PSYCHOMETRIKA—VOL. 17, NO. 3.
SEPTEMBER, 1982

MAXIMUM LIKELIHOOD ADDITIVITY ANALYSIS

YOSHIO TAKANE

MCGILL UNIVERSITY

A maximum likelihood estimation procedure was developed to fit unweighted and weighted additive models to conjoint data obtained by the categorical rating, the pair comparison or the directional ranking method. The scoring algorithm used to fit the models was found to be both reliable and efficient, and the program MAXADD is capable of handling up to 300 parameters to be estimated. Practical uses of the procedure are reported to demonstrate various advantages of the procedure as a statistical method.

Key Words: simple additive model, weighted additive model, method of successive categories, directional ranking method, scoring algorithm, AIC.

Introduction

In many psychological phenomena a variable of concern can be expressed as an additive function of its determining factors. The additive model is simple, often permitting straightforward statistical analyses of data. It is no wonder that the analysis of additivity has drawn the attention of so many researchers (see, for example, Anderson, 1974; Krantz, Luce, Suppes & Tversky, 1971; de Leeuw, Young & Takane, 1976). Recently the simple additive model has been generalized into the weighted additive model [Takane, Young & de Leeuw, 1980], which accounts for individual differences in additivity by a differential weighting of additive factors. However, existing procedures for nonmetric additivity analysis (to obtain additive representations of ordinal data), such as MONANOVA [Kruskal, 1965], ADDIT [Roskam, 1968], TRADEOFF [Johnson, 1975], ADDALS [de Leeuw, et al., 1976], MORALS [Young, de Leeuw, & Takane, 1976], and WADDALS [Takane et al., 1980], are all based on the least squares principle, and are primarily descriptive in nature. Within the least squares framework, it is not an easy task to evaluate the goodness of fit of a model to nonmetric data [Takane, 1978]. The use of a Monte Carlo experiment [like the one conducted by Takane et al. (1980) for their WADDALS procedure] may provide a means of reducing the inconvenience associated with the lack of statistical inference capability in the least squares procedures. However, it must be rather extensive, and often requires painstaking efforts on the part of the users. Furthermore a separate Monte Carlo experiment must be conducted under each specific condition, since no proper error theory has yet been offered which links the results of Monte Carlo studies carried out in different contexts.

In this paper we develop a maximum likelihood estimation procedure for nonmetric additivity analysis, when the data are collected by either the categorical rating method, the pair comparison method or the directional ranking method. We call our procedure MAXADD (*MAX*imum likelihood *ADD*itivity analysis procedure). In contrast to the least

The research reported here was supported by Grant A6394 to the author from the Natural Sciences and Engineering Research Council of Canada. Portions of this research were presented at the Psychometric Society meeting in Iowa City, Iowa, in May, 1980.

Thanks are due to Jim Ramsay, Justine Sergent and anonymous reviewers for their helpful comments.

Two MAXADD programs which perform the computations discussed in this paper may be obtained from the author.

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

0033-3123/82/0900-3090$00.75/0

squares procedures the maximum likelihood method allows relatively straightforward hypothesis testing via the asymptotic chi square statistic derived from the likelihood ratio principle. The MAXADD procedure can fit both the simple and the weighted additive models. We may thus fit the two models, compare their goodness of fit, and choose the one which better fits a particular data set. The MAXADD procedure is also capable of incorporating various constraints on estimated model parameters. For example, certain parameters may be fixed to certain prescribed values, nonlinear equality constraints may be imposed on additive effects, etc. This leads to a number of interesting model comparisons which have not been possible previously. Furthermore the AIC statistic [Akaike, 1974; see also Takane, 1981], may be used to compare models which are not necessarily hierarchically ordered. This feature enables one to investigate additivity of prescribed dimensions in a representation of dissimilarity data under a specific power transformation of the additive model. The appropriate value of the power, on the other hand, may be determined by systematically varying its value [Kruskal, 1964].

A maximum likelihood additivity analysis procedure has been developed for quantitative data by Winsberg and Ramsay [1980]. The MAXADD procedure, primarily designed for nonmetric additivity analysis, should complement their procedure in terms of the variety of data that can be subjected to maximum likelihood additivity analysis. A procedure similar to MAXADD has been developed by Falmagne [1978] for pair comparison data. His procedure has been extended to multiple response data by Hamerle and Tutz [1980]. The MAXADD procedure is much more general in that it can also deal with directional ranking data and the weighted additive model.

The Model and the Data

The Model

The additive model we discuss in this paper postulates that the dependent variable can be expressed as an additive function of the independent variables. For illustrative convenience we focus our attention to the two-factor case. An extension to higher order designs is relatively straightforward. Let α_i and β_j denote the additive effects of the i^{th} level of Factor A and the j^{th} level of Factor B, respectively. The simple additive model (SAM) is then written as

$$\text{SAM:} \qquad \mu_{ij} = \alpha_i + \beta_j, \tag{1}$$

where μ_{ij} is the model prediction representing the combined effect of the i^{th} level of Factor A and the j^{th} level of Factor B. The key feature of SAM is that this combined effect is obtained by simple addition of a contribution from Factor A and a contribution from Factor B.

We also consider the weighted additive model (WAM) [Takane et al., 1980], which is an individual differences extension of SAM. Let w_{kA} and w_{kB} denote the weights individual k attach to Factor A and Factor B, respectively. The WAM can then be stated as

$$\text{WAM:} \qquad \mu_{ijk} = w_{kA}\alpha_i + w_{kB}\beta_j, \tag{2}$$

where α_i and β_j are the same as in (1), and μ_{ijk} is analogous to μ_{ij}. However, it now has a subscript k, implying that this quantity is specific to the k^{th} individual. The WAM assumes that the additive effects are constant across individuals. It also assumes that the individual differences in additivity arise from the difference in the way individuals evaluate those additive effects and combine them into a global judgment. The difference in individuals' evaluation scheme is represented by the weights. The model is similar to Carroll and Chang's [1970] individual differences multidimensional scaling except that the additive effects may not correspond to dimensionwise differences in stimulus coordinates in the present case. Instead the levels of the additive effects can be any classification categories of observations. The nature of this model as well as its relationship to SAM has been fully described elsewhere [Takane et al., 1980], so that the interested reader is referred to that paper.

There is an indeterminacy problem associated with origins (zero points) of the additive effects in SAM. In order to eliminate the indeterminacy we may either "center" the additive effects (i.e., $\sum_{i=1}^{n_A} \alpha_i = 0$ and $\sum_{j=1}^{n_B} \beta_j = 0$ where n_A and n_B are, respectively, the number of levels in Factor A and in Factor B), or simply assign some arbitrary value to a level in Factor A and a level in Factor B. An additional indeterminacy problem exists in the case of WAM. That is, there is a tradeoff between the scale factor of the individual differences weights and that of the additive effects (i.e., for any constant c_A, define $w_{kA}^* = w_{kA} c_A$ and $\alpha_i^* = \alpha_i / c_A$ and we have $w_{kA}^* \alpha_i^* = w_{kA} \alpha_i$; the same for β_j as well.) We may fix the scale of the additive effects by requiring $\sum_{i=1}^{n_A} \alpha_i^2 = n_A$ and $\sum_{j=1}^{n_B} \beta_j^2 = n_B$.

