MULTIDIMENSIONAL SCALING

Forrest W. Young, University of North Carolina

ABSTRACT

In this entry we summarize the major types of multidimensional scaling (MDS), the distance models used by MDS, the similarity data analyzed by MDS, and the computer programs that implement MDS. We also present three brief examples. We do not discuss experimental design, interpretation, or the mathematics of the algorithms. The entry should be helpful to those who are curious about what MDS is and to those who wish to know more about the types of data and models relevant to MDS. It should help the researcher. the statistical consultant, or the data analyst who needs to decide if MDS is appropriate for a particular set of data and what computer program should be used.

For a more complete, but still brief. introduction to MDS, the, reader should turn to Kruskal and Wish [6]. A complete discussion of the topics covered here as well as of experimental design. data analysis, and interpretive procedures can be found in Schiffman et al. [141. An intermediate-level mathematical treatment of some MDS algorithms is given in Davison (1983). An advanced treatment of the theory of MDS, illustrated with innovative applications, is presented by Young and Hamer [211]. Reviews of the current state of the art are presented by Young (1984a; 1984b). Multidimensional scaling is related to principal component analysis, factor analysis, cluster analysis, and numerical taxonomy: the reader is referred to the appropriate entries in this encyclopedia, along with the **SCALING** and **PROXIMITY DATA** entries.

1. OVERVIEW OF MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS) is a set of data analysis techniques that display the structure of distance-like data as a geometrical picture. It is an extension of the procedure discussed in SCALING.

MDS has its origins in psychometrics. where it was proposed to help understand people's judgments of the similarity of members of a set of objects. Torgerson [18] proposed the first MDS method and coined the term. his work evolving from that of Richardson [11]. MDS has now become a general data analysis technique used in a wide variety of fields [14]. For example, the book on theory and applications of MDS by Young and Hamer, [21] presents applications of MDS in

such diverse fields as marketing'. sociology, physics, political science', and biology. However, we limit our examples here to the field with which the author is most familiar, psychology.

MDS pictures the structure of a set of objects from data that approximate the distances between pairs of the objects. The data, which are called similarities. dissimilarities, distances, or proximities, must reflect the amount of (dissimilarity between pairs of the. In this article we use the term *similarity* generically to refer to both similarities (where large numbers refer to great similarity) and to dissimilarities (where large numbers refer to great dissimilarity).

In addition to the traditional human similarity judgment, the data can be an "objective" similarity measure (the driving time between pairs of cities) or an index calculated from multivariate data (the proportion of agreement in the votes cast by pairs of senators). However, the data must always I represent the degree of similarity of pairs of objects (or events).

Each object or event is represented by a point in a multidimensional space. The points are arranged in this space so that the distances between pairs of points have the strongest possible relation to the similarities among the pairs of objects. That is, two similar objects are represented by two points that are close together, and two dissimilar objects are represented by two points that are far apart. The space is usually a two- or three-dimensional Euclidean space, but may be non-Euclidean and may have more dimensions.

MDS is a generic term that includes many different specific types. These types can be classified according to whether the similarities data are qualitative (called nonmetric MDS) or quantitative (metric MDS). The number of similarity matrices and the nature of the MDS model can also classify MDS types. This classification yields classical MDS (one matrix, unweighted model), replicated MDS (several matrices, unweighted model), and weighted MDS (several matrices, weighted model). We discuss the nonmetric-metric and the classical-replicated-weighted classifications in the following sub-sections.

CLASSICAL MDS

The identifying aspect of *classical MDS* (CMDS) is that there is only one similarity matrix. Table I is a matrix of similarity data suitable for CMDS. It contains the flying mileages between 10 American cities. The cities are the "objects." and the mileages are the "similarities." An MDS of these data gives the picture in Fig. 1, a map of the relative locations of these 10 cities in the United States. This map has 10 points, one for each of the 10 cities. Cities that are similar (have short flying mileages) are represented by points that are close together, and cities that are dissimilar (have large mileages) by points far apart.

