Ann. Rev. Psychol. 1984. 35:55–81

SCALING

Forrest W. Young

Psychometrics Laboratory, University of North Carolina, Chapel Hill, North Carolina 27514

CONTENTS

INTRODUCTION

The purpose of scaling is to quantify qualitative data. Scaling procedures attempt to do this by using rules that assign numbers to qualities of things or events. A quality of a thing or event is an attribute (characteristic) of the thing or event observed in circumstances assumed by the observer to be qualitative. Scaling replaces qualitative observations about the attribute of the thing or event with numeric measures of the attribute of the thing or event.

This is my definition of scaling: Scaling is the process that uses rules to assign numbers to attributes of things or events observed in circumstances

0066-4308/84/0201-0055$02.00

assumed by the observer to be qualitative. Scaling produces measurements, and these measurements are called scale values.

My definition of measuring is only very slightly different: Measuring is the process that uses rules to assign numbers to attributes of things or events observed in circumstances assumed by the observer to be quantitative. Measuring produces measurements.

Note that scaling and measuring both produce measurements by applying rules to assign numbers to observed attributes of things or events. The difference is that scaling derives measurements from qualities, whereas measuring derives measurements from quantities. How do we know if the observation circumstances are quantitative or qualitative? We don't know. Rather, we must assume one or the other. In fact, the distinction does not reside in the observation circumstances, but is in the mind of the observer.

My review concentrates on developments which fall within the definition of scaling just given. Specifically, the review focuses on rule-based processes to attach numbers to data which are at least in part qualitative. Contributions are reviewed which involve: (*a*) models (rules) used in scaling; (*b*) algorithms (rule-based processes) for applying models to qualitative data; and (*c*) applications of scaling to empirical situations. No attempt has been made to provide complete coverage. Rather, I have covered what I believe to be the most important developments in scaling during the review period. However, because of the vast application of some developments, applications have been slighted in this review.

My review specifically omits certain models and algorithms: factor analysis and regression analysis are covered only when the developments apply to qualitative data. Latent trait theory, test theory, and cluster analysis are completely excluded from the review, even though these developments often apply to qualitative data. These topics are periodically reviewed elsewhere.

DATA THEORY

Scaling techniques may be structured into two very broad categories according to the kind of data being scaled: The data may be *multivariate* data or *dissimilarity* data. Multivariate data have observations that are (qualitative or quantitative) values on one or more variables (thus including univariate data as well). Dissimilarity data have observations that are (qualitative or quantitative) dissimilarities (or similarities) between pairs of things.

This structure may also be seen according to the model being used to scale the data: scaling techniques for quantifying multivariate data use *linear* models (including bilinear and multilinear). Principal components and multiple regression are examples. Scaling techniques for quantifying qualitative dissimilarity

data use *distance* models. Multidimensional scaling and unfolding are examples.

The two-level structure just given forms the broad outline of this review. The next section of the review is on scaling dissimilarity data via distance models, and then comes a section on scaling multivariate data via linear models. The remainder of the present section of the review briefly introduces and defines some data theory terms used in the rest of the review. The terms are from a data theory that I first discussed in a manuscript circulated in the scaling community in 1975, which will appear in Young & Hamer (1984). Condensations of this data theory have appeared in Young & Lewyckyj (1979), Young et al (1980), Young (1981), Schiffman et al (1981), and Young (1983a). The theory rests on the following basic assumptions:

1. *Data are always categorical:* Data are obtained by a classification process that assures we can decide whether two observations are empirically equivalent. Equivalent observations form an observation category. This assumption underlies the measurement aspects of the data theory.
2. *Data are always designed:* Data are always obtained in an empirical situation that has an objective design. This assumption underlies the empirical aspects of the data theory.
3. *Data are always modeled:* Data are always obtained in the context of a specific model of the empirical situation. This assumption underlies the model aspects of the data theory.

Measurement Aspects

The categorical assumption that underlies data theory provides a very nice organizing principle for three important measurement concepts: measurement process, which concerns the relationships among all of the observations within a single data category; measurement level, which concerns the relationships among all of the observations between different data categories; and measurement conditionality, which concerns the relationships within sets of observation categories.

There are two types of measurement processes: discrete and continuous. The discrete process implies that we believe that all of the observations in a category should be represented by a single number after they have been scaled. Alternatively, the continuous process implies that we believe that all observations in a category should be represented by a bounded interval of real numbers after they have been scaled. Notice that the measurement process assumption concerns what happens within a single observation category.

The familiar measurement level notions correspond to assumptions about what happens between observation categories. Levels of measurement differ from each other in the ways that restrictions are imposed on the numbers that

can be assigned to different observation categories. There are a variety of restraints which could be discussed, but only three are needed to satisfy the characteristics of the four familiar measurement levels.[1]

For the nominal level, there are no measurement level restraints. The characteristics of this level are completely specified by the within-category restrictions implied by the chosen measurement process. For the ordinal level we require, in addition to the process restraints, that the real numbers assigned to observations in different categories reflect the order of the empirical observations. For the numeric (quantitative) levels of measurement (which include interval and ratio, among others) we require, in addition to the process constraints, that the real numbers assigned to observations in different categories be related to each other by some form of numeric function.

The final type of restraints placed on the numers used to scale the observation categories concerns relationships among sets of categories. It may be, for a particular set of data, that one measurement level and process apply to all of the data. Such data are called unconditional. On the other hand, it may be that different levels and processes apply to different portions of the data. These data are called conditional.

Empirical Aspects

One important aspect of the data theory is the shape of the data: data may be either square or rectangular. Square data are in matrices whose rows and columns refer to the same set of things. Usually (perhaps always), square data are "relational" data: data which indicate the degree of relation between the rows and columns of the matrix. Such data include distances, dissimilarities, similarities, etc. Square data can be symmetric or asymmetric. Rectangular data are in matrices whose rows and columns refer to two different sets of things. Thus, the matrix is rectangular in shape. These data are usually called "multivariate" data because the (multiple) columns often refer to variables.

The shape of a set of data can be reexpressed in terms of the number of ways and the number of modes of a set of data. The number of ways of any data matrix is always two (its rows and columns). The shape of the matrix actually refers to whether the two ways are distinct. The number of distinct ways is the number of modes. Thus, a square matrix is one-mode and a rectangular matrix is two-mode. Whereas a *data matrix* is always two-way, *data* may be more than two-way. Multi-way data are organized into several matrices and may be called three-way data, four-way data, etc.

[1]To be faithful to my distinction between scaling and measuring (made in the opening paragraphs) I should refer to nominal and ordinal as levels of "scaling," and interval and ratio as levels of "measuring." However, the phrase "levels of measurement," which is always used to refer to all four levels, is in such common use that I will use it instead of the other more precise terms.

My classification of recent developments in scaling involves the distinction between multivariate data and dissimilarity data. Multivariate data are rectangular data which consist of repeatedly observed values on many variables. The two ways of the data are the repeated observations (oftentimes people) and the variables. Dissimilarity data are usually square, but may also be rectangular. When square, the two ways are usually some type of stimuli. When rectangular, the two ways are usually people and variables (as with multivariate data). Usually, multivariate or rectangular dissimilarity data are two-way and row or column conditional, whereas square dissimilarity data are usually three-way and matrix-conditional.

The fundamental difference between multivariate and dissimilarity data is in the nature of the individual datum. For multivariate data the basic datum is the amount of each variable for each observation. For dissimilarity data the basic datum is the distance (or proximity) between the row-thing and the column-thing. [I use "dissimilarity" generically to include proximity (similarity) data as well as distance (or difference) data.]

Model Aspects

It may seem a bit strange to include the nature of models as one of the three major organizing principles of a data theory. After all, doesn't data theory concern data and not models? Actually, the answer is no; data theory concerns both the data and the model of the data. In particular, it is my view that data in themselves do not possess measurement characteristics. Rather, the measurement characteristics which appear to be possessed by a particular set of data are actually dependent on the interaction of that data with the model chosen to describe the data.