We assume that the model prediction given in (1) or (2) is error perturbed. This error perturbed model prediction is denoted by y_{ijr} or y_{ijkr} depending on the fitted model. Subscript r is the index of replication. (For illustrative convenience we will assume that we are fitting WAM throughout the following discussion. The case of SAM follows with minor modifications.) In the metric approach the y_{ijkr} are assumed directly observable. In the nonmetric approach, on the other hand, only some incomplete information about y_{ijkr} is deemed accessible. The kind of incomplete information contained in the data depends on a particular data collection method employed.

In this paper we consider three data collection methods, the categorical rating, the pair comparison and the directional ranking methods. In the categorical rating method the subject is asked to rate stimulus S_{ij} (defined by the combined effect of the i^{th} level of Factor A and the j^{th} level of Factor B) on a rating scale having a relatively small number of observation categories (up to 7 or 9). In this case only category membership of y_{ijkr} is observed. In the pair comparison method, on the other hand, the subject is asked to compare two stimuli according to some prescribed criterion and to decide which one "dominates" the other. In this case only pairwise ordinal relations between y_{ikjr}'s are observed. Finally, in the directional ranking method the subject is required to rank order the S_{ij}'s in a specific order (either from the largest element to the smallest or entirely in the opposite direction). Thus, only the (directional) rank order information about y_{ijkr} is obtained.

It should be emphasized, however, that although the procedure we discuss in this paper mainly focuses on the three data collection methods described above, it by no means implies that the generalization to other data collection methods is impossible or even difficult. As a matter of fact it has been shown [Takane & Carroll, 1981] that specialized treatments of data conditionalities, missing data and tied observations in the directional ranking method allow a still wider range of data including those obtained by the pick-m largest (or smallest) method. Takane [Note 1] also suggests the categorical rating method of pair comparisons as an interesting possibility. In this method the subject is asked to rate the degree to which one stimulus "dominate" the other.

The Method

As we have seen in the previous section, the y_{ijkr} are not directly observed in the three data collection methods. Some bit of information is assumed lost in the observation process. For each data collection method we postulate a specific information reduction mechanism which is presumed to underlie the subject's judgmental process. The likelihood of observed data given in a particular form is then stated in terms of the likelihood of the error perturbed model prediction via the model of this information reduction mechanism. (This model is called the response model for the data.) We briefly discuss this model corresponding to each data collection method in turn. More comprehensive treatments of this topic can be found in Takane [1978, 1981] and Takane and Carroll [1981].

The Categorical Rating Method

We represent each category on a rating scale by an interval specified by its upper and lower boundaries. We assume that S_{ij} is judged to be in a certain category when the corresponding error perturbed model prediction, y_{ijkr}, happens to fall in an interval corresponding to that category. The probability, p_{ijkm}, that S_{ij} is classified into the m^{th} category by individual k, can then be stated as

$$p_{ijkm} = \Pr(b_{k(m-1)} < y_{ijkr} < b_{km}), \tag{3}$$

where b_{km} and $b_{k(m-1)}$ denote, respectively, the upper and lower boundaries of category m for individual k. Let

$$F_{ijkm} = \Pr(y_{ijkr} < b_{km}). \tag{4}$$

Then we have $p_{ijkm} = F_{ijkm} - F_{ijk(m-1)}$. The F_{ijkm} defined above is a cummulative distribution function of y_{ijkr}, and can be more explicitly written as

$$F_{ijkm} = [1 + \exp\{-s_k(b_{km} - \mu_{ijk})\}]^{-1} \tag{5}$$

under the logistic distributional assumption on y_{ijkr}. The s_k is an inverse measure of dispersion. (When category boundaries are considered random variables, we assume that $y_{ijkr} - b_{km}$ is logistic and (5) follows.)

Notice that the above model is basically Thurstonian, except that the logistic distribution, instead of the normal distribution, is assumed. It actually corresponds with the simplest case of the law of categorical judgment [Torgerson, 1958, p. 209]. The major difference between the classical and the present approaches is that we make a specific structural assumption under μ_{ijk} which is represented by (1) or (2), and directly estimates α_i and β_j rather than μ_{ijk}. Note that it is also possible to incorporate, as in the more general cases of the law of categorical judgment [Torgerson, 1958], various variance components models on s_k (see also Ramsay, 1982).

The number of parameters to be estimated may be reduced by imposing various structural assumptions on the category boundaries [Takane, 1981]. For example, the category boundaries may be assumed linearly related to each other (i.e., $b_{km} = \gamma m + \delta$), or it may be assumed that there are no individual differences in the category boundaries (i.e., $b_{km} = b_m$ for all k). The most appropriate assumption can be chosen empirically by comparing their goodness of fit.

Let Z_{ijkm} denote the observed frequency with which S_{ij} is classified into category m by individual k. The likelihood of Z_{ijkm} ($m = 1, \ldots, M$) is then given by

$$p_{ijk} = \prod_{m=1}^{M} (p_{ijkm})^{Z_{ijkm}}. \tag{6}$$

The joint likelihood of the total set of observations is in turn given by the product of p_{ijk}. This requires statistical independence of y_{ijkr}, so that the judgmental sequence should be carefully arranged so as to minimize possible sequential dependency. For example, successive presentations of stimuli involving the same level of a factor might be avoided.

The Pair Comparison Method

In this case we may assume that S_{ij} is judged to be larger than S_{uv} whenever the error perturbed model prediction corresponding to μ_{ijk} exceeds that corresponding to μ_{uvk}. The probability, p_{ijuvk}, that S_{ij} is judged larger than S_{uv}, is then given by

$$\begin{aligned} p_{ijuvk} &= \Pr(y_{ijkr} > y_{uvkr}) \\ &= \Pr(y_{ijkr} - y_{uvkr} > 0). \end{aligned} \tag{7}$$

Under the assumption of the logistic distribution on $y_{ijkr} - y_{uvkr}$, p_{ijuvk} can be explicitly written as

$$p_{ijuvk} = [1 + \exp\{-s_k(\mu_{ijk} - \mu_{uvk})\}]^{-1}. \tag{8}$$

Let Z_{ijuvk} represent the number of times S_{ij} is judged larger than (or more preferable to) S_{uv} among N_{ijuvk} repeated trials. The likelihood of Z_{ijuvk} is then stated as

$$p^*_{ijuvk} = (p_{ijuvk})^{Z_{ijuvk}}(1 - p_{ijuvk})^{N_{ijuvk} - Z_{ijuvk}}. \tag{9}$$

The likelihood of the total set of observations in turn is given by the product of all p^*_{ijuvk}. The statistical independence of p^*_{ijuvk}'s within each k is obtained, if the two pairs of error perturbed model predictions (y_{ijkr}'s) involved in two distinct judgments have pairwise equal covariances [Takane, 1978].