Table 1   Flying Mileages Between 10 American Cities

| Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington, DC
0	587	1212	701	1936	604	748	2139	2182	543
587	0	920	940	1745	1188	713	1858	1737	597
1212	920	0	879	831	1726	1631	949	1021	1494
701	940	879	0	1374	968	1420	1645	1891	1220
1936	1745	831	1374	0	2339	2451	347	959	2300
604	1188	1726	968	2339	0	1092	2594	2734	923
748	713	1631	1420	2451	1092	0	2571	2408	205
2139	1858	949	1645	347	2594	2571	0	678	2442
2182	1737	1021	1891	959	2734	2408	678	0	2329
543	597	1494	1220	2300	923	205	2442	2329	0
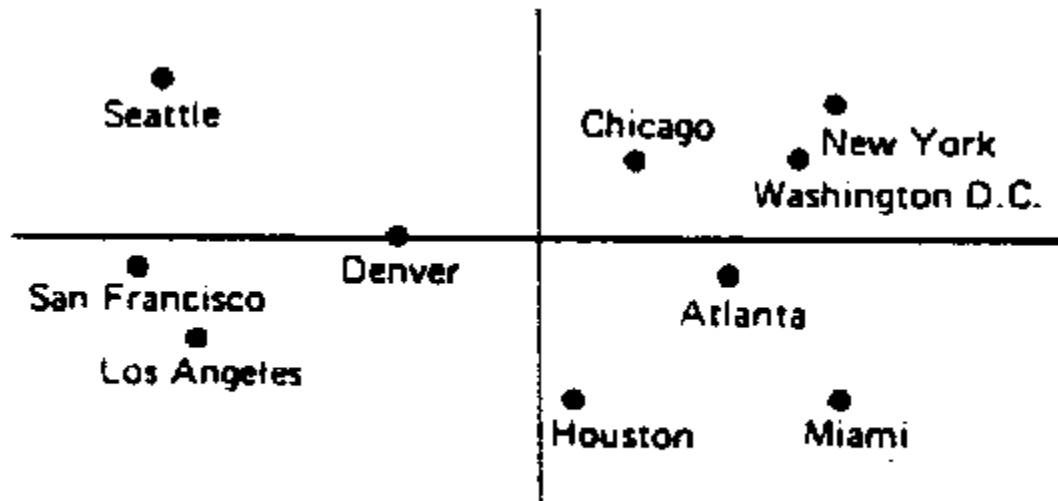
Figure 1   CMDS of flying mileages between 10 American cities.

Generally, CMDS employs Euclidean distance to model dissimilarity. That is, the distance d_{ij} between points i and j is defined as

$$d_{ij} = \sqrt{\sum (x_{ia} - x_{ja})^2}$$

Where x_i. Specifies the position (coordinate) of point i on dimension a. The distance can also be defined according to the Minkowski model:

$$d_{ij} = \left[\sum_a |x_{ia} - x_{ja}|^p \right]^{1/p}$$

Where the value of p (≤ 1) is set by the investigator.

For either definition of distance there are n points, one for each of the n objects. There are also r dimensions, where the value of r is determined by the investigator. The coordinates x_{ia} are contained in the nXr matrix **X**. Using matrix algebra, the Euclidean model can be defined as

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

where x_i is the ith row of X and contains the r coordinates of the ith point on all r dimensions. A simple matrix expression for the Minkowski model is not possible. For both models, the distances d_{ij} are contained in the n X n symmetric matrix D. Finally. The similarity s_{ij} are contained in the matrix S, also n X n.

METRIC CMDS. The first major CMDS proposal [18] was metric (i.e.. the similarities had to be quantitative). Torgerson's development required the data to be at the ratio level of measurement, although this was soon generalized to the interval level [8]. While the data could contain random error, this early type of MDS required that the data be dissimilarities (not similarities), complete (no

missing values), and symmetric (the dissimilarity of objects *I* and *J* had to equal that of objects *J* and *1*). These CMDS proposals also required the distance model to be Euclidean. The flying mileage example is metric CMDS because flying mileages are at the ratio level of measurement.

For metric CMDS the distances D are determined so that they are as much like the dissimilarities S as possible. There are a variety of ways in which "like" is strictly defined, but a common one is a least-squares definition. In this case we define

1 {S} = D + E.