When a set of data are analyzed by some model, the analysis necessarily assumes that the data have certain specific measurement characteristics. As suggested by Takane et al (1977), empirical information about the most appropriate measurement assumptions can be obtained by repeatedly analyzing the data several times, each time changing only the measurement assumptions. But as they point out, the measurement assumptions which appear to be most appropriate necessarily appear this way in a context created by the data analysis model. When a different model is used it can appear that different measurement assumptions are most appropriate. The apparent measurement characteristics of a set of data interact with the chosen data analysis model.

Although many different models are used in data analysis, I confine my review to three broad model classes: the linear, bilinear, and distance models. I confine my review to these three model classes because they are the ones which have been used in most recent developments in scaling. The next two sections cover these three model classes.

SCALING DISSIMILARITY DATA VIA DISTANCE MODELS

This section reviews recent developments in multidimensional scaling and unfolding: scaling methods which use distance to model dissimilarity data. The distinction between multidimensional scaling (MDS) and multidimensional unfolding (MDU) is in the data analyzed. MDS is always based on square dissimilarity data, whereas MDU is always based on rectangular dissimilarity data. However, they both share in common the use of distance models.

For example, the simplest form of MDS involves Euclidean distances between all the members of one set of points:

$$d_{ij}=\{(x_i-x_j)(x_i-x_j)'\}^{1/2},$$

where the x_i and x_j (the parameters of the model) are row vectors each having r elements that specify the location of points i and j in an r-dimensional Euclidean space. The x_i and x_j are the ith and jth rows of the stimulus coordinates matrix X, an n points by r dimensions matrix. Notice that the one set of points (X) is used to model the one mode of the square data. Similarly, the simplest type of MDU involves Euclidean distances from each member of one set of points to each member of another set:

$$d_{ij}=\{(y_i-x_j)(y_i-x_j)'\}^{1/2},$$

where x_j is the jth row of a coordinates matrix X, and y_i is the ith row of a second coordinates matrix Y. Notice that the two sets of points (X and Y) model the two modes of the rectangular data.

There are four major streams of development in MDS during the review period. These are: 1. the continued generalization of MDS models; 2. the introduction of confirmatory (constrained) MDS models and algorithms; 3. the introduction of maximum likelihood MDS models and algorithms; and 4. the improvement of least squares MDS algorithms. Each of these developments is discussed in the following sections, as is multidimensional unfolding.

Generalized Multidimensional Scaling

One research trend during the review period was a continuation of interest in generalizing the very popular weighted Euclidean model discussed by Bloxom (1968), Horan (1969), and Carroll & Chang (1970). (The model is usually known as the INDSCAL model.) Much of this work has continued the early (and unpublished) generalizations of the weights proposed by Harshman (1972) and by Carroll & Chang (1972). During the review period, no work has

generalized the Euclidean aspect of the model. Also, there was no work that continued the earlier extension of MDS to Minkowski models.

The first work published during the review period was by Bloxom (1978), who proposed what he called the generalized Euclidean model (GEM):

$$d_{ijk}=\{(x_i-x_j)W_k(x_i-x_j)'\}^{1/2},$$

where the three subscripts on d_{ijk} refer to the three ways of the data, i and j for the one mode of the basic square matrix (usually stimuli) and k for the third mode (usually subjects). In Bloxom's paper, W_k is positive semidefinite, with the rank of W_k controlled by the user just as he/she controls the dimensionality of X. Bloxom proposes decomposing W_k into

$$W_k=V_kV_k',$$

where V_k and V_k are the upper and lower triangular gram factors of the weights for subject k.

The GEM was also proposed, independently, at about the same time by Young & Lewyckyj (1979) but was not published until recently (Young 1984). The name Principal Directions Scaling (PRINDSCAL) reflected our choice to decompose W_k into its principal components. Note that for either decomposition, V_k can then be applied to the group's stimulus space X to obtain an individual's space X_k by the formula

$$X_k=XV_k .$$

The space X_k is a subspace of the overall space X, and its dimensions constitute the subject's principal dimensions. When oriented in a principal components orientation, as in PRINDSCAL, each successive dimension accounts for as much variance in the subject's data as possible. The space X_k is the subject's principal subspace in the maximum variance sense.

Interesting applications of GEM appear in Bloxom (1978), Young (1984), Jones & MacCallum (1984), Forsyth (1984), Dunn & Harshman (1982), and Easterling (1984). The model is discussed by Young (1982), Ramsay (1982b), and by Lingoes & Borg (1978b). Several authors propose the model in the context of constrained MDS (see below).

Despite great interest in this model, two problems emerge: (*a*) all of the algorithms that have been proposed are still essentially experimental in nature, none having been fully investigated; (*b*) very few applications have appeared, and it may be that the model is too complex to be of much use to those who are not experts on MDS models. The ultimate usefulness of this line of development remains to be seen.

Lingoes & Borg (1978b) proposed an interesting extension of the weighted Euclidean model which they call the perspective model. The model posits that a subject has a unique perspective on the stimulus space X which translates it to a new origin and then dilates the space around the new origin. Algebraically, this is done by adding a constant vector a_k (having r elements, one for each dimension) to X and then premultiplying the translated space by a diagonal weight matrix U_k (an n by n matrix having a row and column for each stimulus). Thus the individual's perspective on the common space is defined as:

$$X_k = U_k(X - 1A_k),$$

where 1 is an n-element column vector of 1s. The Euclidean distances for the subject are calculated in the usual way.

This model can be very useful in, for example, the cognitive geography situation in which people judge the perceived distance between geographic locations. It should be useful here because it is well known that people subjectively expand geographic space around a point with which they are familiar. Unfortunately, no convincing applications are presented by Lingoes & Borg (1978b), and their algorithm is very cumbersome and does not appear to be one that will be very useful. Thus, the final reviews are not in on this development either.

Finally, two somewhat eclectic and very interesting developments should be mentioned. One of these (Holman 1978) is a "completely nonmetric" multidimensional scaling method. By this Holman means that the method is one which (*a*) only uses the order of the dissimilarities and (*b*) does not posit a specific distance metric in which to construct a stimulus space. The first characteristic is common, of course. It is the second characteristic that is eclectic. The notion that replaces a distance metric is "betweeness": stimulus j should be located between stimuli i and k on every dimension if i and k are the farthest apart of the three stimuli. Holman presents numerous convincing examples of his method.

The other somewhat eclectic development (Takane 1980a) is a novel way to scale a particular type of dissimilarity data called sorting data. These data are obtained when subjects are asked to sort a set of stimuli into as many groups as they wish, where the groups consist of "similar" stimuli, and "dissimilar" stimuli are placed in different groups. As mentioned by the author, this type of similarity data is particularly appealing in many common situations. The method is very similar to correspondence analysis, and has the advantages of (*a*) being nonmetric and (*b*) constructing a joint space having both stimulus points and group centroids for each subject. Thus, the model is an individual differences model. Takane presents two interesting examples which show that for this type of data his method is quite useful.

Constrained Multidimensional Scaling

Another major trend in the MDS literature is the introduction of constraints on the parameters of the model. As mentioned in the previous section, the simplest form of MDS occurs when we have two-way, square dissimilarities from just one subject. The analysis invokes the Euclidean model to define distances between all points in a set of points. The parameters of the model are the coordinates x_i and x_j that give the location of points i and j in the Euclidean space. The points have a configuration represented by the complete matrix of coordinates X.

As pointed out by de Leeuw & Heiser (1980) in their lucid review of the use of constraints in MDS, when the data are three-way the most straightforward generalization of the preceding analysis is to apply it to each dissimilarity matrix separately. When this is done we obtain completely independent configuration matrices X_k for each subject k. Each configuration X_k is totally unconstrained by any possible relationship to other X_ks.

A much more interesting way of analyzing such data is via constraints of the form $X_k = XW_k$ where X is now a common stimulus configuration presumed to underlie every subject's judgments, and where W_k is a diagonal matrix of weights for each subject. This is, of course, the INDSCAL model discussed above but discussed from the viewpoint of constraints on X_k. Furthermore, W_k can be generalized to other forms, introducing different constraints on the X_k.

