A REVIEW OF APPLICATIONS OF AIC IN PSYCHOMETRICS

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

1. Introduction

Many theories in psychology are still in early stages of development. It is rare to find a single definitive and uncontested theory developed for a single phenomenon. It is more usual that many competing models are formulated for a single phenomenon giving rise to the necessity of extensive model comparison. Introduction of a conceptually simple and easy-to-use criterion for model comparison, like AIC (Akaike, 1973), is thus considered a blessing in often complicated psychological research. Psychometrics is one of the earlier areas which introduced this statistic. In 1987, *Psychometrika*, a leading journal in psychometrics, published a special issue featuring four papers (Akaike, 1987; Sclove, 1987; Bozdogan, 1987; Takane, Bozdogan, & Shibayama, 1987) presented in a symposium on AIC held at the previous year's annual meeting of the Psychometric Society. This was brought about with the recognition of importance of AIC in psychometric research. The present paper reviews applications of AIC in psychological research, and highlights some of the difficulties in modelling psychological phenomena.

There are two major classes of models that we focus on in this paper, although there are a number of other models in psychometrics to which AIC may potentially be applied. Both classes of models have traditionally been attributed to psychometricians. One is multidimensional scaling (MDS), and the other the latent variable models, such as factor analysis (FA), analysis of covariance structures (ACOVS), latent structure analysis, etc. MDS represents similarity data by a distance model, and is one of the first models to which AIC was applied (Takane, 1978). FA, on the other hand, postulates latent variables to explain covariances among observed (manifest) variables. FA is the model for which AIC was originally conceived (Akaike, 1987). This paper reviews applications of AIC in these two and related methods.

2. MDS: An Informal Introduction

What is MDS? It is a data analysis technique that locates a set of points in a multidimensional space in such a way that points corresponding to similar stimuli are located close together, while those corresponding to dissimilar stimuli are located far apart. To take a simple example, if you are a driver, you know that many road maps have a table of intercity distances somewhere in the corner. What MDS does is to recover a map (relative locations of cities) based on the intercity distances. Given a map

H. Bozdogan (ed.), Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, 379–403.

it is relatively straightforward to measure intercity distances, but the reverse operation, that of recovering a map based on intercity distances, is not so straightforward. The task of MDS is to perform this reverse operation.

In MDS, we typically have some empirical measures of similarity among stimuli, and we represent the stimuli as points in a multidimensional space, so that their mutual distances in some sense best agree with the observed similarity relations. Why would we like to do that? Because a picture is worth a thousand words. In psychology this technique has been useful in understanding some aspects of human cognitive processes.

Here is an example (Takane, 1984). Stimuli are ten digits made up of seven line segments arranged in the form of number 8 (see Figure 1). Each digit is defined by a subset of these line segments. For example, digit 2 is made up of segments 1, 3, 4, 5, and 7. Two stimuli (not necessarily distinct) were briefly presented in each trial, and the subject was asked to indicate as quickly as possible whether the two stimuli presented were "same" or "different." Presumably it takes longer to discriminate more similar stimuli, so that (discrimination) reaction times can serve as similarity measures. MDS in this case locates stimuli with longer reaction times close together, and more discriminable ones far apart. There were two subjects in the study whose data were analyzed separately. Each subject underwent two sessions of 360 trials each. In exactly one half of the trials *same* pairs (pairs of two identical stimuli) were presented (10 *same* pairs presented 18 times each), and in the other half *different* pairs were presented (45 *different* pairs presented 4 times each) in random order. See Sergent & Takane (1987) for a more detailed account of the procedure. The portions of the data pertaining to *same* pairs were excluded from the analysis.

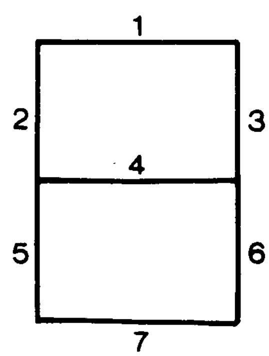

Figure 1. Seven line segments that constitute the digit stimuli.

Figure 2 presents a two-dimensional stimulus configuration obtained from Subject 1. A vertical line and a horizontal line (dotted lines) have been drawn to make important distinctions in the configuration: (1) First, all the stimuli located to the left of the vertical line have something in common. That is, they have all the three horizontal line segments, whereas those stimuli on the other side miss some of them. The horizontal direction in the space is thus related to the three horizontal line segments. (2) Next, all the stimuli below the horizontal line have segments 2, 3, & 6 as opposed to at least some of them missing in the stimuli above the line. The vertical direction is thus related to the vertical line segments. We may say that this subject is distinguishing the stimuli in terms of presence and absence of the horizontal and vertical line segments from which they were constructed.

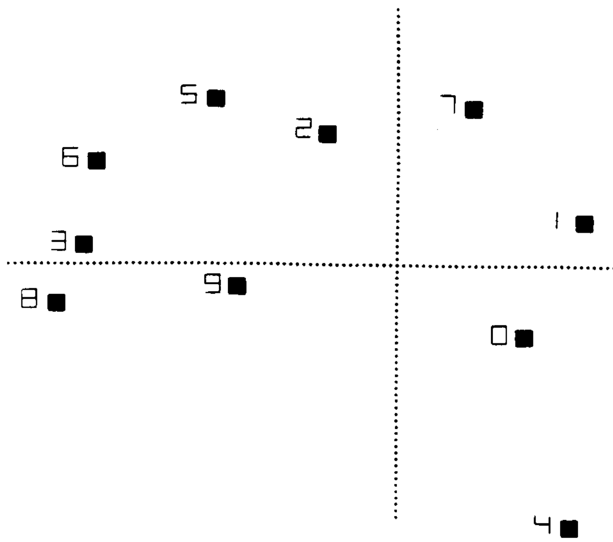

Figure 2. Derived two-dimensional stimulus configuration for Subject 1 for the digit stimuli.

Figure 3 presents the result obtained from Subject 2. This configuration is totally different from the previous one. It appears that the stimuli are arranged from 0 to 9 in a roughly circular manner. This implies that the digits which are numerically similar are less discriminable. That is, this subject is recognizing the stimuli as numbers, and is distinguishing them as such. Note that it is not at all necessary to recognize the stimuli as numbers in order to perform the task. It suffices to judge if the two stimuli presented have all line segments in common or not, as was done by Subject 1. Interestingly, this subject was totally unaware of his own strategy (that he was perceiving the stimuli as numbers). He himself had a great deal of difficulty in interpreting his own stimulus configuration, being preoccupied by the line segment characterization of the stimuli.