Where *I* {S} is read "a linear transformation of the similarities." If the measurement level is ratio, and then the linear transformation has a zero intercept, but can be nonzero when the level is interval. If the data are similarities. The slope of the transformation is negative: if dissimilarities, it is positive.

In the preceding equation, E is a matrix of errors (residuals) that in the least-squares optimization situation, we wish to minimize. Since the distances D are a function of the coordinates X, the goal of CMDS is to calculate the coordinates X so that the sum of squares of E is minimized, subject to suitable normalization of X. We also need to calculate the best linear transformation *I*{S}. Torgerson's method does not actually minimize the sum of squares of E, nor do ALSCAL or

MULTISCALE. The KYST, MINTSSA, and SMACOF programs do. These programs are all discussed in the Computer Programs section.

NONMETRIC CMDS. The second major CMDS proposal [5, 15] was nonmetric. That is. The data could be at the ordinal level of measurement (see ORDINAL DATA). In addition. The data S could be either complete or incomplete. Symmetric or asymmetric, and similarities or dissimilarities.

These nonmetric CMDS proposals extended the distance model to the Minkowski case and generalized the relation between similarities and distances. They enable defining

$m\{S\} = D + E,$

Where m {S} is read "a monotonic transformation of the similarities." If S is actually dissimilarities then m {S} preserves order, whereas if S is similarities, it reverses order. Thus, for nonmetric CMDS, we need to solve for the monotonic (order-preserving) transformation m{S} and the coordinates X. which together minimize the sum of squares of the errors E (after normalization of X). This exact problem is solved by the MINISSA, KYST, and SMACOF programs (discussed in the final section) while ALSCAL and MULTISCALE solve closely related problems.

The nonmetric optimization represents a much more difficult problem to solve than the metric problem and is an important breakthrough in multidimensional scaling. In fact, nonmetric CMDS is the first example of using quantitative models

to describe qualitative data that belongs to the approach discussed by Young [19] *(see* QUALITATIVE DATA. STATISTICAL ANALYSIS).

It is reassuring to know that when we degrade the flying mileages (Table 1) into ranks of flying mileages, and then submit the ranks to nonmetric CMDS, the map that results is indistinguishable from that shown in Fig. 1.

REPLICATED MDS

The next major development, *replicated MDS* (RMDS), permitted the analysis of several matrices of similarity data simultaneously [7]. There are m matrices S, one for each subject $k, k = 1, \ldots, m.$

RMDS uses the same distance models as CMDS, but uses them to describe several similarity matrices rather than one. With RMDS, the matrix of distances D is determined so that it is simultaneously like all the similarity matrices S_k.

For metric RMDS, the least-squares definition of "like" is

$$1_k \{S_k\} = D + E_k,$$

Where $1_k \{S_k\}$ is the linear transformation of the kth similarity matrix S_k which best fits the distances D. The data may be similarities or dissimilarities and may be at the ratio or interval levels, just as in metric CMDS. The analysis minimizes the sum of the squared elements in ail error matrices E_k. subject to normalization of X.

For nonmetric RMDS we minimize the several E_k in

$$m_k \{S_k\} = D + E_k ,$$

Where $m_k \{S_k\}$ is the monotonic transformation of the similarity matrix S_k that is a least-squares fit to the distances in matrix D. The data may be similarities of dissimilarities. Just as in CMDS.

Note that for RMDS each linear or monotonic transformation l_k. or m_k is subscripted, letting each data matrix S_k have a unique linear or monotonic relation to the distances D. Since k ranges up to m. there are m separate linear or monotonic transformations, one for each of the m dissimilarity matrices S_k. This implies that RMDS treats all the matrices of data as being related to each other (through D) by a systematic linear or monotonic transformation (except for a random error component). The KYST and SMACOF programs minimize the sum of squares of E_k, while ALSCAL and MULTISCALE solve other closely related problems. In psychological terms, RMDS accounts for differences in the ways subjects use the response scale (i.e., differences in response bias).

