.P., Swets, J.A., & Green, D.M. (1956). Some general properties of the hearing mechanism. Technical Report No. 30. Electrical Defense Group, Univ. of Michigan.

ter Braak, C.J.F. (1987). Unimodal models to relate species to environment. Wageningen, The Netherlands: Agricultural Mathematics Group.

van der Heijden, P.G.M., and de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. Psychometrika, 1985, 50, 429-447.