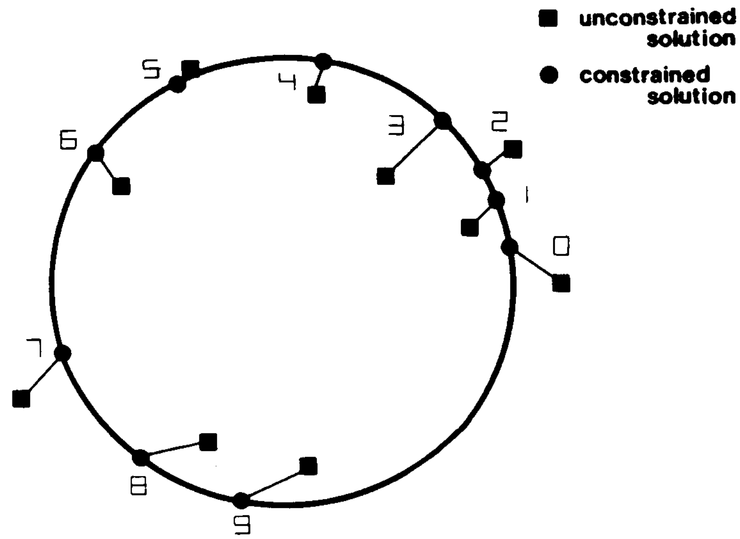

Figure 3. The same as figure 2 for Subject 2.

For the same set of stimuli we found two radically different configurations, which reflected different strategies for performing the task. MDS was effectively used to uncover initially unsuspected individual differences. MDS would be even more powerful if it allowed goodness of fit (GOF) comparisons among different models. For example: (1) So far we have assumed that the two-dimensional configurations are the best for both subjects. Is it really so? (2) Subject 1's judgments seem to have been based on patterns of presence and absence of the line segments from which the stimuli were constructed. Can a distance measure defined on the patterns of presence and absence of the line segments then explain his data adequately? (This will be referred to as hypothesis (2) below.) (3) For Subject 2, the stimuli are located nearly in a circular form. Can we say it is indeed circular? (This will be called hypothesis (3) below.)

These questions can be answered in a straightforward manner by the minimum AIC procedure. First, models that embody various hypotheses are fitted by the maximum likelihood method. Then, values of the AIC statistics are calculated for the purpose of model comparison. The smaller the value of AIC, the better the GOF of the model. An important point is that the model chosen by the minimum AIC criterion is not necessary a "true" model. The minimum AIC merely indicates that the model is best among competing models in the sense that, given the data at hand, the chosen model gives predictions which are closest to those generated by the true model. Thus, for example, in selecting appropriate dimensionality using AIC, the question answered is not what the correct dimensionality is, but how many dimensions can be reliably estimated, given the data at hand.

Table 1 summarizes the AIC values obtained for the digit data. In the table, n(p) indicates the number of independent parameters in the model. For both Subject 1 and Subject 2, the two-dimensional euclidean distance model fits the data better than the three-dimensional counterpart. For Subject 1, the two-dimensional euclidean model is also the best fitting model. Hypothesis (2) suggested above for Subject 1 (labelled "Linear" in the table) does not seem to be supported by the data. The model is called "Linear", because the distance between two stimuli are defined by a linear combination of contributions of features (line segments) on which the two stimuli differ (see section 4). For Subject 2, the constrained solution derived under hypothesis (3) gives the best fitting model. The stimulus configuration may indeed be considered circular for this subject. This solution has been superposed onto the unconstrained solution in Figure 3.

Dimensionality			Subject 1 Unconstrained	Subject 2 Unconstrained	Subject 2 Constrained
2	Euclidean	AIC	8.0*	34.6	23.4*
		n(p)	(17)	(17)	(10)
3	Euclidean	AIC	10.2	40.8	
		n(p)	(24)	(24)	
	Linear	AIC	26.0		
		n(p)	(7)		

*The minimum AIC solution.

Table 1. Summary statistics for the digit data.

3. Maximum Likelihood MDS

To discuss other applications of AIC, a more general account of MDS is necessary. In constructing a maximum likelihood MDS procedure, it is important to distinguish three essential ingredients (Takane, 1981); representation model, error model and response model. The representation model is the model that describes similarity relations between stimuli. The error model describes the way model predictions derived from the representation model are error-perturbed. The response model describes the mechanism by which the error-perturbed model predictions are transformed, when similarity judgments are made. We briefly discuss various possible specifications of these submodels in turn. It should be emphasized, however, that no matter which specifications of submodels may be used, the likelihood function is stated for a specific form of data as a function of parameters in these submodels. The maximum likelihood (ML) method is then applied to obtain estimates of the parameters. Values of AIC can readily be calculated from the maximum likelihood.

The representation model most frequently used in MDS is the euclidean model

$$d_{ij} = \{ \sum_{a=1}^{A} (x_{ia} - x_{ja}) \}^{1/2}, \tag{1}$$

where d_{ij} is the (euclidean) distance between stimuli i and j, x_{ia} is the coordinate of stimulus i on dimension a, and A is the dimensionality of the representation space. This is the model used in the previous example. There are nA parameters in this model, where n is the number of stimuli. However, there are translational and rotational indeterminacies inherent in the euclidean distance model, and exactly A(A+1))/2 parameters can be arbitrarily fixed. This leaves n(p) = nA - A(A+1)/2 independent parameters to be estimated.

In some cases (most typically, when the stimulus dimensions are separable), the city-block distance model defined by

$$d_{ij} = \sum_{a=1}^{A} | x_{ia} - x_{ja} | , \tag{2}$$

may be more appropriate (Sergent & Takane, 1987). There are (n-1)A independent parameters in this model. While there is no rotational indeterminacy in the city-block distance model, the translational indeterminacy is still in effect. Both the euclidean distance and the city-block distance models are special cases of the Minkowski power distance model,

$$d_{ij} = \{ \sum_{a=1}^{A} | x_{ia} - x_{ja} |^{r} \}^{1/r}, \tag{3}$$

where $r \geq 1$. Model (1) follows from (3) with r = 2, and Model (2) with r = 1.

In MDS subjects (when there are more than one) are often treated as replications. However, this is only justifiable when there are no systematic individual differences. When such is not the case, the following model, called weighted euclidean model, is often used (Carroll & Chang, 1970) to represent individual differences in similarity judgments:

$$d_{ijk} = \{ \sum_{a=1}^{A} w_{ka}(x_{ia} - x_{ia})^2 \}^{1/2}, \tag{4}$$

where k indexes subjects and w_{ka} is the weight attached to dimension a by subject k. This model assumes that there is a stimulus configuration common to all subjects, and that individual differences are produced by the differential weightings of the common dimensions by different subjects. Note that this model is still fairly restrictive. Note also that in this model the number of parameters to be estimated increases as the number of observations increases, so that, strictly speaking, the asymptotic theory for ML estimates does not hold.

Distances may sometimes be defined only for objects (stimuli) belonging to two mutually exclusive subsets. That is, stimulus i belongs to one set, and stimulus j to the other. In such situations Model (1) may be modified into

$$d_{ij} = \{ \sum_{a=1}^{A} (x_{ia} - y_{ja})^2 \},^{1/2} \tag{5}$$

where x_{ia} and y_{ja} are distinct. This model is sometimes called unfolding (or ideal-point) model (Coombs, 1964), and is useful for representing individual differences in preference judgments. In this model subjects are assumed to have ideal stimuli represented as ideal points, and distances between stimulus points and the ideal points are assumed inversely related to subjects' preferences for the stimuli. Takane, Bozdogan and Shibayama (1987) and Takane (1987b) used this model for discriminant analysis and analysis of contingency tables. Note that the number of parameters in Model (5) increases linearly with the number of subjects, as in Model (4). Additional constraints are often imposed to avoid inconsistency in ML estimates. For example, subjects' ideal points may be represented as linear combinations of some demographic information about the subjects.