Jacobowitz [4] used RMDS to study the way language develops as children grow to adulthood. In his experiment he asked children and adults to judge the similarity of all pairs of 15 parts of the human body. The judges were five-, seven-. And nine-year-olds, and adults. There were 15 judges at each age. Four separate RMDS analyses were done. One for each age group.
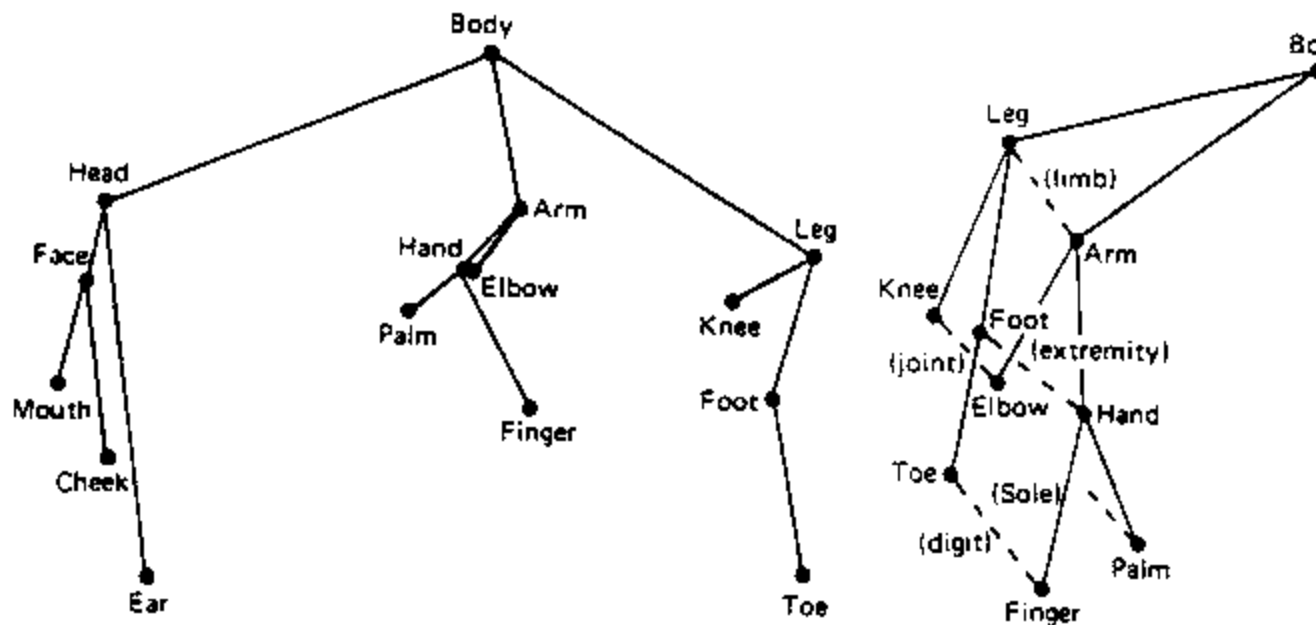
Figure 2 (a) RMDS of children's similarity judgments about 15 body parts; (b) RMDS [...] similarity judgments about 15 body parts.

The RMDS results for the five-year-olds are shown in Fig. 2*a*, and for the adults in Fig. 2*b*. The analysis located the points in the space, but did not draw the lines. The lines were drawn by Jacobowitz to interpret the psycholinguistic structure that people have for body-part words. Jacobowitz theorized that the structure would be hierarchical. We can see that it is. He further theorized that the structure would become more complex as the children become adults. This theory is also supported, since the adults' hierarchy also involves a classification of corresponding arm and leg terms. (In Fig. 2*b* the corresponding terms are linked by dashed lines. the implied classification terms are shown in parentheses, and the word *sole,* which was not a stimulus, is shown in the position that we would predict it to be in if the study were repeated.)

WEIGHTED MDS

The next major MDS development, *weighted MDS* (WMDS), generalized the distance model so that several similarity matrices $\mathbf{S}_k$ could be assumed to differ from each other in systematically nonlinear or nonmonotonic ways. Whereas RMDS only accounts for individual differences in response bias, WMDS incorporates a model to account for individual differences in the fundamental perceptual or cognitive processes that generate the responses. For this reason, WMDS is often called *individual differences scaling (INDSCAL)* and is often regarded as the second major breakthrough in multidimensional scaling.