The form $X_k = XV_k$, where V_k are the factors (of one type or another) of a positive, semidefinite, nondiagonal W_k, is one of the generalizations mentioned in the previous section. This model has been discussed from an explicitly constrained MDS viewpoint by Bloxom (1978), Young & Lewyckyj (1979), de Leeuw & Heiser (1980), Carroll et al (1980), and by Young, et al (1983). In these developments, the matrix X is assumed to be a known matrix of experimental design coefficients or other a priori information about the stimuli, and what is desired is to constrain the analysis of the dissimilarities according to the design or other information. Here V_k can be looked on as the coefficients which constrain the X_k to be linear combinations of the information in X.

Young, et al (1983) present the linear constraints $X_k = X\dot{V}_k$ as an analog of multiple regression in which X is the independent information on multiple measurements for each observation (just as in multiple regression) and the dependent information is dissimilarities between the observations instead of a single value for an observation. Two very interesting applications of this notion are found in Forsyth (1984) and in Jones & MacCallum (1984).

Bentler & Weeks (1978, Weeks & Bentler 1982) develop a different class of constraints. In their work the coordinate x_{ia} can be fixed to an a priori value ($x_{ia} = c_{ia}$), can be proportional to the value of another coordinate ($x_{ia} = c_{ia}x_{jb}$), or can be free to take on an unconstrained value. They discuss the interesting special cases of coordinates being set to zero or being proportional to externally

provided information, and the case of stimuli having unique dimensions. These types of constraints appear in more primitive form in several generally distributed algorithms and in the published work of de Leeuw & Heiser (1980), Takane (1978a, 1981), Takane & Carroll (1981), and Bloxom (1978). Superb applications appear in the work of Takane, especially in 1981. Circular constraints are the focus of work by Lee & Bentler (1980), and ordinal constraints are discussed by Noma & Johnson (1977). These last two notions are also discussed briefly by de Leeuw & Heiser (1980).

Maximum Likelihood Multidimensional Scaling

A major trend during the review period is the development of maximum likelihood multidimensional scaling. There are two major lines of development, one by Ramsay and collaborators (1977, 1978a,b, 1980a,b; Winsberg & Ramsay 1981), the other by Takane and collaborators (1978a,b, 1980b, 1981, 1982a,b, Takane & Carroll 1981, Takane & Sergent 1983). Zinnes & MacKay (1983) have also worked on the topic.

These lines of development are similar but not the same. The work of Ramsay and his collaborators is fundamentally metric since they assume that the error process introduces erroneous dissimilarity values. The work of Zinnes and MacKay is similar in this regard. On the other hand, the work of Takane and his collaborators is fundamentally nonmetric since they assume that the error process introduces violations of monotonicity.

Ramsay discusses a variety of transformations of the data, especially power transformations and monotone splines. The splines permit monotonic transformations of the dissimilarities, but the data are still fundamentally at the interval or ratio levels of measurement. The data must be matrix conditional and may be asymmetric. Takane discusses optional constraints on the data transformations which allow him to deal with data at the ordinal, interval, or ratio levels of measurement. He also discusses asymmetry and a wide variety of conditionality situations. It does not appear that Zinnes and MacKay permit any transformation of the data in their work; thus they apparently assume that the data are ratio (and perhaps interval). Their data may be unconditional or matrix conditional and may be asymmetric.

Ramsay extends his developments to a very wide range of distance models, including GEM. He explicitly discusses (1982b) the simple and replicated Euclidean models, the INDSCAL-type model, and the full and reduced rank GEM models. The developments provided by Takane and by Zinnes and MacKay are restricted to the simple and replicated Euclidean models.

The motivation for developing maximum likelihood MDS is clearly presented by Ramsay in his first paper (1977). It is his view that in almost all data analysis there is some curiosity about the way the observed data vary around the fitted values. He asserts that "a maturing of a data analysis technology usually

brings a desire for . . . an explication of the error model involved in the fitting process."

Maximum likelihood satisfies this desire in a way which many feel has one major advantage: the approach changes multidimensional scaling from a descriptive tool into an inferential tool. Associated with this change is the introduction of significance tests. In particular, if we feel comfortable adopting certain assumptions, then maximum likelihood multidimensional scaling allows significance tests to determine the appropriate dimensionality, the appropriate MDS model, and the appropriate error model. Also, this approach provides confidence regions for the stimuli and, with weighted models, for subjects.

The crucial assumption that we must accept in order to believe the conclusions suggested by the significance test is the specific nature of the error model posited by the chosen maximum likelihood MDS method. If the model is an appropriate mirror of the actual error processes active in the empirical situation, then the significance tests may be meaningful. If it is not, then they are not. This aspect is particularly important since the three teams of researchers working on MLMDS have proposed five error models, Ramsay and Takane two each and Zinnes and MacKay another.

Throughout their work, Ramsay and Takane actually posit the same pair of error models, but they differ in their assumptions about how these models influence the data. They both argue that error is normally distributed and is either added to or multiplied times the true distance. Thus, their procedures are alike in providing the user with the choice of assuming that the dissimilarity judgments are normally (additive model) or log-normally (multiplicative model) distributed about the true distance. However, the two procedures differ in the implied effect of the two assumptions: Ramsay's implies that the data are erroneous dissimilarity *values*. Takane's implies that the data are erroneous dissimilarity *orders*.

In contrast, Zinnes and MacKay make the fundamentally different assumption that the *stimuli* (not their distances) are normally perturbed during the subject's judgment process. This implies that the dissimilarities have noncentral chi-square distribution around the true distance. Like Ramsay, Zinnes and MacKay then proceed to assume that the data are erroneous dissimilarity values (not orders).

Although the mathematical virtuosity of the preceding developments is truly impressive, I question just what has been gained. From the viewpoint of the user, what seems to have been gained is that the user no longer has to decide on the "proper" distance model (dimensionality and weighting scheme) because the significance tests automate the decision. Furthermore, with two of the algorithms the program even indicates which error model is best. But this state of affairs may lead the unsophisticated user into a false state of confidence.

Of particular concern are those users who do not know that selecting a specific computer program implies selecting a specific choice of error models. What such a user knows is that the chosen program provides guidelines, via a series of significance tests, to choose the "proper" error and distance models. But what this user does not know is that what appears to be the "proper" error and distance models depends on the program doing the analysis. This occurs because (*a*) no two programs incorporate the same error models, and (*b*) the results of the significance tests for selecting the distance and error models will differ between programs because of the different error models. Unfortunately, there is no way to use internal consistency as a guideline because none of the approaches provide such a test. Consequently, these methods and their associated significance tests and confidence regions may induce a false sense of confidence in the results.

Furthermore, there are several technical problems with the significance tests. One technical problem is the assumption made by Ramsay and by Zinnes and MacKay that the dissimilarity judgments are independent. Their likelihood functions are critically dependent on this assumption. If this assumption is violated, then the significance levels are wrong. This problem has been addressed directly by Takane (1978a). He has proposed several different likelihood functions, each one specifically designed for the dependencies induced into the dissimilarity judgments by a specific experimental design, so his work is not subject to this criticism.

There is also the question, in the work of all three researchers, of whether the samples are large enough to justify the inferential claims. The question arises from the fact that hypothesis testing is only justified for very large sample sizes because the tests are based on asymptotic characteristics of maximum likelihood estimation. In many cases the sample size (number of subjects usually) is small, thus suggesting that the tests are not asymptotic.

As pointed out by Ramsay (1980b), the sample size problem is exacerbated for those models which have parameters for subjects, as in the very popular INDSCAL model. Clearly, for these models a large sample size can never imply asymptotic estimates of the parameters because the number of parameters increases without bound as the number of subjects increases. For these models the inferential aspects will never be completely appropriate. The importance of this problem is unknown, although Ramsay (1980b) suggests it is minor. Note that this is not a problem for the simplest model discussed by Ramsay, nor for any of the work of Takane and of Zinnes and MacKay, because in these cases there are no parameters for subjects.

Finally, it is unclear to me just exactly what the confidence regions mean. Do they mean that for other samples of subjects the recovered stimulus locations will be in the region 95% of the time? I am not sure. Furthermore, there are several complicated technical issues which affect the presentation of these

regions. By comparing the various papers discussed in this section, you will see that the developers of these techniques do not agree on the presentation method.