Again, the above model is essentially Thurstonian (Case V of the law of comparative judgment). The logistic assumption makes it equivalent to the BTL model [Bradley & Terry, 1952; Luce, 1959]. The BTL model (as well as Thurstone's Case V) presupposes what Krantz [1967] called "simple scalability". That is, all aspects of a stimulus pertinent to choice probability can be represented by a single number (μ_{ijk}) irrespective of comparison stimuli. This assumption is satisfied reasonably well, when the stimuli to be compared are relatively homogeneous regarding the similarity between them. However, when this assumption is untenable, the above model breaks down to varying degrees depending on the seriousness of violation (see, for example, Restle, 1961; Tversky & Russo, 1969; Tversky, 1972; Halff, 1976). When two stimuli are more similar than others, they are more comparable, and consequently choice probability involving them tends to be more extreme than what is predicted from (8). This situation may be remedied by incorporating a comparability index into s_k (in a manner similar to Takane [1980], Edgell and Geisler [1980], Strauss [1981] and Colonius [1981]), though the estimation procedure in this case would naturally be much more complicated. We may avoid heterogeneous stimulus pairs by excluding, for example, such pairs as S_{ij} and S_{iv}, which is easier to compare than S_{ij} and S_{uv} ($i \neq u, v; j \neq u, v$), since the former shares the i^{th} level of Factor A.

Another kind of problem may potentially arise, when the subject takes a different response strategy. We are assuming in this paper that stimuli are evaluated first for subsequent global (stimulus-by-stimulus) comparisons. We are also assuming that the evaluation process is additive (as indicated by (1) or (2)), and that the global comparison process is subtractive (as indicated by (7)). But what happens, if the subject first compares stimuli within dimensions (factors) and then, based on the dimensionwise comparisons, forms a global pair comparison judgment? In this case the additive representation, (1) or (2), no longer holds as it were, except under specialized conditions [Tversky, 1969].

Let us suppose, for simplicity, that the dimensionwise comparison process is subtractive, and that the evaluation (integration) process is additive. Then we obtain the general (weighted) additive-difference model [Tversky, 1969], which is written as

$$p_{ijuvk} = G[\phi_A\{w_{kA}(\alpha_i - \alpha_u)\} + \phi_B\{w_{kB}(\beta_j - \beta_v)\}] \tag{10}$$

for the two-factor case, where ϕ_A and ϕ_B are monotonically increasing skew-symmetric functions [i.e., $\phi(-x) = -\phi(x)$], and G is a distribution function of a symmetric distribution about zero. If ϕ_A and ϕ_B are linear (in which case ϕ_A and ϕ_B may be assumed to be identity without loss of generality), the argument in G above simplifies into:

$$\begin{aligned} w_{kA}(\alpha_i - \alpha_u) + w_{kB}(\beta_j - \beta_v) &= (w_{kA}\alpha_i + w_{kB}\beta_j) - (w_{kA}\alpha_u + w_{kB}\beta_v) \\ &= \mu_{ijk} - \mu_{uvk}. \end{aligned} \tag{11}$$

If we further assume that G is logistic, (10) reduces to (8), and the additive-difference model in this case is completely equivalent to the additive model postulated in this paper.

The additive model is thus a special case of the additive-difference model. On the other hand, by taking the dimensionwise differences as levels of additive factors (see the second example in the result section) and imposing appropriate equality and inequality restrictions on the additive effects, the additive-difference model (10) can be directly fitted, as if it were a special case of the additive model. Thus, it is possible, and may also be interesting, to compare the goodness of fit of the two models and to see if the subject follows the holistic (stimulus-by-stimulus) comparison process or analytic (dimension-by-dimension) comparison process (though, as mentioned above, in certain cases they are not empirically distinguishable).

The Directional Ranking Method

One of the most critical features of the directional ranking method is the directionality of ranking process [Takane & Carroll, 1981]. That is, ranking is performed in a specific direction, either from the largest element to the smallest or vice versa according to the instruction. In this case it may be assumed that a rank order is obtained by successive first choices. Suppose that the ranking is performed from the largest to the smallest, and that there are M objects to be rank-ordered. The subject is first asked to choose the largest element among the M objects, then to choose the next largest element among $M - 1$ remaining objects, and so on until a complete ranking is obtained among the M objects.

Let $\mu_k^{(m)}$ be the model prediction judged to be m^{th} largest by individual k at a particular time. (If the m^{th} largest element happens to be μ_{ijk}, then $\mu_k^{(m)} = \mu_{ijk}$.) Let $y_{kr}^{(m)}$ denote the error perturbed model prediction corresponding to $\mu_k^{(m)}$ generated at the r^{th} first choice. We may assume that the stimulus corresponding to $\mu_k^{(m)}$ is chosen as the m^{th} largest element (among $M - m + 1$ remaining elements after the $m - 1$ largest elements have been eliminated), when $y_{km}^{(m)}$ exceeds all $y_{km}^{(m')}$ for $m' = m + 1, \ldots, M$. Then the probability, $p_k^{(m)}$, that the stimulus corresponding to $\mu_k^{(m)}$ is chosen as the m^{th} largest element is given by

$$p_k^{(m)} = \Pr(y_{km}^{(m)} > y_{km}^{(m+1)}, \ldots, y_{km}^{(m)} > y_{km}^{(M)}). \tag{12}$$

We use Luce's model for first choice to state $p_k^{(m)}$ more explicitly [Luce, 1959]. That is,

$$p_k^{(m)} = \left[1 + \sum_{j=m+1}^{M} \exp\{-s_k(\mu_k^{(m)} - \mu_k^{(j)})\}\right]^{-1}, \tag{13}$$

which is the distribution function of the multivariate logistic distribution. The likelihood of a ranking is then defined as the product of the likelihoods of successive first choices. This requires the statistical independence of successive first choices, which is obtained under a similar condition to that in the pair comparison method. (A precise statement of the condition can be found in Takane and Carroll [1981].) The joint likelihood of multiple rankings is given by the product of the likelihoods of those rankings. When the ranking is performed in the other direction (i.e., from the smallest to the largest), we only need to change $-s_k$ in (13) to s_k.

Two major approaches have been proposed to the treatment of ties in a ranking [Takane & Carroll, 1981]. Suppose two stimuli corresponding to $\mu_k^{(m)}$ and $\mu_k^{(m')}$ are tied. The first approach defines $p_k^{(m)}$ excluding $\exp\{-s_k(\mu_k^{(m)} - \mu_k^{(m')})\}$, and $p_k^{(m')}$ excluding $\exp\{-s_k(\mu_k^{(m')}$

$-\mu_k^{(m)})\}$ in (13). Then $\mu_k^{(m)}$ and $\mu_k^{(m')}$ will not constrain each other. If, on the other hand, we include those terms in defining $p_k^{(m)}$ and $p_k^{(m')}$, $\mu_k^{(m)}$ and $\mu_k^{(m')}$ will be forced to be as close as possible to each other. The first approach is analogous to Kruskal's primary approach to ties and the latter to his secondary approach to ties [Kruskal, 1964].