WMDS invokes the following definition of weighted Euclidean distance:

$$d_{ij} = \left[\sum_a w_{ka}(x_{ia} - x_{ja})^2 \right]^{1/2}$$

Which, in matrix algebra is

$$d_{ij} = [(x_i - x_j) W_k (x_i - x_j)']^{1/2}$$

Where $\mathbf{W_k}$ is a *r x r* diagonal matrix. The diagonal values, which must not be negative, are weights for subject k on each of the *r* dimensions.

WMDS is appropriate for the same type of data as RMDS. However, RMDS generates a single distance matrix D, while WMDS generates m unique distance matrices D_k, one for each data matrix S_k. The distances D_k are calculated so that

they are all as much like their corresponding data matrices S_k as possible. For metric WMDS, the least-squares problem is

$$l, \{S_k\} = D_k + D_k ,$$

And for nonmetric WMDS, the problem is

$$m_k \{S_k\} = D_k + E_k..$$

Thus, for WMDS, we need to solve for the matrix of coordinates X, the m diagonal matrices of weights Wk, and the m transformations Mk or 1. We wish to do this so that the sum of squared elements in all error matrices, E, is minimal subject to normalization constraints on X and Wk.

Neither of the two most commonly used computer programs solves either of the problems defined. (These programs and others are discussed in the last section.) The INDSCAL program, by Carroll and Chang [1], provided the first metric WMDS solution. However, it optimizes the fit of scalar products to a transformation of the data. The ALSCAL program, by Takane et al. [17] (see also Young and Lewyckyj [22] and Young et al. [24], provided the first and still the only algorithm to incorporate both nonmetric and metric solution to WMDS and optimize the fit of squared distances to the data. In fact, ALSCAL is still the only algorithm to provide the user with nonmetric and metric solutions to the CMDS, RMDS, and WMDS situations discussed, and it is regarded as the third major breakthrough in multidimensional scaling.

The MULTISCALE algorithm by Ramsay (101 provided the first metric WMDS solution to optimize the preceding index (it fits distances to the data). Finally, the SMACOF algorithm [2, 3] and its associated program [16], which is still under development, will more than likely be the first program to be able to fit distances to the data so that the sum of squares of E is strictly minimized, where the distances may be CMDS, RMDS, or WMDS distances, and where the transformation may be metric or nonmetric.

While WMDS incorporates the RMDS notion of individual differences in response bias (via m_k and l_k), the important aspect of WMDS is that it provides specific parameters for individual variation in cognitive or perceptual processes. These parameters are the weights. The weights are interpreted as the importance, relevance. or salience of each dimension to each individual. A large weight means that the dimension is important to the individual, a small weight means the dimension is unimportant. If the similarity matrices correspond to experimental conditions, say, rather than to people, the interpretation is that the weights reflect the importance of each dimension in the various experimental conditions.