The distance model may not always be the best model for representing similarity data (Tversky, 1977), and similarity models other than the distance model have been used in the literature. These includes the feature matching model (Tversky, 1977), various kinds of tree structures (Sattath & Tversky, 1977), and so on.

The d_{ij} defined above are assumed error-perturbed. The error model prescribes the nature of this error perturbation process. The error models that have been most frequently used are the normal error model,

$$\delta_{ijr} = d_{ij} + e_{ijr}, \text{ where } e_{ijr} \sim N(0, \sigma_{ij}^2), \tag{6}$$

and the log-normal error model (Ramsay, 1977),

$$\delta_{ijr} = d_{ij}e_{ijr}, \text{ where } \ln e_{ijr} \sim N(0, \sigma_{ij}^2), \tag{7}$$

where r indexes replications. The log-normal model is particularly attractive in MDS, since δ_{ijr} is always positive, it is positively skewed, and the variance increases with d_{ij}. The log-normal error model was the one assumed for reaction time data in the previous example. A variety of variance component models may further be assumed for σ_{ij}^2 (Ramsay, 1982). For example,

$$\sigma_{ij}^2 = \sigma^2 d_{ij}^s, \tag{8}$$

with s = 0, 1 or 2, and

$$\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2, \tag{9}$$

etc.

There are a number of different ways in which similarity data are collected in MDS. In the simplest case (dis)similarity data are taken on relatively continuous rating scales. In this case, δ_{ijr} (or at least a simple transformation of it) may be assumed directly observable (Ramsay, 1977). The likelihood function can then be stated in a relatively straightforward manner. However, in other cases such as reaction time data (Takane & Sergent, 1983), stimulus identification data (Takane & Shibayama, 1986, 1992), categorical rating data (Takane, 1981), pair comparison data (Takane, 1978), rank-order data (Takane, 1982; Takane & Carroll, 1981), etc., δ_{ijr} is not directly observed. It is transformed in a specific way depending on the particular methods for collecting the similarity data. The response model describes this transformation mechanism. Consequently, a specific response model has to be constructed for each specific data collection method. Response models have been constructed for all types of similarity data mentioned above. We refer to the papers cited above for specific response models.

In the remainder of this section we briefly describe the response model for the reaction time data (Takane & Sergent, 1983) as an example. For simplicity, we assume Models (1) and (6) with a constant variance, σ^2, and consider only *different* pairs. Let

$$Y_{ijr} = \begin{cases} 1, & \text{if stimulus i and j are judged to be "same" in replication r} \\ 0, & \text{otherwise} \end{cases}$$

T_{ijr} : the observed reaction time for stimuli i and j in replication r.

Let b denote the response threshold for the same-different judgments. It is assumed that a "same" judgment is elicited whenever $\delta_{ijr} < b$, otherwise a "different" judgment is observed. Let P_{ij} denote the probability that $\delta_{ijr} < b$. Then,

$$P_{ij} = \int_{-\infty}^{v_{ij}} \phi(z)dz$$

where $\phi(.)$ is the density function of the standard normal distribution, and $v_{ij} = (\ln b - \ln d_{ij})/\sigma$. The reaction time is assumed to take its maximum when $\delta_{ijr} = b$, and decrease as δ_{ijr} gets far away from b. Specifically,

$$\ln T_{ijr} \sim N\,(q(\ln d_{ij} - \ln b) + u,\ q^2\sigma^2)$$

for a "same" judgment (i.e., when $\delta_{ijr} < b$), and

$$\ln T_{ijr} \sim N\,(q(\ln b - \ln d_{ij}) + u,\ q^2\sigma^2)$$

for a "different" judgment (i.e., when $\delta_{ijr} > b$). (Here, q (< 0) and u are additional parameters.) Let $g_{ijr}^{(s)}(T_{ijr})$ and $g_{ijr}^{(d)}(T_{ijr})$ denote the density functions of T_{ijr} conditional on $Y_{ijr} = 1$ and $Y_{ijr} = 0$, respectively. The joint density of T_{ijr} and Y_{ijr} is then given by

$$f_{ijr}(T_{ijr}, Y_{ijr}) = \{g_{ijr}^{(s)}(T_{ijr})P_{ij}\}^{Y_{ijr}}\,\{g_{ijr}^{(d)}(T_{ijr})(1 - P_{ij})\}^{1-Y_{ijr}}.$$

Finally, the likelihood function for the entire set of observations is stated as

$$L = \prod_i \prod_j \prod_r f_{ijr}(T_{ijr}, Y_{ijr}), \tag{10}$$

assuming the independence among the observations.

The likelihood is maximized with respect to x_{ia} (i = 1,..., n; a = 1,..., A) in the representation model, σ^2 in the error model and b, q, and u in the response model by some iterative optimization procedure. Once the maximum likelihood is obtained the value of AIC for model p is readily calculated by

$$AIC(p) = -2 \max(\ln L) + 2n(p), \tag{11}$$

where n(p) is the effective number of parameters.

The results reported in the previous section were obtained by the method just described.

4. Various Model Comparisons in MDS

We are now in a position to overview applications of AIC in MDS in a wider perspective. This review is bound to be somewhat sketchy, because any in-depth discussion of the topic would necessarily involve substantive issues. Such an in-depth discussion is impossible due to space limitation. The reader is encouraged to refer to original papers for more details.

Selection of the best representation model provides a rich source of interesting model comparisons in MDS. The two common ones are (1) selection of dimensionality and (2) tests of specific structural hypotheses about stimulus configurations. Both of these have been illustrated in the previous example. Linear structural hypotheses can generally be expressed as

$$C'x = 0, \tag{12}$$

where x is the vector of stimulus coordinates (a matrix of stimulus coordinates strung out in a vector form), and C a known constraint matrix. This form of constraints is convenient for imposing equality constraints on certain stimulus coordinates, and hence for incorporating factorial structures. Takane (1981) and Sergent & Takane (1987), for example, used the above form of constraints in testing various hypotheses about perceived similarities among such stimuli as rectangles, parallelograms, circles, colors, etc. Takane, Bozdogan, & Shibayama (1987) and Takane (1987b) used the constraints in multi-sample cluster analysis (Bozdogan, 1986), that is, in testing the equality among criterion groups in discriminant analysis.

An alternative form of the constraints is stated as

$$x = Gw, \tag{13}$$

where w is a vector of weights, and G a matrix of "predictor" variables. This is analogous to regression analysis, where predictions on the dependent variable are "constrained" to be a linear function of the predictor variables. This form of constraints is sometimes more convenient, and is extensively used in ideal point discriminant analysis (Takane, 1987b, 1989a, 1989b; Takane, Bozdogan, & Shibayama, 1987), where points corresponding to rows of a data matrix are represented as a linear combination of the predictor variables. By manipulating G in various ways, we may also investigate effects of discretization of certain predictor variables, interactions among them, partialling out of the effects of one set of variables from the other, etc.

Constraints of the above form can be combined with (12). That is, the weight vector w in (13) may be further constrained by

$$C^{*\prime}w = 0. \tag{14}$$

where C* is a known constraint matrix analogous to C in (12). This can be used effectively, for example, in testing a significance of the effects of certain predictor variables in G.