It should be emphasized that we are *not* proposing the above model as a general ranking model. There are many different ways a ranking is obtained, and different processes of ranking may require different process models. The above model should be applied only when the directionality of ranking processes is strictly observed.

Note that despite all sorts of precautions the postulated models (for the data collection methods) may fail in certain situations. Ideally, alternative models should be constructed, and based on the same (model comparison) principle, the best fitting model should be selected for the response model. However, this seems to be a challenging process and will not be pursued in this paper.

Brief Algorithmic Considerations

Two MAXADD programs, one for categorical rating data and the other for pair comparison and directional ranking data, have been written which maximize the log of the likelihood defined in the previous section. Both programs use Fisher's scoring algorithm to solve maximum likelihood equations. The scoring algorithm is an iterative procedure which, starting from some initial estimates $\boldsymbol{\theta}^{(0)}$, updates the current parameter estimates $\boldsymbol{\theta}^{(q)}$ by solving

$$\varepsilon^{(q)}\mathbf{I}(\boldsymbol{\theta}^{(q)})(\boldsymbol{\theta}^{(q+1)} - \boldsymbol{\theta}^{(q)}) = \mathbf{u}(\boldsymbol{\theta}^{(q)}) \tag{14}$$

for $\boldsymbol{\theta}^{(q+1)}$, where $\mathbf{u}(\boldsymbol{\theta}) = (\partial \ln L/\partial\boldsymbol{\theta})$ is the score vector, $\mathbf{I}(\boldsymbol{\theta}) = E[(\partial \ln L/\partial\boldsymbol{\theta})(\partial \ln L/\partial\boldsymbol{\theta})']$ is Fisher's information matrix, and ε is the step-size parameter. This algorithm has been found to work very well under the present circumstance; the convergence is generally very quick and smooth.

There is one qualification necessary to the above statement. The total number of parameters to be estimated may sometimes be quite large. For example, in one of the examples to be discussed in the next section nearly 300 parameters have to be estimated. With so many parameters the scoring algorithm will not be very efficient, if used in an unmodified form. Fortunately this problem is largely circumvented by partitioning the parameters into several groups and updating each group successively within each iteration (see below). As a consequence the total number of iterations to convergence tends to increase (the convergence is at most linear as opposed to the quadratic convergence of the original scoring method), but the speed at which each iteration can be performed with reduced numbers of parameters usually more than offsets the larger number of iterations.

There are four distinct sets of parameters in the model; additive effects, individual differences weights, category boundaries and dispersion parameters. Of these the individual differences weights appear only in the WAM. The category boundaries are relevant only when the data are categorical rating data. Moreover, the dispersion parameters do not have to be explicitly estimated. We may arbitrarily set $s_k = 1$ for all k, and let model predictions compensate for their size. After the convergence is reached, we normalize the model predictions and adjust the values of the dispersion parameters accordingly. Thus, there are up to three sets of parameters to be estimated. We update these three sets of parameters conditionally and once in each iteration. We first update the additive effects with the other set(s) of parameters being fixed, then the individual differences weights (if WAM is fitted), and finally the category boundaries (if the data are obtained by the categorical rating method). The order in which the three sets of parameters are updated is rather arbitrary, however, since it does not seem to make very much difference in terms of the speed of convergence and the number of iterations required.

Individual-specific parameters such as w_{kA} and b_{km} (when they are to be estimated) are independent across individuals, provided that other parameter estimates are given. For example, w_{kA} and w_{kB} (for individual k) do not affect $w_{k'A}$ and $w_{k'B}$ for other individuals ($k' \neq k$). Consequently they can be updated separately for each individual. The derivatives of p_{ijk} in (6), p^*_{ijstk} in (9) and $p_k^{(m)}$ in (13) with respect to $w_{k'A}$ and $w_{k'B}$ are always zero for $k \neq k'$, so that the information matrix is block-diagonal, each block consisting of a matrix of order equal to the number of additive factors. This means that (14) can be solved for each block separately, and thus there is great economy of both storage space and computation time. Basically the same thing holds for the individual-specific category boundaries (b_{km}), except that in this case each block consists of an $M - 1$ by $M - 1$ matrix.

Examples of Application

In this section we report some empirical results obtained by the method described in the previous sections. We analyze two sets of data. One is the data collected by Kempler [1971] in his study of developmental change in children's perceptual structure of rectangles. The second data set pertains to dissimilarity data obtained by the method of triadic combinations. A set of stimuli employed are the well-known colors originally used by Torgerson [1958] in his study of classical multidimensional scaling. The data are analyzed from the viewpoint that the Minkowski power distance model is a kind of additive conjoint measurement [Beals, Krantz & Tversky, 1968].

Analysis of Kempler's Data

Kempler [1971] studied a systematic developmental change in the structure of weights children attach to height and width of rectangles when they make perceived largeness judgments. He constructed a set of 100 rectangles by factorially combining 10 height levels and 10 width levels each ranging from 10 inches to 14.5 inches in half inch intervals. He had four groups of children (1st, 3rd, 5th and 7th graders) judge each of the 100 rectangles as to whether it looked "large" or "small", From "conservation" literature he contended that younger children tended to put more emphasis on the height of rectangles than older children.

Takane et al [1980] fitted the weighted additive model to Kempler's data by alternating least squares (WADDALS), and found that the weights attached to height decrease and those attached to width increase rather consistently as a function of age, confirming Kempler's contention. However, the group differences were the primary focus of their analysis, and possible individual differences within the groups were completely disregarded. The purpose of reanalysis of Kempler's data here is to highlight the possible individual differences within the groups.