The Jacobowitz data already discussed provide a nice example of WMDS. An analysis of the 15 five-year-olds together with the 15 adults provided the results displayed in Fig. 3. In Fig. 3a, we see that the stimulus structure is the anticipated hierarchy. In Fig. 3b, which is the weight space. we see that the children and adults occupy different parts of the space, showing that the children and adults have different cognitive structures for parts of the body.
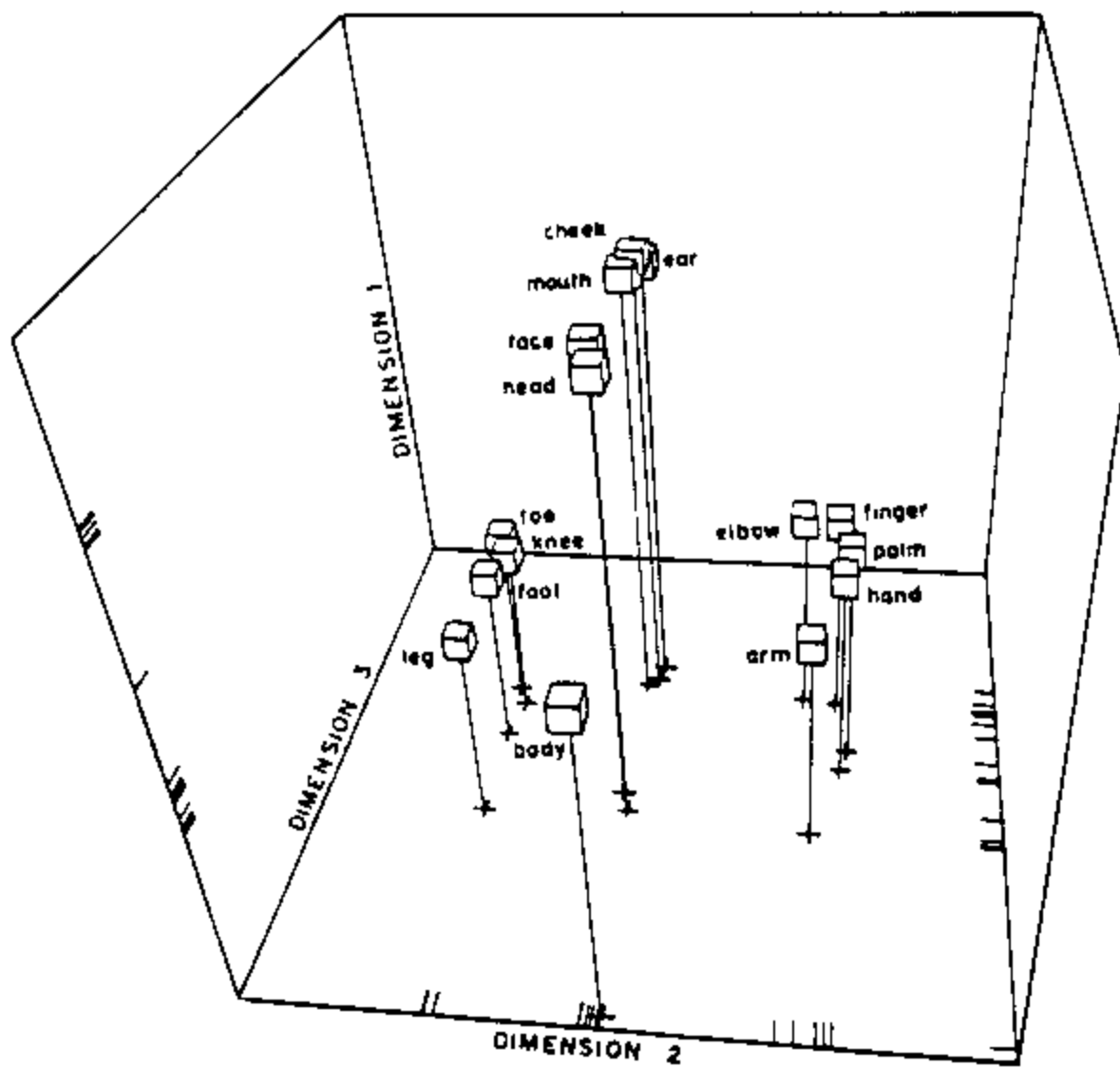
Figure 3 (a) WMDS of children's and adults' similarity judgments about 15 body parts.
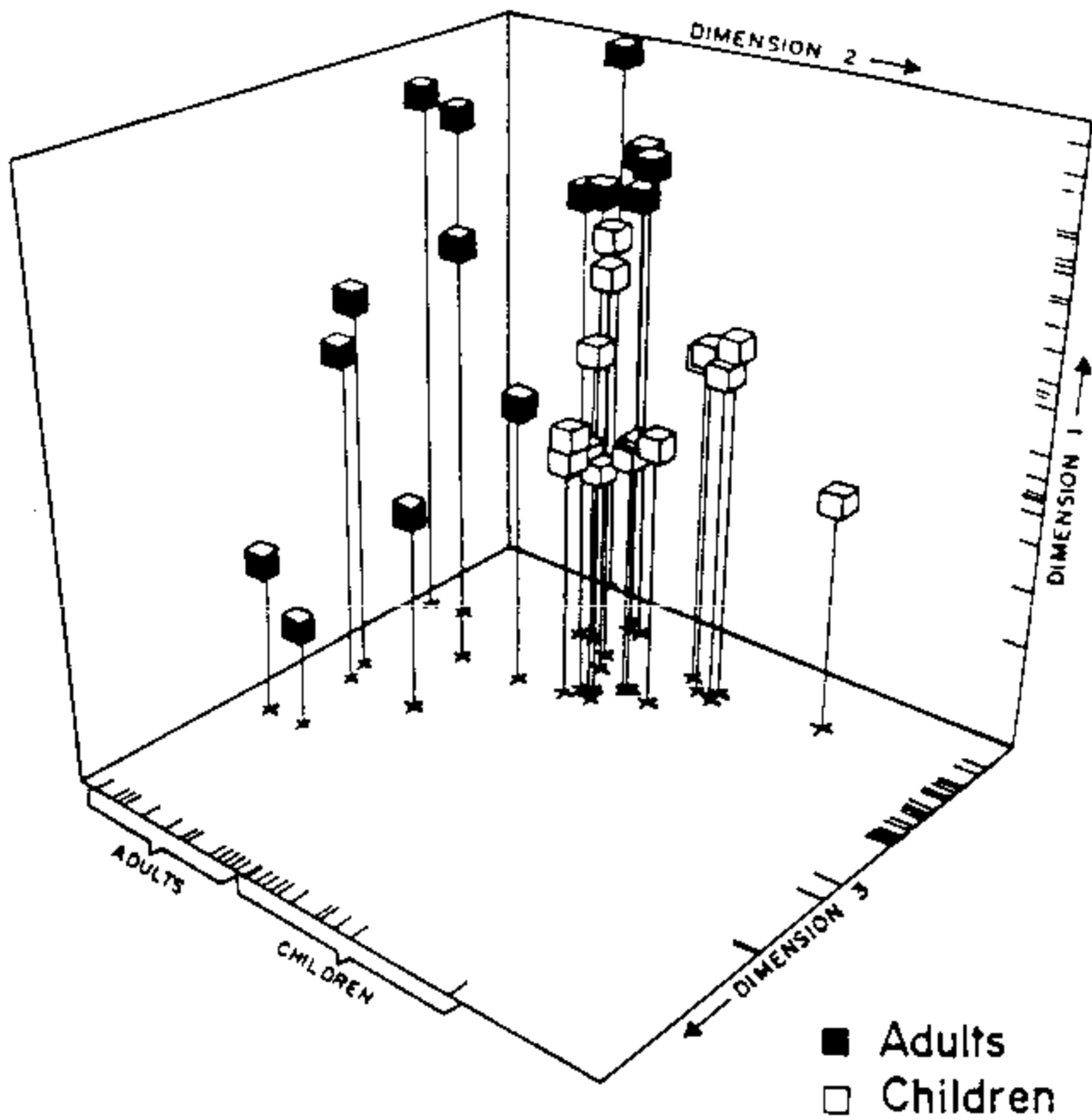
Figure 3 (b) Subject weights from WMDS analysis.