Fundamentally, what the user usually needs from MDS is not inferences but a picture. Most often the probabilistic (statistical) notion is irrelevant to the application, and what is needed is an exploratory (descriptive) graphical technique. In these cases MDS does just fine without the excess baggage introduced by inferential notions based on questionable and potentially misleading error models. True, in some cases the theoretical understanding of the phenomenon being studied is sufficiently advanced (perhaps through previous use of descriptive MDS) that the error processes are well understood. Here the maximum likelihood MDS may prove particularly useful as a confirmatory analysis following earlier exploratory MDS analyses. However, during the earlier stages of scientific development the significance tests of these new procedures should be used with care.

Multidimensional Scaling Algorithms

During the review period there was much effort focused on developing new multidimensional scaling algorithms. This effort has resulted in "third-generation" algorithms which are faster, easier to use, more general, and more accessible than previous algorithms. Because of this, the application of multidimensional scaling techniques has increased markedly.

ALSCAL is the program which is certainly the most widely distributed and is among the most general and easiest to use (Takane et al 1977, Young & Lewyckyj 1979, 1983, Verhelst 1981, Young 1982). It is incorporated into SAS, a major statistical system (SAS Institute 1982), further increasing its availability and usefulness. It performs metric or nonmetric analyses of two- or three-way square or rectangular data with unweighted, diagonal, or general Euclidean models. The algorithm is convergent, optimizing the fit of squared (weighted) Euclidean distances to transformations of the data. This is currently the program of choice because of its availability and generality. However, more efficient algorithms exist, and ALSCAL fits squared Euclidean distances, which is not as desirable as fitting the distances themselves.

SMACOF is another very general program. It is based on a very simple, fast and elegant algorithm (de Leeuw & Heiser 1980, Stoop & de Leeuw 1982). While not widely distributed, this algorithm is certainly the most efficient general algorithm currently available. It performs metric and nonmetric analyses of two- or three-way data, but is currently limited to the unweighted Euclidean model. It is a convergent least squares program, optimizing the fit of the Euclidean distances to transformations of the data. If this algorithm, which is still under development, is extended to a wider selection of models and is made available in a major statistical system, it will be the program of choice, at least among the least squares programs.

MULTISCALE is another very general and easy to use program (Ramsay 1977, 1978a,b, 1982a). It performs metric analyses of two- or three-way square data with unweighted, diagonal, or general Euclidean models. Arrangements are being made to incorporate this program into a major statistical system. There are two MULTISCALE algorithms, both of which are maximum likelihood algorithms that maximize the fit of log (weighted) Euclidean distances to tranformations of the data. Generally, maximum likelihood estimation takes more computer resources (time and memory) than least squares estimation, and that seems to be true for both MULTISCALE algorithms when compared to ALSCAL or SMACOF. MULTISCALE-I suffers from certain convergence problems and inefficiencies and has been replaced with MULTISCALE-II. This newer algorithm is certainly the most completely developed and tested general maximum likelihood algorithm. However, it is still a fairly new algorithm and should be used with some caution until it has received further study. Also, please note the recommendations given at the end of the previous section.

A number of other algorithms have been proposed, but the programs based on them are either not as easily available or not yet fully mature. The most important of these are the many algorithms proposed by Takane (see references above, plus Takane 1981, 1984); however, they still seem to be maturing. The PROSCAL program by MacKay & Zinnes (1982) is another promising maximum likelihood algorithm, but only very recently proposed. Saito (1978, 1982) is developing a new approach to constructing least squares procedures, as are Null & Sarle (1982). The SUMSCAL algorithm (de Leeuw & Pruzansky 1978), a very efficient way of optimizing the function used by the INDSCAL program, has not received the attention it deserves. The DISTREG procedure (Young et al 1983) is an interesting new development in constrained multidimensional scaling. The books by Schiffman, Reynolds & Young (1981) and by Coxon (1982) evaluate a number of the older algorithms and some of the newer ones.

There has been some work during the period to investigate the robustness of various multidimensional scaling procedures. The main work has been that of Sibson (1978, 1979), Sibson et al (1981), and MacCallum (1978, 1981). Weeks & Bentler (1979) have compared nonmetric and metric approaches. Mardia (1978) and Takane (1977) have investigated the mathematical characteristics of several algorithms. Best et al (1979) defined an index of the potential robustness of a stimulus space.

Multidimensional Unfolding

There continues to be some work attempting to resolve the algorithmic problems associated with multidimensional unfolding (MDU), still with only mixed results. The most advanced and encouraging work is that of Heiser, who

proposed and evaluated many improved unfolding algorithms appropriate to rectangular preference data (Heiser 1981, Heiser & Meulman 1983) and pairwise preferences (Heiser & de Leeuw 1981). He and his co-workers, investigate the relationship of unfolding to MDS, Principal Components, and Correspondence Analysis, and have proposed algorithms based on notions in those areas. Other work on unfolding appears in papers by Winsberg & Ramsay (1981), DeSarbo & Carroll (1980), Ramsay (1980a), Young (1982), and Rodgers & Young (1981).

Books and Reviews

The last article on scaling (Carroll & Arabie 1980) in the *Annual Review of Psychology* provided extensive coverage of the developments in multidimensional scaling up to 1978. Since that review appeared there have been many review articles. Certainly the one to reach the widest general scientific audience was the review of models and applications of multidimensional scaling and cluster analysis published in *Science* by Shepard (1980).

There have been a number of reviews tailored for various kinds of specific scientific audiences. Reviews of the models and algorithmic methods used in multidimensional scaling have been written for statisticians (Wish & Carroll 1981), computer scientists (Kruskal 1977), psychologists (Carroll 1980), ethologists (Spence 1978), and in numerous other fields. At least three encyclopedia entries have been written (Carroll & Kruskal 1978, Jones 1983, Young 1983a). Two extensive mongraphs have been written, one on unidimensional scaling (McIver & Carmines 1981) and the other on multidimensional scaling (Kruskal & Wish 1978).

The first introductory textbook on the theory, application, and methods of multidimensional scaling was written during the review period and has received uniformly very positive reviews (Schiffman et al 1981). Books have also been written on multidimensional scaling in German (Borg 1981b) and in Japanese (Takane 1980c). Collections of previously published papers have been edited by Lingoes et al (1979), Lingoes (1977), and by Davies & Coxon (1982). Collections of new papers written by invited contributors appear in Young & Hamer (1984) and in Law et al (1984). Conference proceedings have been edited by Golledge & Raynor (1982) and by Lantermann & Feger (1980). A system of unified scaling programs is documented in Coxon (1982), and an interesting treatise on multidimensional scaling was written by Borg (1981a).

SCALING MULTIVARIATE DATA VIA LINEAR AND BILINEAR MODELS

This section reviews recent developments in scaling methods which use either a linear or bilinear model of qualitative multivariate data. The distinction between the two types of models can be understood from the following equation:

$$Y=XA+E\ ,$$

where Y is an (n by r) matrix; X is (n by p); A is (p by r); and E is (n by r). This one equation can be either a linear or a bilinear equation, depending on the specific nature of its matrices.

1. Linear model:
   1. Y is observations from n sources (subjects) on r variables.
   2. X is observations from the same n sources on p different variables.
   3. A contains r coefficients of p linear equations that are the parameters of the linear model. These are estimated to optimize some fit criterion.
   4. E is a matrix of residual error from perfect fit.
2. Bilinear model:
   1. Y is still observations from n sources (subjects) on r variables.
   2. X is now no longer observed, but represents estimates of scores of the n sources on p latent (unobserved, hypothetical) variables.
   3. A is still r coefficients of p linear equations that are the parameters of the bilinear model. These are estimated to optimize fit, as before.
   4. E is still a matrix of residual error from perfect fit.

Note that the crucial difference in these models is that X is observed in the linear model but is unobserved in the bilinear model. Both models have observed multivariate data in Y, and both models posit a set of linear equations to model this multivariate data. Since both models use linearity as the basic idea, they both have the word "linear" in their name. The "bi" aspect of the bilinear model refers to the fact that it requires the estimation of *two* matrices (X and A).

There are four major streams of development in scaling with linear and bilinear models during the review period. The first two of these, conjoint analysis and multiple (or canonical) regression, use the linear model. The second two, principal components and correspondence analysis, use the bilinear model.