We first analyzed individual (nonaggregated) data for each age-group separately. Note that the individual data in this case consists of a set of two-category judgments ("large" or "small"). The main results are reported in the upper portions of Table 1. The first column of the table shows the results of fitting the weighted additive model (WAM), the second column those of the simple additive model (SAM) with individual differences in dispersion (σ_k), and the third column those of SAM without individual differences in dispersion (σ). (In WAM the dispersion parameters are necessarily allowed to vary over individuals.) In all cases category boundaries (there is only one category boundary for each individual) were allowed to vary over individuals. Three figures are reported in each cell of the table; the top one (designated as -2(LL)) is minus twice the log likelihood, the middle one (designated as

Table 1

Summary of MAXADD Analyses of Kempler's Data

Separate Analysis		WAM σ_k	SAM σ_k	SAM σ
	-2(LL)	1933.5	2140.8	2191.6
Grade 1	df	64	50	34
(N=16)	AIC	2061.5	2240.8	2259.6
	-2(LL)	3005.5	3166.2	3284.9
Grade 3	df	88	66	42
(N=24)	AIC	3181.5	3298.2	3368.9
	-2(LL)	3178.5	3283.3	3350.7
Grade 5	df	88	66	42
(N=24)	AIC	3354.5	3415.3	3434.7
	-2(LL)	2302.1	2354.0	2439.6
Grade 7	df	76	58	38
(N=20)	AIC	2454.1	2470.0	2515.6
	-2(LL)	10419.6	10944.3	11266.8
Total	df	316	240	156
	AIC	11051.6	11424.3	11578.8
Joint	-2(LL)	10513.2	11118.8	11477.5
Analysis	df	268	186	102
	AIC	11049.2	11490.8	11681.5

df) is the effective number of parameters in the fitted model, and the bottom figure is the value of the AIC statistic [Akaike, 1974], defined by

$$\mathrm{AIC} = -2(\mathrm{LL}) + 2\ \mathrm{df}. \tag{15}$$

The AIC defined above is a badness of fit measure. It has been devised as a descriptive index for comparing models having different numbers of parameters. Models with a larger number of parameters naturally tend to fit the data better (i.e., a larger value of likelihood is obtained). The AIC statistic avoids this superfluous improvement in fit by adding twice the degrees of freedom of the model to $-2(\mathrm{LL})$.

The effective number of parameters in the model is calculated as follows under the present circumstance. Let N_s be the number of individuals. Let there be n_A levels in Factor A and n_B levels in Factor B. Then the df of the simple additive model with constant dispersion is given by $n_A + n_B - 2 + N_s(M - 1)$, where, as before, M is the number of observation categories on a rating scale. In the present case we have $n_A = n_B = 10$, $M = 2$, and $N_s = 16$, 24, 24 and 20 for the 1st, 3rd, 5th and 7th graders, respectively. Additional N_s parameters are necessary to allow individual differences in dispersion in SAM. Additional

$2(N_s - 1)$ parameters are necessary to specify the weighted additive model, where 2 is the number of factors.

According to the AIC criterion (the smaller it is, the better is the fit) the weighted additive model fits to the data best in all age-groups, implying that there are substantial individual differences in the weight structure *within* each group. The nature of the group differences previously found with the WADDALS procedure should be reinterpreted in the light of this new evidence.

An interesting observation to be made is that the differences in the AIC values between WAM and SAM tend to diminish as the age levels go up. This implies that WAM is more essential to the younger age-groups than to the older age-groups. That is, the individual differences weights are more heterogeneous in the younger groups than in the older groups. Figure 1 shows the estimated individual differences weights in the four age-groups. One

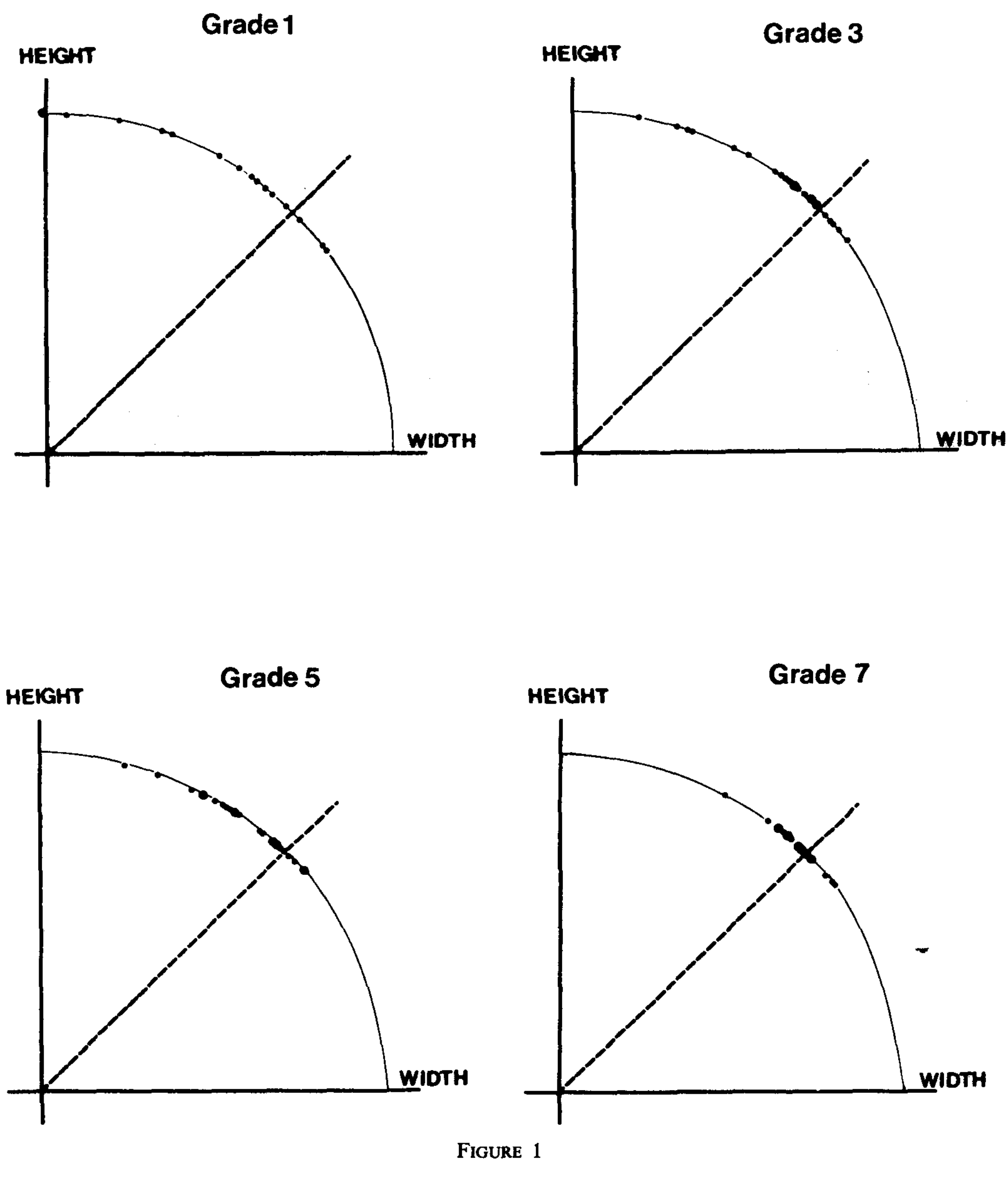

FIGURE 1

tendency is clear; the plots of the weight estimates tend to converge toward the 45° lines (dotted lines) between horizontal and vertical axes, as the age levels go up. For example, in grade 1 there are quite a few children who put disproportionately large weights on height including those two who totally ignore the width dimension, while those extreme subjects decrease in number and in its degree until a majority of children put approximately equal weights on both height and width of rectangles. Thus, the group differences we previously found in the WADDALS analysis of Kempler's data seem to be largely due to the difference in the group composition; the groups consist of heterogeneous individuals within themselves but the degree of this heterogeneity within the groups tends to decrease with age.