(b)

2. COMPUTER PROGRAMS

Several computer programs have become a significant part of the MDS discipline. These programs and several of their characteristics are listed in Table 2. A complete reference for each program is given in the bibliography.

Table 2   Characteristics of Several MDS Computer Programs

	MINISSA	KYST-2a	INDSCAL	ALSCAL-83	MULTISCL-2	S
Similarity	Yes	Yes	Yes	Yes	Yes	
Asymmetric	No	Yes	Yes	Yes	Yes	
Missing	Yes	Yes	No	Yes	Yes	
Two-way	Yes	Yes	No	Yes	Yes	
Measurement	N	MN	M	MN	M	
Model	C	CR	WO	CRWO	CRWO	
Fit	D	D	P	S	DL	
Algorithm	L	L	L	L	M	
Converge	No	No	Yes	Yes	No	
Stimuli	100	100	dyn	dyn	50	
Matrices	dyn	dyn	dyn	dyn	100	
Elements	4950	4000	dyn	dyn	15000	
Dimensions	10	6	10	6	10	

The first four rows of Table 2 refer to the type of data each program can analyze. The information includes whether each program can analyze similarity data in addition to dissimilarity data; asymmetric data in addition to symmetric data; data with missing elements in addition to data without; and two-way in addition to three-way data.