The circularity constraint in the example discussed in section 2 can be handled by neither forms of constraints given above. It requires a re-expression of the stimulus coordinate matrix in terms of the polar coordinate system, and to impose certain equality constraints. See also Takane (1981).

There are other interesting comparisons of model specifications worth mentioning. They include the comparison between the euclidean distance model (1) and the city-block model (2) (Sergent & Takane, 1987), or more generally, selection of r in the Minkowski power distance model (3) (Takane & Sergent, 1983), and the comparison between the distance model and other models of similarity data (e.g., the feature matching model, trees, etc.). The latter models can often be expressed as linear models by constructing appropriate predictor matrices. The linear model used in section 2 is one such instance. Set $x_{ia} = 1$ if stimulus i owns feature (line segment) a, $x_{ia} = 0$, otherwise. Define the n(n - 1)/2 by A matrix of $|x_{ia} - x_{ja}|$. The vector of predicted distances is then obtained by a linear combination of the columns of this matrix. See Takane & Sergent (1983), Takane & Shibayama (1986, 1992) for more examples of this kind.

For the error model, the normal error model (6) and the log-normal error model (7) have been extensively compared. One general finding is that the former fits better to the data collected by the methods that involve comparisons among two or more similarities, such as the pair comparison method and the rank-order method (Takane, 1978, 1982; Takane & Carroll, 1981), while the latter fits better to the rating data (Ramsay, 1977; Takane, 1981). Variance component models, (8) and (9), can also be subjected to model comparison.

The response model constructed for each specific form of similarity data is supposed to simulate the process by which the particular form of the data are generated. A realistic model of this process plays a crucial role in constructing a successful MDS procedure. An important class of similarity data (e.g., stimulus identification data, pair comparison data, rank-order data) requires the subject to compare similarities and choose the one that best fits a criterion. Several choice models have been tried to capture the choice processes, and have been compared in terms of GOF to actual data sets. These include: (1) Takane & Shibayama (1992) compared Luce's (1959) similarity choice model and Nakatani's (1972) confusion choice model for stimulus identification data. The two models were found to fit about equally well in all cases attempted. (2) Takane & Shibayama (1992) also compared the exponential ($\exp(-d_{ij})$) and the Gaussian ($\exp(-d_{ij}^2)$) response strength models in Luce's similarity choice model. Characteristics of situations were identified which favored one model over the other. (3) Takane (1989a) compared Luce's choice model and Thurstone's successive categories model (Torgerson, 1958) for ordinal criterion variables (columns) in contingency table analysis. The former was found to fit slightly but consistently better than the latter.

5. Evaluation

AIC has been applied in a variety of model comparison situations in MDS. These applications are reasonably successful, although still limited in scope. In this section we first briefly discuss possible reasons for the apparent "success", and then point out some of the problems yet to be resolved.

The reasons for the "success" may be summarized into three points. First of all, the stimuli used are relatively simple, constructed mostly by manipulating a few known physical attributes. The data analyzed are collected very carefully in controlled laboratory environments. Secondly, an elaborate response model which closely simulates the data generation process has been constructed for each specific kind of similarity data. Thirdly, different distributional assumptions are incorporated under which different solutions can be obtained. The most appropriate assumption can be chosen empirically by comparing the GOF of the solutions.

Not everything is easy, however. There are difficult problems as well. In MDS, sample size is often too small to rely on the large-sample theory for ML estimation. Obtaining a large sample is not easy with the single-subject design. Employing the between-subjects design leads to another kind of difficulty (see below). Observations are assumed independent (see (10)), despite the use of multiple-judgment sampling within subjects. The problem of correlated observations is less serious with the single-subject design. It can also be mitigated by taking comparative judgments (Takane & Carroll, 1981) and by measuring asymptotic performance by providing enough training trials so that subject's performance no longer improves during a data collection session. Still, the assumption of independence is at best only approximately satisfied.

Simultaneous analyses of between-subjects data with multiple-judgment sampling within subjects are difficult due to systematic individual differences. If no provisions are made in the model to account for the differences, the problem of correlated observations becomes more pronounced. If, on the other hand, parameters are introduced in the model to account for the individual differences, another kind of difficulty arises. The number of parameters may increase indefinitely with the number of observations, and no asymptotic properties of ML estimates are obtained (Kiefer & Wolfowitz, 1956). There are two possible ways to get out of the difficulty. One is to examine the small sample behavior of ML estimates by the Monte-Carlo methods, and deduce appropriate correction factors in AIC as functions of the number of stimuli and the number of subjects (see Ramsay, 1980). The other is to make reasonable distributional assumptions on the individual differences parameters (Basu, 1977; Kalbfleisch & Sprott, 1970) and marginalize them out. Although this second approach is theoretically more appealing, it creates additional problems of specifying reasonable prior distributions on the individual differences parameters, and of developing an optimization procedure that may involve a large number of numerical integrations.

6. Latent Variable Models

We now turn to the second major area of psychometrics; i.e., latent variable models such as FA and ACOVS. FA postulates a set of latent variables to account for covariances (or correlations) among observed (manifest) variables. Let t denote a random vector of n manifest variables. The FA model is stated as

$$t = m + Gu + e, \tag{15}$$

where m is the mean vector (m = E(t)), G is the matrix of factor loadings, u and e are random vectors of common factors and error components, respectively, and

$$\begin{pmatrix} u \\ e \end{pmatrix} \sim N\left(0, \begin{bmatrix} M^2 & 0 \\ 0 & Q^2 \end{bmatrix}\right). \tag{16}$$

In the standard FA model, it is further assumed that $V(e) = Q^2$ is diagonal. This is called (linear) local independence (LI) assumption. Matrix M^2 is the variance-covariance matrix of common factors. Without loss of generality, it may be assumed that the common factors are uncorrelated, i.e., $M^2 = I$ (an identity matrix). It follows that

$$\Sigma = V(t) = GG' + Q^2. \tag{17}$$

FA obtains estimates of G and Q^2, given a sample estimate S of Σ. There is a rotational indeterminacy in G, so that the effective number of parameters in Model (17) is equal to $n(A + 1) - A(A - 1)/2$, where A is the number of common factors.