The above observation may be subjected to a more rigorous hypothesis testing. If each individual's weights are considered as indicating a direction on a plane, the test developed by Stephens [1969] can be applied to determine whether the degree of heterogeneity in the weight structure is homogeneous over the age-groups [Jones, Note 2]. Let $\mathbf{w}_{ki}$ be the vector of estimated weights for the k^{th} individual in the i^{th} group. Let $\mathbf{w}_i = \sum_{k=1}^{N_i} \mathbf{w}_{ik}$ (where N_i is the number of individuals in the i^{th} group) and $R_i = \| \mathbf{w}_i \|$ indicate the euclidean norm of $\mathbf{w}_i$ (i.e., $R_i = (\mathbf{w}_i' \mathbf{w}_i)^{1/2}$.) Then the test statistic is given by $Z = Z^*/C$, where

$$Z^* = D \ln\left\{ \sum_{i=1}^{s} d_i(N_i - R_i) \right\} - D \ln D - \sum_{i=1}^{s} \ln (N_i - R_i),$$

and

$$C = 1 + \frac{\left\{ \sum_{i=1}^{s} \frac{1}{d_i} - \frac{1}{D} \right\}}{3(s-1)}.$$

(We are restating Stephens' [1969] result here, since there is an important typographical error in his paper.) In the above formula s is the number of groups, $d_i = N_i - 1$ and $D = \sum_{i=1}^{s} d_i$. Under the null hypothesis that there are no differences in the spread of directions among the groups the statistic Z is known to be approximated by a chi square with $s - 1$ degrees of freedom. In the present case Z turned out to be 21.8 with 3 df. It indicates a significant departure from the equal spread hypothesis (well beyond the .001 significance level). The observed differences in the degree of heterogeneity in the spread of individual weights across different age-groups are thus substantiated.

There seem to be group differences in the weight structure as well as individual differences within the groups. But how about additive effects? Are there any significant group differences in the additive effects? In the above analysis the data for each group were analyzed separately, but this would not have been necessary, if it could be assumed that there were no group differences in the additive effects. In order to find out whether this assumption is tenable the data from all age-groups were analyzed simultaneously. The results are reported at the bottom of Table 1. As expected, WAM has been found to be the best fitting model among the three fitted models (WAM, SAM with σ and SAM with σ_k). Note that WAM applied to all age-groups simultaneously assumes that the additive effects are common to all age-groups. Separate analyses by groups, on the other hand, allow additive effects to vary over the different groups. The comparison of the goodness of fit between the two, then, should indicate whether the additive effects may be assumed common across groups. The AIC from the joint analysis of all age-groups by WAM has turned out to be 11049.2, which is slightly smaller than the joint AIC of 11051.6 from the separate analyses of the age-groups. This latter AIC is obtained by adding the group AIC's over all age-groups. The difference is small (2.4). Nonetheless it clearly indicates that the additive effects may be assumed common to all age-groups.

Additivity Analysis of Dissimilarity Data

Dissimilarity data are typically analyzed by multidimensional scaling (MDS). MDS finds a stimulus configuration in a multidimensional space in such a way that distances between stimuli best agree with the observed dissimilarities. The Minkowski power metric model, most frequently used in MDS, can also be viewed as a special type of additive model [Tversky & Krantz, 1979], in which dimension-wise differences in stimulus coordinates are taken as the additive effects. Therefore, if a set of stimuli have a factorial structure (not necessarily complete) according to some prescribed dimensions, we may test, whether the observed dissimilarities may be represented as an additive function of those dimensions.

Figure 2 displays the Munsell configuration of the nine colors (all red) originally employed by Torgerson [1958]. The stimuli are factorially arranged in terms of the Munsell Value (brightness) and Chroma (saturation) dimensions. If those Munsell dimensions represent the real psychological dimensions by which people judge dissimilarities (as advocated by the Munsell system), then dissimilarities between the colors have to be expressed as a monotonic function of the sum of the effects of intervals on the two dimensions.

Let the interval bounded by V_a and V_b (see Figure 2) be denoted by V_{ab} on the Value dimension. Let $\alpha(V_{ab})$ represent the contribution of interval V_{ab}. Similarly the contribution of interval C_{xy} on the Chroma dimension is denoted by $\beta(C_{xy})$. Under the additivity assumption the dissimilarity $\delta_{ab,\,xy}$ defined on V_{ab} and C_{xy} should be monotonically related to

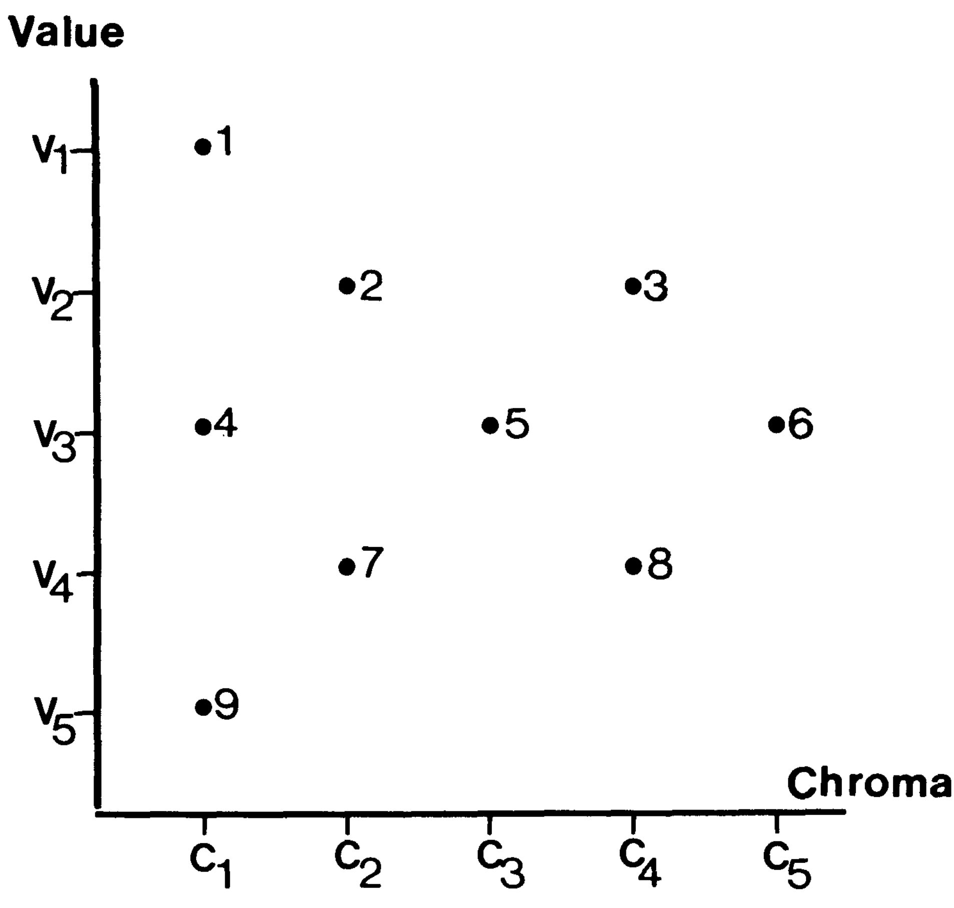

FIGURE 2

$\alpha(V_{ab}) + \beta(C_{xy})$. Let us assume, as in the Minkowski power metric model, that this function is a power transformation. Then we may write

$$\delta_{ab,\,xy} = \{\alpha(V_{ab}) + \beta(C_{xy})\}^{1/p}, \qquad (p \geq 1). \tag{16}$$