Conjoint Analysis

Conjoint analysis, in its simple form, is nothing more than main effects ANOVA performed on ordinal data. Stated in the usual ANOVA fashion, the simplest conjoint model is

$$y^*_{ik} = a_i + b_j,$$

where y^*_{ij} is the model's prediction of y_{ij}, the ordinal level observation in the experimental condition produced when level i of factor A is empirically combined with level j of factor B. The prediction represents the simple additive

combination of a_i, the effect of the ith level of factor A and b_j, the effect of the jth level of factor B.

Naturally, the simple additive model underlying conjoint analysis is a special case of the general linear model formula given in matrix form above. The matrix X is a design matrix specifying a main effects ANOVA design (one column for each experimental variable), and where Y is univariate (has only one column) and is at the ordinal level of measurement. (Some researchers refer to ANOVA on interval level data as conjoint analysis, but this only confuses the discussion because such an analysis is simply ANOVA. Since these developments do not fall into my definition of scaling I do not review them.)

During the review period, de Leeuw et al (1976) proposed ADDALS, an efficient least squares algorithm for conjoint analysis. Their approach is notable for its great flexibility with regards to the measurement characteristics of the data (Y). These data may be at the binary, nominal, ordinal, or interval levels of measurement (or may be mixtures of each); may be generated by a discrete or continuous process; may have known degrees of imprecision; may be balanced or unbalanced; and may have any arbitrary type of conditionality. (The last characterisitc permits the data to be partitioned into any number of arbitrary subsets, each with its own separately stated measurement characteristics.)

In addition to the flexibility in the observed data, the experimental variables (i.e. X) may have order or linear constraints placed on them, thus also permitting them to be measured at the binary, nominal, ordinal, or interval levels of measurement. The main limitation of ADDALS is in the model: it must be the univariate main effects model specified above. That is, there can only be one Y variable, and X can represent only main effects. Perreault & Young (1980) present an overview of this approach to conjoint analysis.

As you would expect, more complicated ANOVA models have been discussed in the context of conjoint analysis. The WADDALS algorithm (Takane et al 1980) extends ADDALS to the weighted additive model:

$$y^*_{ijk} = w_{ka}a_i + w_{kb}b_j,$$

where a_i and b_j are the same as before, and where the subscript k on y^*_{ijk} indicates that the prediction is for the kth individual. Note that the weights w_{ka} and w_{kb} are specific to indivdual k, and that there is one weight for each way of the experimental design. The general linear model matrix formula has a column in the observed data Y for each subject k. Thus, Y is now multivariate. The model posits that each subject (column in Y) has a weight for each way of the experimental design (column in X). This is the nonmetric main effects multivariate analysis of variance model.

The %CONJOINT algorithm, a SAS macro written in the MATRIX language by me and my students (Young 1983b), extends ADDALS in a different

direction. It permits generalized (optional interaction terms) univariate conjoint analysis. This is a full-fledged nonmetric ANOVA program. MANOVALS, a program developed by de Leeuw (personal communication), is a complete nonmetric MANOVA program that incorporates both the multivariate and generalized (interaction term) notions.

A very promising new direction has been taken by Takane (1982a) in his work on MAXADD, a maximum likelihood program for conjoint analysis. This work covers the univariate, main-effects, weighted or unweighted additive model. MAXADD can analyze nominal data or certain types of ordinal data. Takane (1984) compares MAXADD to ADDALS and WADDALS. Falmagne (1978) has developed a procedure similar to MAXADD, though it is more limited.

Takane (1982a) assumes that nominal and ordinal data represent incomplete information about interval data. In particular, a metric (interval) process is assumed to underlie the nonmetric (nominal or ordinal) data, with the metric information getting lost in the observation process, leaving only nonmetric information. Takane then proposes specific models linking the unobserved metric process with the observed nonmetric data. A different model is proposed for categorical data, paired comparison data, and directional rankings data. In each case the model is Thurstonian, but with the normal error model replaced with a (univariate or multivariate) logistic error model. In the case of paired comparison data the model is equivalent to the BTL model discussed by Bradley & Terry (1952) and by Luce (1959).

The main advantage of MAXADD is that it allows various statistical inferences not possible with the least squares procedures discussed above. The assumptions underlying the significance tests are essentially the same as in Takane's maximum likelihood MDS developments, and the comments made above about those developments apply here. In particular, the lack of empirical independence seems to be very well handled by his developments. However, small samples violate an assumption required by the asymptotic significance tests and may adversely affect the significance levels they generate.

Nonmetric Regression Analysis

Nonmetric regression refers to a very general class of situations in which multiple or canonical regression is applied to multivariate data having at least one ordinal variable. The work of Young et al (1976) is a simple extension of their conjoint analysis work. They propose CORALS, a least squares algorithm that extends canonical regression so that the variables may be measured at the binary, nominal, ordinal, or interval levels of measurement; may be discrete or continuous; and may have any type of conditionality. There is no restriction on the mixture of measurement characteristics. Of greatest importance is the fact that each variable has its own separately defined measurement characteristics.

Since the canonical regression model includes all other linear models as special cases, all of the least squares conjoint analysis developments discussed in the previous section are subsumed under CORALS.

The article by Young and co-workers on CORALS proposes an algorithm limited to obtaining the first pair of canonical variates. This limitation was eliminated by de Leeuw (personal communication) in later developments. My students and I (Young 1983b) have written %CORALS (a SAS macro in the MATRIX language) which also removes this limitation as well as providing nonmetric maximum redundancy analysis for the first time. Tenenhaus (1979) presents an interesting mathematical comparison of ADDALS, MORALS (the multiple regression version of CORALS), and Kruskal's MONANOVA (1981). Cunningham (1982) has discussed four monotone multiple regression models, two of which are special cases of MORALS and two of which are not cases of the general linear model.

Breiman & Friedman (1982) have independently developed a nonmetric multiple regression procedure named ACE that is also very flexible with regard to the measurement characteristics of the variables. Like MORALS, ACE extends multiple regression to ordinal variables. However, the two approaches differ in the way these variables are transformed. Whereas the MORALS approach transforms ordinal variables according to Kruskal's least square monotonic transformation, the ACE approach transforms ordinal variables according to Tukey's smoothing transformations. This implies that the overall behavior of the ACE algorithm is not quite as "nice" as MORALS (since smoothing is not least squares), but that ACE obtains "nicer" transformations of the ordinal variables (since least squares monotonic transformations are step functions which are not smooth).

CORALS extends the descriptive (but not the inferential) power of the canonical model to data that include ordered variables. This is an important development because for many decades canonical analysis has been possible on data with categorical (nominal) and continuous (interval) variables but not on data with ordinal variables. There is some work that also extends the inferential power of the canonical model to ordered variables. This work is by Winsberg & Ramsay (1980), who propose a maximum likelihood procedure for nonmetric multiple regression with nominal, ordinal, or interval level variables. The proposal uses B-splines to transform ordinal variables monotonically so that the likelihood function is maximized. This follows on previous work by Ramsay (1977) which permitted a limited class of monotonic transformations of the dependent variable.

As an apporach to defining monotonic transformations, B-splines are attractive alternatives to either least squares monotonic transformations or Tukey-type smoother transformations. B-splines provide flexibly shaped but smooth monotonic transformations, and they have relatively small number of easily

identified parameters. So from this viewpoint, B-splines are a valuable tool for constructing monotone transformations.

While this approach extends the inferential power of multiple regression to ordinal variables, I question the robustness of the inferential process to violations of the assumptions about independence and sample size. Furthermore, it is the monotonically transformed variables that must have a known (and tractable) probability density function, implying that the residuals of the transformed variables from the model are, for example, normally distributed. I would feel more comfortable with the statistical inferences if the framework taken by Takane in MAXADD were extended to the canonical situation, since Takane's inferential framework seems to be more in keeping with the basic nature of ordinal data.

Nonmetric Principal Components Analysis

Principal components analysis is a well-known data analysis technique that uses the bilinear model to analyze multivariate data whose variables are all quantitative (at the interval level of measurement). There are several developments during the review period concerning nonmetric PCA, also called nonlinear PCA. These developments apply the same bilinear model to multivariate data that have at least one variable which is qualitative (ordinal or nominal).