The following criteria are commonly used for estimating parameters in FA. The maximum likelihood (ML) method amounts to the minimization of

$$F_{ML}(\Sigma, S) = \ln|S| - \ln|\Sigma| + tr(S\Sigma^{-1}) - n. \tag{18}$$

The generalized least squares (GLS) criterion minimizes

$$F_{GLS}(\Sigma, S) = (1/2)\, tr(I - S^{-1}\Sigma)^2. \tag{19}$$

The above two criteria are asymptotically equivalent under the multivariate normality (MVN) of t. Furthermore, if (17) is correct, both $(N-1)\hat{F}_{ML}$ and $(N-1)\hat{F}_{GLS}$ (where N is the sample size, and a hat indicates the minimized value of the criterion) follows asymptotically chi-square with degrees of freedom equal $n(n + 1)/2 - n(A + 1) + A(A-1)/2$ under suitable regularity conditions. A generalization of GLS is the weighted least squares (WLS) criterion defined by

$$F_{WLS}(\Sigma, S) = (vec(S) - vec(\Sigma))W^{-1}(vec(S) - vec(\Sigma)), \tag{20}$$

where W is a positive-definite weight matrix, and the vec vectorizes nonduplicated elements of a symmetric matrix. When elements of W are taken to be covariances between elements of vec(S), F_{WLS} reduces to F_{GLS} (which in turn is asymptotically equivalent to F_{ML} under the MVN). By modifying the weights in W appropriately, however, F_{WLS} may be used under distributions other than the MVN, while still retaining its asymptotic equivalence to F_{ML} and F_{GLS} (Browne, 1982, 1984; Kano, Berkane & Bentler, 1990). Whichever criterion is employed, the value of AIC for model p can be calculated by

$$AIC(p) = (N - 1)\hat{F} + 2n(p), \tag{21}$$

where $\hat{F}$ is the minimized value of F_{ML}, F_{GLS}, or F_{WLS}, and $n(p) = n(A+1) - A(A-1)/2$. (Note that the values of AIC calculated from the different criteria are likely to differ for a finite sample.)

The basic FA model (15) can be generalized in various ways. We discuss two possibilities here for later reference. First, just as the covariance matrix is structured in FA, the mean vector m may also be structured. That is,

$$m = HCb, \tag{22}$$

where H and b are a known matrix and a vector of "predictors", respectively, and C a matrix of parameters to be estimated.

Secondly, a model analogous to (15) may be assumed for u as well. This leads to the second-order FA model or ACOVS (Jöreskog, 1970) in which a FA model is nested within another FA model. Let $u = Bs + w$ be the FA model for u, where s, w and B are analogous to u, e and G in (15), respectively, and

$$\begin{pmatrix} s \\ w \end{pmatrix} \sim N\left(0, \begin{bmatrix} K^2 & 0 \\ 0 & R^2 \end{bmatrix}\right). \tag{23}$$

Then,

$$t = m + G(Bs + w) + e, \tag{24}$$

and

$$\Sigma = G(BK^2B' + R^2)G' + Q^2. \tag{25}$$

In FA and ACOVS, AIC has been used to select the number of common factors (latent variables) and to test a variety of structural hypotheses for the mean vector and the covariance matrix.

7. ACOVS Pair Comparison Models

We discuss two interesting special cases of ACOVS with structured means. There is a phenomenon called similarity effect in pair comparison judgments. Similar stimuli are easier to compare, and consequently pair comparison judgments involving similar stimuli tend to be more extreme than those involving dissimilar stimuli, for a given difference in preference values. Takane (1980) and Heiser and de Leeuw (1981) independently proposed the so-called THL model that takes into account the differential comparabilities among stimulus pairs. This model maps stimuli into a multidimensional space, and defines the stimulus comparability as the reciprocal of the euclidean distance between the stimuli.

De Soete and Carroll (1983) later proposed a similar model, called wandering vector model (WVM), in which not only the stimulus comparabilities, but also preference values are functions of the stimulus configuration. Specifically, they are assumed given by the projections of the stimulus points onto a vector (called the mean subject preference vector) pointing in a particular direction in the space.

Takane (1987a; see also Takane & de Leeuw, 1987) generalized both THL model and WVM to accommodate systematic individual differences in preference judgments. He introduced a random vector pertaining to the systematic individual differences, which was then marginalized (integrated) out, and arrived at ACOVS formulations of the two models. We briefly describe them here for pairwise preference rating data. In the pairwise preference rating method, two stimuli are presented to subjects in each trial. The subjects are asked to indicate, on a relatively continuous rating scale, the degree to which they prefer one stimulus over the other. Let G be a design matrix for pair comparisons. Each row of G corresponds with a stimulus pair. If that pertains to stimuli i and j, the row has 1 and -1 in the ith and the jth positions and 0's elsewhere. An example of G is

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

for $n = 4$.

Let c be the mean stimulus preference vector, and let X be the matrix of coordinates of the stimulus points in a multidimensional space. Then, the ACOVS THL model may be written as

$$m = Gc, \tag{26}$$

and

$$\Sigma = G(XX' + K^2)G' + Q^2, \tag{27}$$

and the ACOVS WVM as

$$m = GXv, \tag{28}$$

with the same covariance structure as (27), where v is the mean subject preference vector. The WVM is a special case of the THL model in which c in (26) is constrained to be Xv.

8. An Example of ACOVS Analysis

The THL model and the WVM were fitted to an actual data set. The stimuli were nine celebrities: 1. Brian Mulroney (Prime Minister of Canada), 2. Ronald Reagan (Ex-President of the US), 3. Margaret Thatcher (Ex-Prime Minister of the UK), 4. Jacqueline Gareau (twice winner of the Boston Marathon in the woman's division), 5. Wayne Gretzky (NHL hockey player), 6. Steve Podborski (former champion of the World Cup downhill ski race), 7. Paul Anka, (male vocalist), 8. Tommy Hunter (country song singer), 9. Ann Murray (female vocalist). These stimuli were deliberately chosen to create heterogeneity in similarity among them to test the similarity effect in pair comparison judgments. There are three politicians, three athletes, and three entertainers/singers, and the stimuli in a same group are deemed more similar to each other than those in different groups. Subjects were 119 McGill undergraduates who were asked, for each of 36 pairs of stimuli, to indicate the extent to which they preferred one stimulus over the other on a 25-point rating scale.

Figure 4 presents the two-dimensional ACOVS solution from the THL model. (The ML method was used in all the analyses reported in this section.) This shows the plot of stimulus coordinates (the X matrix). The three groups of people used as stimuli nicely cluster together, as anticipated. This means that stimuli within a cluster are more comparable (having larger covariances) than those in different clusters. Estimated mean preference values are: 2.6(BM), 2.4(RR), 2.7(MT), -2.6(JG), 0.9(WG), 0.9(SP), -1.9(PA), -4.2(TH), and -0.9(AM). It seems that they are fairly highly correlated with a particular direction (lower right) in the space, although two least preferred stimuli, JG and TH, are somewhat displaced. This motivates to fit the WVM.

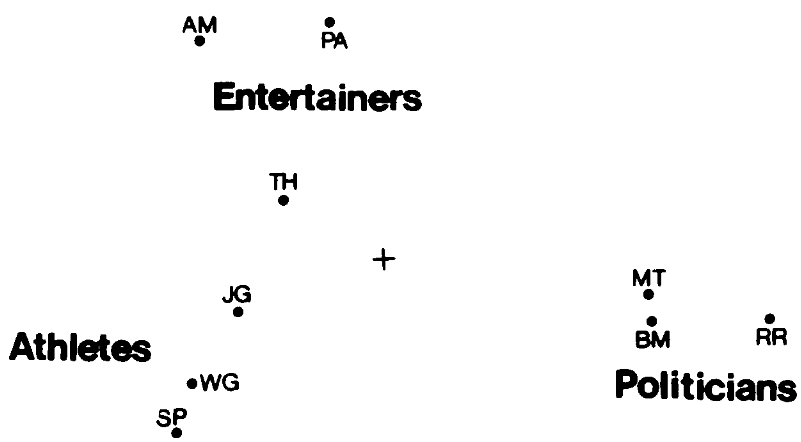

Figure 4. Two-factor solution from the THL model for the celebrity data.