For example, the dissimilarity between stimuli 1 and 3 are defined on V_{12} and C_{14}. Hence, $\delta_{12,\,14} = \{\alpha(V_{12}) + \beta(C_{14})\}^{1/p}$. For a given value of p we may perform an additivity analysis of the dissimilarity data by estimating $\alpha(V_{ab})$ and $\beta(C_{xy})$, and taking $1/p^{\text{th}}$ power of $\alpha(V_{ab}) + \beta(C_{sy})$. We require that α and β are both nonnegative, and are zero for (and only for) null intervals (e.g., $\alpha(V_{aa}) = \beta(C_{xx}) = 0$).

Note that (16) is slightly more general than the Minkowski power metric model. To make it equivalent to the latter we need some constraints on α and β; the $1/p^{\text{th}}$ power of α or β should be equal to the width of that interval in the Minkowski power metric model. Consequently we need to consider only those intervals having no intervening boundaries (e.g., V_{12}, V_{23}, V_{34}, etc.: the same for Chroma). The remaining intervals can be expressed as a sum of those basic intervals. For example,

$$\alpha(V_{13})^{1/p} = \alpha(V_{12})^{1/p} + \alpha(V_{23})^{1/p}. \tag{17}$$

The above equation serves as a constraint imposed on $\alpha(V_{13})^{1/p}$. Analogous constraints should also be imposed on β. By incorporating constraints we can evaluate the goodness of fit of the Minkowski power metric model in reference to that of the unconstrained additive model.

The degrees of freedom associated with the unconstrained additive model is 20 under the present circumstance. There are 15 intervals on each dimension, five of which are null intervals whose effects are set to zero, and there are two dimensions. They are reduced to eight in the Minkowski power metric model. Only four basic intervals need be estimated on each dimension. The current MAXADD programs use the penalty function method (see Zangwill, 1969; see also Ramsay, 1978; Lee & Bentler, 1980) to impose the constraints.

Dissimilarity data to be analyzed were collected by the method of triadic combinations. In this method stimuli are presented in triads, and the subject is asked to choose the most similar stimulus pair, and then the most dissimilar pair. Since choosing the most dissimilar pair among two remaining stimulus pairs (after the most similar stimulus pair is eliminated) is equivalent to choosing the most similar pair among the two remaining pairs, the task involved is essentially rank-ordering the three dissimilarities defined on a triad of stimuli from the smallest to the largest. Hence it is considered a special case of the directional rank order method [Takane & Carroll, 1981]. Six replicated observations were obtained from a single subject (male adult, normal vision) on a complete set of 84 triads. The results of MAXADD analyses of the data are summarized in Table 2. Both the unconstrained additive model and the Minkowski power metric model with prescribed dimensions were fitted with the value of p varied systematically. The basic structure of the table is the same as in Table 1. Three values are reported in each cell, which are, from the top to the bottom, minus twice the log likelihood, the *df* of the model and the value of the AIC statistic. The minimum AIC criterion indicates that the optimum value of p is somewhere between 2.0 and 2.5 in the Minkowski power metric model. (The values of AIC are 695.1 and 695.6 for $p = 2.0$ and 2.5, respectively.) This basically agrees with Kruskal's previous finding [Kruskal, 1964] in a similar situation. Our result, however, is contingent upon the veracity of the Munsell dimensions as a model of judged dissimilarities between the colors. For the two values of p ($p = 1.0$ and 2.0) with which the unconstrained additive model was fitted, the smaller values of AIC were consistently found in the unconstrained model than in their constrained counterparts. The difference, however, is relatively small for a near optimal value of p (i.e., $p = 2.0$). This may indicate that the constraint set implied by

Table 2

Summary of MAXADD analyses of the color data

Power		Minkowski power metric model	Unconstrained additive model
p=1.0	-2(LL)	813.0	678.6
	df	8	20
	AIC	829.9	718.6
p=1.5	-2(LL)	704.3	
	df	8	
	AIC	720.3	
p=2.0	-2(LL)	679.1	654.2
	df	8	20
	AIC	695.1	694.2
p=2.5	-2(LL)	679.6	
	df	8	
	AIC	695.6	
p=3.0	-2(LL)	685.1	
	df	8	
	AIC	701.1	

the Minkowski power metric model (like the one shown in (17)) is at least not too radically violated, given $p \approx 2.0$ and under the additivity hypothesis.

The next question is the additivity of the two Munsell dimensions without unidimensional constraints. In order to answer this question MAXSCAL-4 [Takane & Carroll, 1981] was applied to the same set of data. MAXSCAL-4 is a maximum likelihood MDS procedure specifically designed for dissimilarity data obtained by the directional ranking method. The MDS analyses do not assume that the Munsell dimensions are correct though, due to a limitation in the MAXSCAL-4 procedure, we had to assume $p = 2.0$, which was partly justified by the fact that the optimum value of p was near 2.0 in the previous additivity analysis. In the MDS analysis the best fitting model was found to be the three dimensional solution with the AIC value of 540.8. This value is distinctly smaller than that of the best solution obtained from the additivity analysis. Not only was the additivity assumption (i.e., additivity of the Munsell dimensions) drastically violated, but also was the derived stimulus configuration found to be curved in an interesting way in the three dimensional euclidean space. It is like a two dimensional surface embedded in the three dimensional space. More precisely, the configuration looks like a valley (or a ridge) between two mountains.

The third dimension did not emerge so distinctly in the previous studies [Torgerson, 1958; Nakatani, 1972; Saito, 1977; Takane, 1978] in which the same set of stimuli were employed. The apparent disagreement may be related to the fact that the present study obtained replications within a single subject rather than over different subjects. If the nature of individual differences is such that the individual's cognitive maps of colors differ in the way the configuration is curved along the third dimension, the data aggregated over different individuals will not yield a clear third dimension. On the other hand, the three dimensional configuration within the subject may have been caused by the noneuclidean (at least locally) nature of the cognitive map of colors [Takane, 1982]. The important point here is that, whatever the discrepancy is, the additivity of the two Munsell dimensions is clearly rejected with the current data set.