The paper by Young et al (1978) proposed PRINCIPALS, a nonmetric PCA algorithm (each variable can be binary, nominal, ordinal, or interval, and can be discrete or continous). The algorithm is in the tradition of metric PCA because it maximizes the variance accounted for by the first several components. It transforms the variables to maximize the variance accounted for by a stated number of components, with the added proviso that each variable's transformation must strictly satisfy the restrictions implied by the variable's stated measurement characteristics. Unlike metric PCA, the user must specify the desired number of components as well as the measurement characteristics of each variable.

Tenenhaus (1977) independently and simultaneously proposed PRINQUAL, an algorithm that is identical to PRINCIPALS with two exceptions. The two differences are (*a*) PRINQUAL has a superior initialization approach for nominal variables, and (*b*) it does not handle ordinal variables at all. A SAS macro named %PRINQUAL has been written by me and my students (Young 1983b) in the MATRIX language. %PRINQUAL adds the superior Tenehaus initialization to the PRINCIPALS algorithm. Tenenhaus (1982) has thoroughly studied the mathematical properties of the resulting algorithm, and he and I (Tenenhaus & Young 1983) have compared it to Correspondence Analysis. The PRINCIPALS notion has been extended to three-way, three-mode data (i.e. many matrices of multivariate data) by Sands & Young (1980) and Kroonenberg & de Leeuw (1980) in independent but closely related develop-

ments. An algorithm that fits the common-factor model to mixed measurement level data was proposed by Takane et al (1979).

While the %PRINQUAL approach has the desirable feature of maximizing the variance accounted for by an algorithm that is very stable and is convergent, it has the undesirable feature of obtaining transformations of qualitative variables which are discontinuous step functions. Two similar but separate developments have appeared that use splines to obtain transformations of the qualitative variables which are smooth but which do not strictly maximize the variance accounted for by the stated number of components.

de Leeuw and co-workers (1982) use the least squares framework and linear combinations of B-splines to transform the variables. Such transformations are not necessarily monotone, implying that the variables are nominal. They investigate the mathematical relationship of their approach to the %PRINQUAL approach and to Correspondence Analysis.

Ramsay & Winsberg (1984) employ maximum likelihood estimation and use non-negative combinations of monotone splines (M-splines) to obtain the transformations of the variables. Since the resulting transformations are monotonic, the implication is that all variables are thought of as being ordinal. They proceed to develop significance tests on the foundation of their maximum likelihood algorithm. While I have reservations about significance tests in this context, I do think that M-splines are a very nice way of obtaining smooth monotonic transformations. Perhaps it would be desirable to develop an algorithm which permits B-splines for nominal variables and M-splines for ordinal variables.

Correspondence Analysis

As pointed out by Nishisato (1980a, 1982), the basic idea underlying correspondence analysis has existed for at least 50 years. However, the technique has suffered from being independently invented and named by a large number of people during those years. Some common names for this type of analysis are: method of reciprocal averages, appropriate scoring, additive scoring, Hayashi's theory of quantification, principal components of qualitative data, Guttman scaling, optimal scaling, dual scaling, biplot, and nonlinear multivariate analysis.

As can be imagined, there are nearly as many ways to describe the analysis as there are inventors and names. Yet they all have in common the fact that the data are fundamentally categorical (nominal or binary), although such data can be presented easily as frequencies or as contingencies. The analysis can be described alternatively as submitting such data to (*a*) a singular-value decomposition; (*b*) a principal components analysis; (*c*) a simple additive (conjoint) analysis; or (*d*) a canonical analysis. To add to the confusion, the method has been studied extensively in several countries, including the USA, Canada,

Britain, Australia, South Africa, France, Japan, and the Netherlands. Naturally, papers have been written in several languages (English, French, and Japanese, predominately).

During the review period, work has continued on the topic, with emphasis on a synthesis of the many different developments. I do not presume to review all, or even most, of the literature, but simply to make reference to those whose work I know the best. Nishisato has published extensively on "Dual Scaling" both in English (1979a,b, 1980a,b) and in Japanese (1982). Some portions of this work appear to represent new directions, while other portions are histories or syntheses. Tenenhaus (1981) and Tenenhaus & Young (1983) present syntheses covering the French and English literature (in English). Gifi (1981) and Meulman (1982) are representative of the work in The Netherlands (in English). Heiser (1981) relates unfolding to correspondence analysis. All of these workers have developed computer programs that perform correspondence analysis of one type or another.

Books and Reviews

There are not very many books and reviews focusing on linear scaling methods for multivariate data. This is an accurate reflection of the state of the art: linear scaling methods have not yet matured to the level where such books and reviews would be appropriate. The main exception to this is the area of correspondence analysis, where most work seems to be synthesizing information from various sources. Here there is one introductory book and at least two advanced reviews.

Nishisato (1980a) has written a very readable, current, and accurate introductory book on correspondence analysis that is appropriate for courses taught at the graduate level. Gifi (1981) (a pseudonym for de Leeuw and his coworkers) presents a very comprehensive treatment of non-linear multivariate analysis, including major chapters on correspondence analysis, multidimensional scaling, and many innovative scaling methods based on the linear and bilinear (and multilinear) models. Tenenhaus (1982) presents a very advanced and general mathematical treatment of scaling based on the linear and bilinear models (in French). The last two books are excellent for the researcher interested in detailed information.

POSTSCRIPT

Comparing this review with the two previous ones on scaling in the *Annual Review of Psychology* (Cliff 1973, Carroll & Arabie 1980) reveals that the predominate trend during the last decade is an explosive growth in MDS methods that peaked somewhere around 1980 and has begun to subside. In fact, this trend is reflected in the titles of these three reviews ("Scaling," "Multi-

dimensional Scaling," and "Scaling," respectively). While the major portion of the present review has been on MDS developments, I see the developments as ones which consolidate earlier gains (improved algorithms, generalized models) or strengthen earlier foundations (constrained and maximum likelihood MDS).

I do not mean to belittle these developments. They are critical developmental stages of a maturing methodology, but they may also be the final significant methodological developments. Kruskal, who wrote his foreword to the Schiffman, Reynolds & Young book on MDS in 1980, stated that "at the age of 20-odd, multidimensional scaling should be in the full vigor of youth—and I am happy to report that it is just graduating from college and doing very well."

In the 4 or 5 years since that was written, MDS has become a vigorous 30-year-old. It has, as Kruskal hoped, begun to develop further the "self-critical tools of assesment and diagnosis," as can be seen in the development of maximum likelihood methods, certainly the single most important trend during the review period (no matter how harshly I criticized them). And I expect that MDS will continue a healthy adulthood for the next several decades in essentially its current form.

As an individual in the family of scaling methods, MDS appears to be reaching full maturity, while the family continues to expand and grow. Perhaps Conjoint Analysis is now the budding teenager, soon to reach maturity. And many new linear scaling models have been born recently, with their futures yet to unfold. I expect that we will hear much more from Conjoint Analysis in the near future, and from its younger sisters and brothers after that.

If I were to venture a guess as to the identity of the next member of the family, I would say that it will be methods for graphically displaying the results of scaling analyses rather than new scaling methods as such. I can see that taking full advantage of the current revolution in high-resolution color graphics technology will greatly enhance the appeal of scaling methods. Three-dimensional color graphics displays that can be interacted with in real time, that are based on quantitative analyses of qualitative data (scaling), and that are based on sound psychological and perceptual principles, should be very useful and attractive to a wide audience, Thus, I think they will become the next member of the scaling family.

Literature Cited[1]

Bentler, P. M., Weeks, D. G. 1978. Restricted multidimensional scaling models. *J. Math. Psychol.* 17:138–51

Best, A. M., Young, F. W., Hall, R. G. 1979. On the precision of a Euclidean structure. *Psychometrika* 44:395–408

Bloxom, B. 1968. Individual differences in multidimensional scaling. *Educ. Test. Serv. Res. Bull. 68–45*

Bloxom, B. 1978. Constrained multidimensional scaling in N spaces. *Psychometrika* 35:283–319

Borg, I. 1981a. *Multidimensional Data Representations: When and Why*. Ann Arbor: Mathesis

Borg, I. 1981b. Anwendungsorientierte Multidimensionale Skalierung. Berlin: Springer (In German)

Borg, I., Lingoes, J. C. 1980. A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika* 45:25–38

Bradley, R. A., Terry, M. E. 1952. The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* 39:324-45

Breiman, L., Friedman, J. H. 1982. Estimating optimal transformations for multiple regression and correlation. *Dep. Stat. Orion 010.* Stanford Univ.