Results from the WVM are displayed in Figure 5. This stimulus configuration is quite similar to the previous one, though clusters are perhaps less distinct. Estimated mean preference values (2.6 for BM, 2.5 for RR, 2.3 for MT, -2.0 for JG, 0.9 for WG, 0.9 for SP, -22.1 for PA, -3.4 for TH and -2.0 for AM) are now proportional to projections of the stimulus points onto the mean subject preference vector depicted in the configuration. JG and TH are now moved up somewhat to make them less preferable. The correlation between these preference values and those from the THL model was .976.

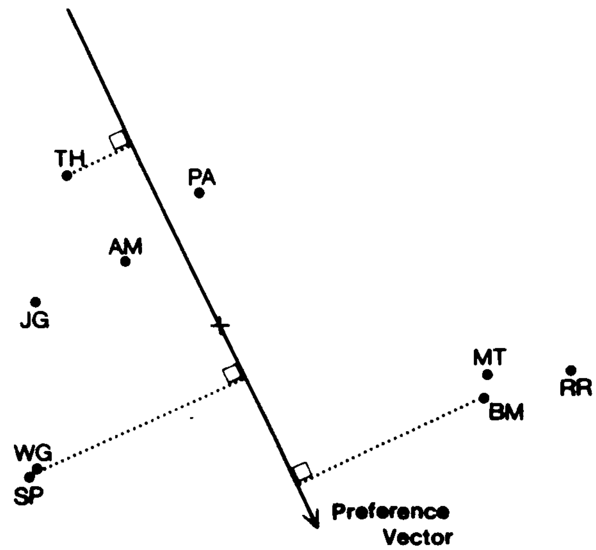

Figure 5. The same as Figure 4 from the WVM.

The question now is how well these models fit overall, and how they compare with each other. Two additional models were fitted to benchmark the models. One is a completely unstructured saturated model in which sample means and covariances are used as estimates of their population counterparts. The other model, called Model (0), is one in which means are modelled by Gc, but covariances are left unstructured. To summarize, the fitted models are:

1) Neither means nor covariances are structured. (Saturated Model)
2) Only means are structured by Gc. (Model 0)
3) In addition to 2) above, covariances are structured by (27). (Model 1: THL model)
4) In addition to 3) above, mean stimulus preference values are structured by c = Xv. (Model 2: WVM)

In 3) and 4) above, diagonal elements of Q^2 are all assumed equal to avoid improper solutions.

Results are reported in Table 2, which summarizes AIC. As it has turned out, Model (0) is the best fitting model. This means that the structure on means is acceptable, but covariances are rather poorly modelled by (27). We obtained three solutions each . the

THL model and the WVM, varying the number of factors from one to three, but none of them are anywhere near the GOF of Model (0).

		Model (1) (THL Model)	Model (2) (WVM)
Saturated Model (Unconstrained sample means & covariances)	AIC	6.3	
	n(p)	(702)	
Model (0) (Only structured means: Gc)	AIC	0.7*	
	n(p)	(674)	
dim = 1	AIC	224.8	306.4
	n(p)	(18)	(11)
dim = 2	AIC	183.7	226.9
	n(p)	(25)	(19)
dim = 3	AIC	183.0	187.9
	n(p)	(31)	(26)

*The minimum AIC solution.

Table 2. Summary of GOF statistics for the celebrity data.

Neither the THL model nor the WVM fits to the data very well. Obviously something is missing in these models, yet what is missing is not obvious. Both THL model and WVM are relatively well conceived models in that there is a strong theoretical reason for the models. Structures expected under these models have been clearly borne out in the resulting configurations (see Figures 4 and 5). Nonetheless the models provide rather crude approximations to the data.

9. Discussion

Latent variable models are currently very popular in social and behavioral science research. As demonstrated in the previous section, they provide a rich source of interesting model comparisons. Curiously, however, systematic applications of AIC in this area have been rather scarce (see, however, Ichikawa, 1988). This, combined with the finding in the previous section, may suggest some general difficulties in applying AIC to FA and ACOVS.

Many phenomena in social and behavioral sciences are extremely complicated involving so many intricate factors intertwined with each other. Models fitted are often oversimplified relative to the complexity of the phenomena. They capture some, but not all aspects of data structures, providing only crude approximations to the data. Models like FA and ACOVS (and MDS) tend to compensate for the lack of fit by increasing the number of latent variables, whereas the cause for the misfit may lie elsewhere. What is likely to happen then is that an increasingly more complex model tends to be chosen, as the sample size increases, until one of the following events occurs:

1. An improper solution (i.e., a solution with negative estimates of variances) is encountered, before a better model is found than the saturated model.

2. A better fitting proper solution is found, but some of the factors extracted are uninterpretable.

3. No better solutions are found than the saturated model.

Event 3 indeed took place in the previous example, when the sample size was subsequently increased from 119 to 501.

In FA and ACOVS subjects are usually taken as replications. This may be justified, if they are randomly drawn from a homogeneous population of subjects. However, this condition is rarely satisfied. Often, there are heterogeneous subsamples of subjects mixed together in a sample. This may cause non-MVN of the distribution. Under such circumstances, it may not be sensible to fit a single covariance matrix to the entire sample. Subjects may have to be stratified, and a separate covariance matrix fitted to each stratified sample. In practice, however, it may be difficult to identify the variable(s) by which the subjects are to be stratified. Mixture models (e.g., DeSarbo, Howard, & Jedidi, 1991) may be applied to avoid this difficulty. These models simultaneously "stratify" the subjects, and fit the covariance structure models. No initial stratification of the subjects is thus necessary. However, some difficulty arises in applying AIC to these models, as discussed by McLachlan & Basford (1988).

In FA and ACOVS the MVN distribution is almost always assumed. However, this assumption is often unrealistic. Micceri (1989), for example, examined distributions of 440 psychometric measures from achievement tests, attitude surveys, personality inventories, etc., and found that none of them passed the test of univariate normality at the prescribed significance level of .01. He concluded that in psychology the normal distribution was just "as improvable as the unicorn." Figure 6 displays empirical distributions of two of the variables from the previous example. These histograms were constructed from the augmented data. They both look very similar to one of the histograms presented by Micceri, and described as an "asymmetric, lumpy, multimodal distribution."

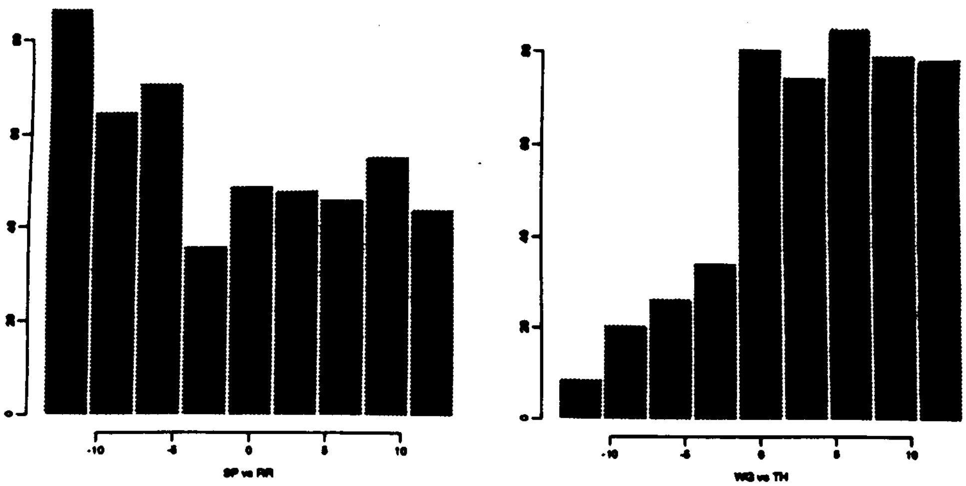

Figure 6. Histograms of two variables from the celebrity data.