Conclusion

We have seen two examples of analyses with the MAXADD procedure. These are, of course, just two examples, and a host of other interesting applications could be found without much difficulty in scientific literature; in psychophysics [Anderson, 1970; Falmagne, 1976], in decision making [Wallsten, 1976], in human judgments [Cliff, 1959; Anderson, 1974] and in other social sciences [Green & Rao, 1971; Johnson, 1974]. Indeed such examples abound in scientific literature.

With so many potential applications the MAXADD procedure discussed in this paper adds a new phase to the analysis of additivity. It can fit both the simple and the weighted additive models, allowing interesting model comparisons which have not been possible previously. The asymptotic chi square and the AIC statistics make these model comparisons much easier to perform than with any previous least squares procedures in which they are not available. It can also deal with three major types of data, categorical rating, pair comparison and directional ranking data. Although there still are other types of ordinal data (e.g., nondirectional ranking data and categorical ratings of pair comparisons [Takane, Note 1]) that cannot be properly handled within the current MAXADD procedure, the above three types of data should cover most of the representative types of ordinal data which may arise in psychology. Furthermore, the constrained optimization feature of the MAXADD procedure enables one to perform interesting model comparisons which were not possible previously, including the comparison between the additive model and the additive-difference model.

REFERENCE NOTES

1. Takane, Y. Statistical procedures for nonmetric multidimensional scaling. Unpublished doctoral dissertation, The University of North Carolina, 1977.
2. Jones C. L. Statistical inference with directional data. Handout for the talk given at the Psychometric Society meeting, Iowa, 1980.

REFERENCES

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

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review* 1970, *77*, 153–170.

Anderson, N. H. Algebraic models in perception. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*, Vol. II. New York: Academic Press, 1974.

Beals, R., Krantz, D. H. & Tversky, A. Foundations of multidimensional scaling. *Psychological Review*, 1968, *75*, 127–142.

Bradley, R. A. & Terry, M. E. The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 1952, *39*, 324–345.

Carroll, J. D. & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, *35*, 283–319.

Cliff, N. Adverbs as multipliers. *Psychological Review*, 1959, *66*, 27–44.

Colonius, H. A new interpretation of stochastic test models. *Psychometrika*, 1981, *46*, 223–225.

de Leeuw, J., Young, F. W., and Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 1976, *41*, 471–503.

Edgell, S. E., & Geisler, W. S. A set-theoretic random utility model of choice behavior. *Journal of Mathematical Psychology*, 1980, *21*, 265–278.

Falmagne, J. C. Random conjoint measurement and loudness summation. *Psychological Review* 1976, *83*, 65–79.

Falmagne, J. C. Probabilistic choice behavior theory: axioms as constraints in optimization. In J. N. Castellan, Jr., & F. Restle (Eds.), *Cognitive theory*, Vol. 3. New York: Erlbaum Assoc. 1978.

Green, P. E., & Rao, V. Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 1971, *8*, 355–363.

Halff, H. M. Choice theories for differentially comparable alternatives. *Journal of Mathematical Psychology*, 1976, *14*, 244–246.

Hamerle, A. & Tutz, G. Goodness of fit tests for probabilistic measurement models. *Journal of Mathematical Psychology*, 1980, *21*, 153–167.

Johnson, R. M. Trade-off analysis of consumer values. *Journal of Marketing Research*, 1974, *11*, 121–127.

Johnson, R. M. A simple method for pairwise monotone regression. *Psychometrika*, 1975, *40*, 163–168.

Kempler, B. Stimulus correlates of area judgments: a psychophysical developmental study. *Developmental Psychology*, 1971, *4*, 158–163.

Krantz, D. H. Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, 1967, *4*, 226–245.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement*. Vol. I. New York: Academic Press, 1971.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, *29*, 1–27.

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of data. *Journal of the Royal Statistical Society*, Series B, 1965, *27*, 251–265.

Lee, S. Y. & Bentler, P. M. Functional relations in multidimensional scaling. *British Journal of Mathematical and Statistical Psychology*, 1980, *33*, 142–150.

Luce, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.

Nakatani, L. H. Confusion-choice model for multidimensional psychophysics. *Journal of Mathematical Psychology*, 1972, *9*, 104–127.

Ramsay, J. O. Confidence regions for multidimensional scaling analysis. *Psychometrika*, 1978, *43*, 145–160.

Ramsay, J. O. Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society, Series A*, 1982, in press.

Restle, F. *Psychology of judgment and choice*. New York: Wiley, 1961.

Roskam, E. E. *Metric analysis of ordinal data in psychology*. Voorschoten, Holland: VAM, 1968.

Saito, T. Multidimensional Thurstonian scaling with an application to color metrics. *Japanese Psychological Research*, 1977, *19*, 78–89.

Stephens, M. A. Multi-sample tests for the Fisher distribution for directions. *Biometrika*, 1969, *56*, 169–181.

Strauss, D. Choice by features: an extension of Luce's model to account for similarities. *British Journal of Mathematical and Statistical Psychology*, 1981, *34*, 50–61.

Takane, Y. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—theory and evaluations. *Japanese Psychological Research*, 1978, *20*, 7–17 and 105–114.

Takane, Y. Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment. *Japanese Psychological Research*, 1980, *22*, 188–196.

Takane, Y. Multidimensional successive categories scaling: a maximum likelihood method. *Psychometrika*, 1981, *46*, 9-27.

Takane, Y. The method of triadic combinations: a new treatment and its applications. *Behaviormetrika*, 1982, *11*, 37–48.

Takane, Y. & Carroll, J. D. Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, 1981, *46*, 389–405.

Takane, Y., Young, F. W. & de Leeuw, J. An individual differences additive model: an alternating least squares method with optimal scaling features. *Psychometrika*, 1980, *45*, 183–209.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Tversky, A. Intransitivity of preferences. *Psychological Review*, 1969, *73*, 31–48.

Tversky, A. Elimination by aspects: a theory of choice. *Psychological Review*, 1972, *79*, 281–299.

Tversky, A. & Krantz, D. H. The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, 1979, *7*, 572–596.

Tversky, A. & Russo, E. Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, 1969, *6*, 1–12.

Wallsten, T. S. Using conjoint-measurement models to investigate a theory about probabilistic information processing. *Journal of Mathematical Psychology*, 1976, *14*, 144–185.

Winsberg, S. & Ramsay, J. O. Monotonic transformations to additivity using splines. *Biometrika*, 1980, *67*, 669–674.

Young, F. W., de Leeuw, J., & Takane, Y. Regression with qualitative and quantitative variables: an alternating least squares method with optimal scaling features. *Psychometrika*, 1976, *41*, 505–529.

Zangwill, W. I. *Nonlinear programming: a unified approach.* Englewood Cliffs, N. J.: Prentice-Hall, 1969.

Manuscript received 4/21/81
Final version received 3/4/82