Carroll, J. D. 1980. Models and methods for multidimensional analysis of preferential choice (or other dominance) data. See Lantermann & Ferger 1980

Carroll, J. D., Arabie, P. 1980. Multidimensional scaling. *Ann. Rev. Psychol.* 31:607–49

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

Carroll, J. D., Chang, J. J. 1972. *IDIOSCAL (individual differences in orientation scaling): A generalization of INDSCAL allowing idiosyncratic reference systems.* Presented at Psychom. Soc. Meet., Princeton, N.J.

Carroll, J. D., Kruskal, J. B. 1978. Multidimensional scaling. In *International Encyclopedia of Statistics,* ed. W. H. Kruskal, J. M. Tanur, pp. 892–907. New York: Free Press

Carroll, J. D., Pruzansky, S., Kruskal, J. B. 1980. Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* 45:3–24

Cliff, N. 1973. Scaling. *Ann. Rev. Psychol.* 24:473–506

Coxon, A. P. M. 1982. *The User's Guide to Multidimensional Scaling.* Exeter, NH: Heinemann

Cunningham, J. P. 1982. Multiple monotone regression. *Psychol. Bull.* 92:791–800

Davies, P. M., Coxon, A. P. M., eds. 1982. *Key Texts in Multidimensional Scaling.* Exeter, NH: Heineman

de Leeuw, J., Heiser, W. 1980. Multidimensional scaling with restrictions on the configuration. In *Multivariate Analysis,* ed. P. R. Krishnaiah, 5:501–22. Amsterdam: North-Holland

de Leeuw, J., Prunzansky, S. 1978. A new computational method to fit the weighted Euclidean model. *Psychometrika* 43:479–90

de Leeuw, J., van Rijckevorsel, J., van der Wouden, H. 1982. Non linear principal components analysis with B-splines. *Methods Oper. Res.* 43:379–94

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

DeSarbo, W. S., Carroll, J. D. 1980. Three-way metric unfolding. In *Marketing Measurement and Analysis,* ed. J. W. Keon, pp. 157–83. Providence, RI: TIMS Coll. Marketing

Dunn, T. R., Harshman, R. A. 1982. A multidimensional scaling model for the size-weight illusion. *Psychometrika* 47:25–44

Easterling, D. V. 1984. Ideological shifts in the US Senate between 1971 and 1978: Principal directions scaling. See Young & Hamer 1984

Falmagne, J.-C. 1978. Probabilistic choice behavior theory: Axioms as constraints in optimization. In *Cognitive Theory,* Vol. 3, ed. J. N. Castellan Jr., F. Restle. New York: Erlbaum

Forsyth, B. 1984. The subjective attributes of cups and bowls: A principal directions analysis. See Young & Hamer 1984

Gifi, A. 1981. *Non-Linear Multivariate Analysis.* Leiden, The Netherlands: Dep. Data Theory, Univ. Leiden

Golledge, R. G., Raynor, J. N., eds. 1982. *Proximity and Preference: Problems in the Multidimensional Analysis of Large Data Sets.* Minneapolis: Univ. Minn. Press

Harshman, R. A. 1972. Foundations of the PARAFAC procedure: Models and condi-

[1]A key-word indexed bibliography of about 250 papers published during the review period is available upon request. This bibliography includes all the papers reviewed here, plus many not reviewed. It also includes a section on other bibliographies.

tions for an "explanatory" multi-modal factor analysis. *UCLA Work. Pap. Phonetics 16*

Heiser, W. J. 1981. *Unfolding Analysis of Proximity Data*. Leiden, The Netherlands: Univ. Leiden Dep. Datatheorie Press

Heiser, W. J., de Leeuw, J. 1981. Multidimensional mapping of preference data. *Math. Sci. Hum.* 19:39–96

Heiser, W. J., Meulman, J. 1983. Analyzing rectangular tables by joint and constrained multidimensional scaling. *Ann. Appl. Econometrics*. In press

Holman, E. W. 1978. Completely nonmetric multidimensional scaling. *J. Math. Psychol.* 20:1–15

Horan, C. B. 1969. Multidimensional scaling combining observations when individuals have different perceptual spaces. *Psychometrika* 34:139–65

Jones, L. V. 1983. Psychological scaling. In *Encyclopedia of Statistical Sciences*, Vol. 5, ed. S. Kotz, N. L. Johnson. New York: Wiley

Jones, M. R., MacCallum, R. C. 1984. An application of principal directions scaling to auditory pattern perception. See Young and Hamer 1984

Kroonenberg, P. M., de Leeuw, J. 1980. Principal component analysis of three-way data by means of alternating least squares algorithms. *Psychometrika* 45:69–98

Kruskal, J. B. 1977. Multidimensional scaling and other methods for discovering structure. In *Statistical Methods for Digital Computers*, ed. K. Enslein, H. S. Ralston, H. S. Wilf, 3:296–339. New York: Wiley

Kruskal, J. B. 1981. Multilinear models for data analysis. *Behaviormetrika* 10:1–20

Kruskal, J. B., Wish, M. 1978. *Multidimensional Scaling*. Beverly Hills, Calif: Sage

Lantermann, E. D., Feger, H., eds. 1980. *Similarity and Choice: Papers in Honour of Clyde Coombs*. Bern, Switzerland: Huber

Law, H. G., Snyder, C. W., MacDonald, R. P., Hattie, J., eds. 1984. *Three-Mode Models for Data Analysis*. New York: Praeger. In press

Lee, S.-Y., Bentler, P. M. 1980. Functional relations in multidimensional scaling. *Br. J. Math. Stat. Psychol.* 33:142–50

Lingoes, J. C. 1977. *Geometric Representations of Relational Data*. Ann Arbor: Mathesis

Lingoes, J. C., Borg, I. 1978. A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika* 43:491–519

Lingoes, J. C., Borg, I. 1983. A quasi-statistical model for choosing between alternative configuations derived from ordinarily constrained data. *Br. J. Math. Stat. Psychol.* 36: In press

Lingoes, J. C., Roskam, E., Borg, I., eds. 1979. *Geometric Representations of Relational Data*. Ann Arbor: Mathesis. 2nd ed.

Luce, R. D. 1959. *Individual Choice Behavior*. New York: Wiley

MacCallum, R. C. 1978. Recovery of structure in incomplete data by ALSCAL. *Psychometrika* 44:69–74

MacCallum, R. C. 1981. Evaluating goodness of fit in nonmetric multidimensional scaling by ALSCAL. *Appl. Psychol. Meas.* 5:377–82

MacKay, D. B., Zinnes, J. L. 1982. PROSCAL: A program for probabilistic scaling. *Discussion Paper 218*, Grad. Sch. Bus., Indiana Univ., Bloomington

Mardia, K. V. 1978. Some properties of classical multi-dimensional scaling. *Commun. Stat.* 7:1233–41

McIver, J. P., Carmines, E. G. 1981. *Unidimensional Scaling*. Beverly Hills, Calif: Sage

Meulman, J. 1982. *Homogeneity Analysis of Incomplete Data*. Leiden, The Netherlands: Univ. Leiden DSWO Press

Nishisato, S. 1979a. Dual scaling and its history. *Math. Sci.* 190:76–83

Nishisato, S. 1979b. Dual scaling and its variants. In *Analysis of Test Data (New Directions for Testing and Measurement)*, ed. R. E. Traub, pp. 1–12. San Francisco: Jossey-Bass

Nishisato, S. 1980a. *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto: Univ. Toronto Press. 276 pp.