The problem of non-normal distributions in FA and ACOVS has been somewhat alleviated by recent robustness studies on the normal likelihood (Anderson & Amemiya, 1988; Browne & Shapiro, 1988; Shapiro & Browne, 1987). There have also been attempts to relax the distributional assumption. Several procedures have been developed that require less stringent distributional assumptions than the MVN. These include the ADF (Asymptotically Distribution-Free) method (Browne, 1984), and those based on elliptical distributions (Browne, 1982; Kano, Berkane, & Bentler, 1990). While these developments are certainly encouraging, even a larger sample size is necessary for the asymptotic results to hold under the relaxed conditions. This brings us back to the state of affairs described at the beginning of this section.

FA and ACOVS typically assume linear local independence (LI). That is, manifest variables are assumed independent (uncorrelated) given the latent variables. This assumption is often problematic. It tends to give rise to improper solutions when a large number of latent variables have to be extracted. The problem is aggravated when the real cause for the large number of latent variables is due to other misfits between the model and the data. Akaike (1987) proposed an ingenious method to constrain estimates of uniqueness variances to take only positive values, thereby avoiding the improper solutions. An effort is yet necessary to make the method more widely available.

10. Concluding Remarks

This paper reviewed the use of AIC in two representative areas of psychometrics. The message of the review seems clear. The model comparison approach plays an important role in psychometrics, although, as has been pointed out, there are a number of difficult problems yet to be resolved as well. Obviously, a lot more efforts have to be expended for improving the quality of data and models.

Finally, a comment is in order on so-called sample-size independent GOF indices recently proposed for use in latent variable models (Bentler, 1990; McDonald, 1989; McDonald & Marsh, 1990; Steiger, 1990). These indices are all based on the "bias" incurred by approximating the true (saturated) model by a reduced model. Figure 7 depicts the relationships among AIC, the bias (Bias) and -2LLR (minus twice the log likelihood ratio between the estimates of the reduced and the saturated models). This figure was drawn by closely following the derivation of AIC in Bozdogan (1987). The saturated model, Θ_K^*, has K parameters, and the reduced model, Θ_k^*, k parameters with the corresponding ML estimates, $\hat{\Theta}_K$ and $\hat{\Theta}_k$, respectively. Symbols such as K and k indicate squared lengths of the sides of the prism. (To simplify the view, the figure was drawn in the identity metric rather than in the metric of the Fisher information matrix.) Asymptotic expectations were taken for the squared lengths of the sides of front and end triangles, while no expectations were taken for the sides and diagonals connecting the two triangles. The latter are estimates based on a particular sample, and vary over samples. -2LLR corresponds with the squared distance between $\hat{\Theta}_k$ and $\hat{\Theta}_K$. The bias (Bias) corresponds with the squared distance between Θ_k^* and Θ_K^*. AIC corresponds with the squared distance between $\hat{\Theta}_k$ and Θ_K^*. Let -2LLR = d. Then, by the Pythagorean theorem, Bias = d + k - K and AIC = d + 2k - K.

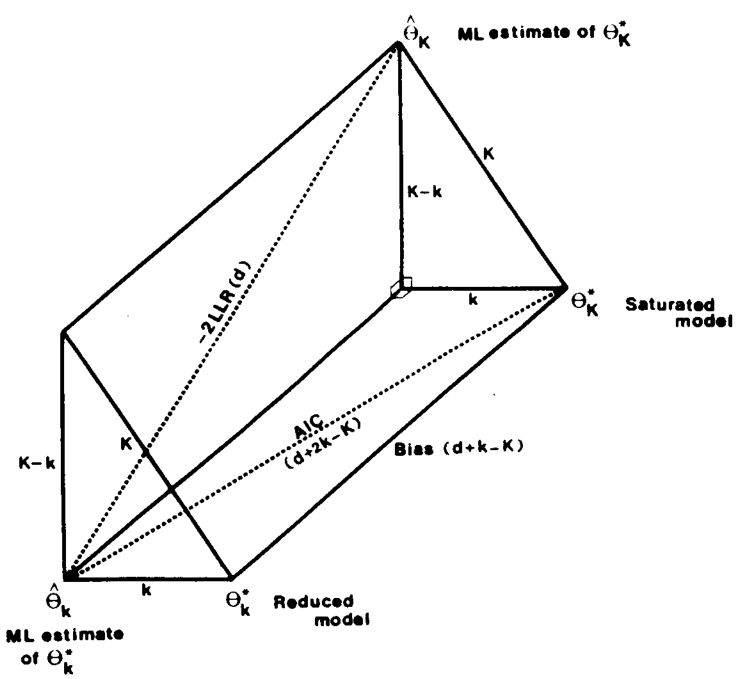

Figure 7. Geometric representation of AIC and the bias (noncentrality).

It can be seen from Figure 7 that Bias lies somewhere between -2LLR and AIC. However, it is known that it behaves much more like -2LLR (e.g., Cudeck & Henly, 1991). In particular, it decreases monotonically as k increases. Bias is thus similar to R^2 (the coefficient of determination) in multiple regression analysis, and should be treated like R^2. While I do not completely deny some descriptive value of R^2 as a GOF index (after all every output from multiple regression analysis carries this information), exclusive use of such an index is quite dangerous. With such an index it is possible to select a model with an excellent fit to the data at hand, but with little predictability for future data. The GOF to the data at hand can always be improved by increasing the number of parameters. For better predictions, however, it may be better to select a model with fewer parameters. This is precisely why AIC was called for in the first place. A sensible compromising view is presented in Cudeck & Henly (1991) and by Browne & Cudeck (1992).

References

Akaike , H. (1973). Information Theory and the Maximum Likelihood Principle, *Second International Symposium on Information Theory*, (eds. B.N. Petrov and F. Csaki), Akademiai Kiado, Budapest.

Akaike, H. (1987). Factor Analysis and AIC, *Psychometrika*, 52, 317-332.

Anderson, T.W., and Amemiya, Y. (1988). The Asymptotic Normal Distribution of Estimators in Factor Analysis Under General Conditions, *Annals of Statistics*, 16, 759-771.

Basu, D. (1977). On Elimination of Nuisance Parameters. *Journal of the American Statistical Association*, 72, 355-366.

Bentler, P. (1990). Comparative Fit Indices in Structural Equation Models, *Psychological Bulletin*, 107, 238-246.

Bozdogan, H. (1986). Multi-Sample Cluster Analysis as an Alternative to Multiple Comparison Procedures. *Bulletin of Informatics and Cybernetics*, 22, 95-130.

Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytic Extensions, *Psychometrika*, 52, 345-370.

Browne, M.W. (1982). Covariance Structures, *Topics in Applied Multivariate Analysis*, (ed. D.M. Hawkins), 72-141, Cambridge University Press, Cambridge.

Browne, M.W. (1984). Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures, *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Browne, M.W., and Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods and Research*, 21, 230-258.

Browne, M.W., and Shapiro, A. (1988). Robustness of Normal Theory Methods in the Analysis of Linear Latent Variable Models. *British Journal of Mathematical and Statistical Psychology*, 41, 193-208.

Carroll, J.D., and Chang, J.J. (1970). Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of "Eckart-Young" Decomposition. *Psychometrika*, 35, 283-319.

Coombs, C.H. (1964). *A Theory of Data.* Wiley, New York.

Cudeck, R., and Henly, S.J. (1991). Model Selection in Covariance Structures Analysis and the "Problem" of Sample Size: A Clarification, *Psychological Bulletin*, 109, 512-519.

DeSarbo, W.S., Howard, D.J., and Jedidi, K. (1991). MULTICLUS: A New Method for Simultaneously Performing Multidimensional Scaling and Cluster Analysis. *Psychometrika*, 56, 121-136.

De Soete, G., and Carroll, J.D. (1983). A Maximum Likelihood Method for Fitting the Wandering Vector Model. *Psychometrika*, 48, 553-566.

Heiser, W., and de Leeuw, J. (1981). Multidimensional Mapping of Preference Data, *Mathematique et Sciences Humaines*, 19, 39-96.

Ichikawa, M. (1988). Empirical assessment of AIC procedure for model selection in factor analysis. *Behaviormetrika*, 24, 33-40.

Jöreskog, K.G. (1970). A General Method for Analysis of Covariance Structures, *Biometrika*, 157, 239-251.

Kalbfleisch, J.G., and Sprott, D.A. (1970). Application of Likelihood Methods to Models Involving a Large Number of Parameters. *Journal of the Royal Statistical Society*, Series B, 32, 175-208.

Kano, Y, Berkane, M., and Bentler, P.M. (1990). Covariance Structure Analysis with Heterogeneous Kurtosis Parameters, *Biometrika*, 77, 575-585.

Kiefer, J., and Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *Annals of Mathematical Statistics*, 27, 887-906.

Luce, R.D. (1959). *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.

McDonald, R.P. (1989). An Index of Goodness-of-Fit Based on Noncentrality, *Journal of Classification*, 6, 97-103.

McDonald, R.P., and Marsh, H.W. (1990). Choosing a Multivariate Model: Noncentrality and Goodness of Fit, *Psychological Bulletin*, 105, 156-166.

McLachlan, G.J., and Basford, K.E. (1988). *Mixture models*. New York: Marcel Dekker.

Micceri,, T. (1989). The Unicorn, the Normal Curve, and Other Improvable Creatures. *Psychological Bulletin*, 105, 156-166.

Nakatani, L.H. (1972). Confusion-Choice Model for Multidimensional Psychophysics, *Journal of Mathematical Psychology*, 9, 104-127.

Ramsay, J.O. (1977). Maximum Likelihood Estimation in Multidimensional Scaling, *Psychometrika*, 42, 241-266.

Ramsay, J.O. (1980). Some Small Sample Results for Maximum Likelihood Estimation in Multidimensional Scaling. *Psychometrika,* 45, 141-146.

Ramsay, J.O. (1982). Some Statistical Approaches to Multidimensional Scaling Data. *Journal of the Royal Statistical Society*, Series A, 145, 285-312.

Sattath, S., and Tversky, A. (1977). Additive Similarity Trees, *Psychometrika*, 42, 319-345.

Sclove, S.L. (1987). Application of Model-Selection Criteria to Some Problems in Multivariate Analysis. *Psychometrika*, 52, 333-343.

Sergent, J., and Takane, Y. (1987). Structures in Two-Choice Reaction Time Data, *Journal of Experimental Psychology, Human Perception and Performance*, 13, 300-315.

Shapiro, A., and Browne, M.W. (1987). Analysis of Covariance Structures Under Elliptical Distributions, *Journal of the American Statistical Association*, 82, 1092-1097.

Steiger, J.H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, 25, 173-180.

Takane, Y. (1978). A Maximum Likelihood Method for Nonmetric Multidimensional Scaling: 1. The Case in Which All Empirical Pairwise Orderings are Independent-Theory and Applications, *Japanese Psychological Research*, 20, 7-17 and 105-114.

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. (1981). Multidimensional Successive Categories Scaling: A Maximum Likelihood Method, *Psychometrika*, 46, 9-28.

Takane, Y. (1982). The Method of Triadic Combinations: A New Treatment and Its Applications, *Behaviormetrika*, 11, 37-48.

Takane, Y. (1984). Basic Concepts and Applications of Multidimensional Scaling, *Proc. 1984 Japan Applied Statistical Society Meeting*, 30-41.

Takane, Y. (1987a). Analysis of Covariance Structures and Probabilistic Binary Choice Data, *Communication and Cognition*, 20, 45-61. (Also, *New Developments in Psychological Choice Modelling* (eds. G. De Soete et al.), 139-160, North Holland, Amsterdam, 1989.)

Takane, Y. (1987b). Analysis of Contingency Tables by Ideal Point Discriminant Analysis, *Psychometrika*, 52, 493-513.

Takane, Y. (1989a). Ideal Point Discriminant Analysis and Ordered Response Categories, *Behaviormetrika*, 26, 31-46.

Takane, Y. (1989b). Ideal Point Discriminant Analysis: Implications for Multiway Data Analysis, *Multiway '88*, (eds. R. Coppi et al.), 287-299, North Holland, Amsterdam.

Takane, Y., Bozdogan, H., and Shibayama, T. (1987). Ideal Point Discriminant Analysis, *Psychometrika*, 52, 371-392.

Takane, Y., and Carroll, J.D. (1981). Nonmetric Maximum Likelihood Multidimensional Scaling from Directional Rankings of Similarities, *Psychometrika*, 46, 389-405.

Takane, Y., and de Leeuw, J.(1987). On the relationship Between Item Response Theory and Factor Analysis of Discretized Variables, *Psychometrika*, 52, 393-408.

Takane, Y., and Sergent, J. (1983). Multidimensional Scaling Models for Reaction Times and Same-Different Judgments, *Psychometrika*, 48, 393-423.

Takane, Y., and Shibayama, T. (1986). Comparison of Models for Stimulus Recognition Data, *Multidimensional Data Analysis* (eds. J. de Leeuw et al.), 119-148, DSWO Press, The Netherlands (with discussion).

Takane, Y., and Shibayama, T. (1992). Structures in Stimulus Identifiçation Data, *Probabilistic Multidimensional Models of Perception and Cognition* (ed. F.G. Ashby), 335-362, Lawrence Erlbaum Associates, Hillsdale, NJ.

Torgerson, W.S. (1558). *Theory and Methods of Scaling*. Wiley, New York.

Tversky, A. (1977). Features of Similarity, *Psychological Review*, 84, 327-352.