Nishisato, S. 1980b. Dual scaling of successive categories data. *Jpn. Psychol. Res.* 22:134–43

Nishisato, S. 1982. *Quantifying Qualtitive Data: Dual Scaling and its Applications*. Tokyo: Asakura Shoten. 241 pp. (In Japanese)

Noma, E., Johnson, J. 1977. Constraining nonmetric multidimensional scaling configurations. *Tech. Rep. 60*. Ann Arbor: Hum. Perform. Cent., Univ. Mich.

Null, C. H., Sarle, W. S. 1982. *Robust multidimensional scaling*. Presented at Spring Psychom. Soc. Meet., Montreal

Perreault, W. D. Jr., Young, F. W. 1980. Alternating least squares optimal scaling: Analysis of nonmetric data in marketing research. *J. Mark. Res.* 17:1–13

Ramsay, J. O. 1977. Maximum likelihood estimation in multidimensional scaling. *Psychometrika* 42:241–66

Ramsay, J. O. 1978a. Confidence regions for multidimensional scaling analysis. *Psychometrika* 43:145–60

Ramsay, J. O. 1978b. Multiscale: Four programs for multidimensional scaling by the method of maximum likelihood. Chicago: Natl. Educ. Resources

Ramsay, J. O. 1980a. The joint analysis of

direct ratings, pairwise preferences and dissimilarities. *Psychometrika* 45:149–65

Ramsay, J. O. 1980b. Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika* 45:139–44

Ramsay, J. O. 1982a. *Multiscale II Manual.* Montreal: Dep. Psychol., McGill Univ.

Ramsay, J. O. 1982b. Some statistical approaches to multidimensional scaling data. *J. R. Stat. Soc. Ser. B* 145:285–312

Ramsay, J. O., Winsberg, S. 1984. Monotone spline transformations for dimension reduction. *Psychometrika* 49: In press

Rodgers, J. L., Young F. W. 1981. Successive unfolding of family preferences. *Appl. Psychol. Meas.* 5:51–62

Saito, T. 1978. The problem of the additive constant and eigenvalues in metric multidimensional scaling. *Psychometrika* 43: 193–201

Saito, T. 1982. Contributions to e_{ij}-type quantification and development of a new method of multidimensional scaling. *Behaviormetrika* 12:63–83

Sands, R., Young, F. W., 1980. Component models for three-way data: An alternating least squares algorithm with optimal scaling features. *Psychometrika* 45:39–67

SAS Institute. 1982. *SAS User's Guide: Basics, 1982 Edition.* Cary, NC: SAS Inst.

Schiffman, S. S., Reynolds, M. L., Young, F. W. 1981. *Introduction to Multidimensional Scaling: Theory, Methods and Applications.* New York: Academic

Shepard, R. N. 1980. Multidimensional scaling, tree-fitting, and clustering. *Science* 210:390–98

Sibson, R. 1978. Studies in the robustness of multidimensional scaling: Procrustes statistics. *J. R. Stat. Soc. Ser. B* 40:234–38

Sibson, R. 1979. Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *J. R. Stat. Soc. Ser. B* 41:217–29

Sibson, R., Bowyer, A., Osmond, C. 1981. Studies in the robustness of multidimensional scaling: Euclidean models and simulation studies. *J. Stat. Comput. Simul.* 13:273–96

Spence, I. 1978. Multidimensional scaling. In *Quantitative Ethology,* ed. P. W. Colgan, New York: Wiley

Stoop, I., de Leeuw, J. 1982. *How to Use SMACOF-IB.* Dep. Datatheorie, Univ. Leiden, The Netherlands

Takane, Y. 1977. On the relations among four methods of multidimensional scaling. *Behaviormetrika* 4:29–43

Takane, Y. 1978a. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—theory. *Jpn. Psychol. Res.* 20:7–17

Takane, Y. 1978b. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—evaluations. *Jpn. Psychol. Res.* 20:105–14

Takane, Y. 1980a. Analysis of categorizing behavior by a quantification method. *Behaviormetrika* 8:75–86

Takane, Y. 1980b. Maximum likelihood estimation in the generalized case of Thurstone's model of comparative judgment. *Jpn. Psychol. Res.* 22:188–96

Takane, Y. 1980c. *Tajigen Shakudoho* (Multidimensional Scaling). Tokyo: Univ. Tokyo Press (In Japanese)

Takane, Y. 1981. Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika* 46:9–28

Takane, Y. 1982a. Maximum likelihood additivity analysis. *Psychometrika* 17:225–41

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

Takane, Y. 1984. The weighted additive model. See Law et al 1984

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

Takane, Y., Sergent, J. 1983. Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika.* In press

Takane, Y., Young, F. W., de Leeuw, J. 1977. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 42:7–67

Takane, Y., Young, F. W., de Leeuw, J. 1979. Nonmetric common factor analysis: An alternating least squares method with optimal scaling feature. *Behaviormetrika* 6:45–56

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

Tenenhaus, M. 1977. Analyse en composantes principales d'un ensemble de variables nominales ou numeriques. *Rev. Stat. Appl.* 25:39–56 (In French)

Tenenhaus, M. 1979. La regression qualitative. *Rev. Stat. Appl.* 27:5–21 (In French)

Tenenhaus, M. 1981. Multiple correspondence analysis and duality schema: A synthesis of different approaches. *Cah. Rech. CESA No. 693,* Jouy-en-Josas, France

Tenenhaus, M. 1982. Analyse canonique generalisee de p cones polyedriques convexes: Application a l'analyse en composantes principales qualitative. *Cah. Rech. CESA,* Jouy-en-Josas, France (In French)

Tenenhaus, M., Young, F. W. 1983. Multiple correspondence analysis and the principal components of qualitative data. *Psychom. Lab. Rep. 169*. Univ. NC, Chapel Hill

Verhelst, N. D. 1981. A note on ALSCAL: The estimation of the additive constant. *Psychometrika* 46:465–68

Weeks, D. G., Bentler, P. M. 1979. A comparison of linear and monotone multidimensional scaling models. *Psychol. Bull.* 86:349–54

Weeks, D. G., Bentler, P. M. 1982. Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika* 47:201–8

Winsberg, S., Ramsay, J. O. 1980. Monotonic transformations to additivity using splines. *Biometrika* 67:669–74

Winsberg, S., Ramsay, J. O. 1981. Analysis of pairwise preference data using integrated B-splines. *Psychometrika* 46:171–86

Wish, M., Carroll, J. D. 1981. Multidimensional scaling and its applications. In *Handbook of Statistics, Volume 2: Classification, Pattern Recognition, and Reduction of Dimension,* ed. P. R. Krishnaiah. Amsterdam: North-Holland

Young, F. W. 1981. Quantitative analysis of qualitative data. *Psychometrika* 46:357–87

Young, F. W. 1982. Enhancements in ALSCAL-82. SUGI Proc.

Young, F. W. 1983a. Multidimensional scaling. In *Encyclopedia of Statistical Sciences,* Vol. 5. New York: Wiley

Young, F. W. 1983b. SAS %MACRO's for conjoint analysis, principal components of qualitative data, and canonical analysis of qualitative data. *SAS Tech. Rep.* Cary, NC: SAS Inst.

Young, F. W. 1984. The general Euclidean model. See Law et al 1984

Young, F. W., de Leeuw, J., Takane, Y. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika* 41:505–29

Young, F. W., de Leeuw, J., Takane, Y. 1980. Quantifying qualitative data. See Lantermann & Ferger 1980

Young, F. W., Hamer, R. M. 1984. *Multidimensional Scaling: Theory and Applications*. Hillsdale, NJ: Erlbaum

Young, F. W., Lewyckyj, R. 1979. *Principal directions scaling: A new individual differences model.* Presented at Ann. Psychom. Soc. Meet., Monterey, Calif.

Young, F. W., Lewyckyj, R. 1983. ALSCAL: A procedure for multidimensional scaling. In *SAS Supplemental Users Guide*. Cary, NC: SAS Inst

Young, F. W., Null, C. H., Hamer, R. M. 1983. DISTREG: A procedure for distance regression. In *SAS Supplemental Users Guide, 1982 Edition*. Cary, NC: SAS Inst.

Young, F. W. Takane, Y., de Leeuw, J. 1978. The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika* 43:279–81

Zinnes, J. L., MacKay, D. B. 1983. Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika* 48:27–48