The next two rows of Table 2 refer to the types of analyses each program can provide. The Measurement row refers to whether the program can provide only nonmetric analysis (N), only metric analyses (M), or both (MN). The Model row

refers to whether the program can provide analyses that are classical (C), replicated (R), weighted (W), or other types (0).

The next three rows of Table 2 refer to several aspects of the iterative algorithm employed by each program. The Fit row refers to the aspect of the model that is fit to (a transformation of) the data (D indicates distances. P scalar products, S squared distances, and L log distances). The Algorithm row indicates whether the program is a leastsquares program (L) or a maximum likelihood program (M). The Converge row shows whether the algorithm is convergent (each iteration must improve the fit index being optimized) or not.

The final four rows specify the maximum size problem that can be analyzed by each program. Some programs place specific limits on the number of stimuli. matrices, dimensions, or total number of data elements. These limits are indicated by a number. Other programs are dynamic and place no limit. These are indicated as "dyn."

The ALSCAL-84, MULTISCALE-2, and SMACOF-IB programs are the current state-of-the-art programs. Of the programs listed in Table 2, ALSCAL-84 [23] is the most flexible, fitting the widest range of models to the widest range of data. The ALSCAL algorithm is convergent (which is desirable) and is faster than MULTISCALE but slower than SMACOF. ALSCAL is the only MDS program currently available in major statistical systems (SAS and SPSS) and is the easiest program to use. However, the algorithm optimizes the fit of squared

distances to the dissimilarities. which is not the most desirable optimization criterion. ALSCAL is descriptive, having no inferential aspects.

MULTISCALE-2 [10] has the unique feature that it is based on the maximum likelihood principle. Of the programs listed in Table 2. it is the only one that enables statistically based significance tests and that can be used for inferential purposes. MULTISCALE provides the user with a selection of models that is smaller than that provided by ALSCAL, but larger than that provided by SMACOF. However, of the three programs, MULTISCALE is the least flexible in the types of data that can be analyzed, and the slowest. Also it has a nonconvergent algorithm.

SMACOF-IB [16], clearly the fastest of these three programs. optimizes the fit of distances to dissimilarities by a convergent algorithm. The algorithm [2, 3] is the simplest and most elegant of any program listed in Table 2. It fits the CMDS. RMDS, and WMDS models and is as flexible as ALSCAL in the types of data it can analyze. However, SMACOF is currently under active development: it is difficult to use and is not available in any statistical package. When fully mature, SMACOF will be the program of choice.

REFERENCES

[1] Carroll, J. D. and Chang. J. J. (1970). *Psychometrika.* 35. 238-319. (A key paper: Provides the first workable WMDS algorithm. and one that is still in very

wide use. Generalizes singular value (Eckart-Young) decomposition to N-way tables.)

[2] Deleeuw, J. (1977). In: *Recent Developments in Statistics. 1.* R. Barra et al. eds. North-Holland. Amsterdam. (Advanced mathematical paper that proposes the SMACOF algorithm and proves its convergence. Difficult but elegant.)

[3] Deleeuw. J. and Heiser. W. J. (1977). In *Geometrical Representatioms of Relational Data.* J. C. Lingoes. ed. Mathesis Press. Ann Arbor, M[. (Continues the work published in the preceding reference.

[4] Jacobowitz. D. (1973). "Development of semantic structures." Unpublished Ph.D. dissertation. Universitv of North Carolina at Chapel Hill.

[5] Kruskal. J. B. *(1964). Psrchometrika.* 29. 1-27; 115-129. (Completes the second major MDS breakthrough started by Shepard by placing Shepard's work on a firm numerical analysis foundation. Perhaps the most important paper in the MDS literature.)

[6] Kruskal, J. B., and Wish. M. (1977). *Multidimensional Scaling.* Sage Publications. Beverly Hills. CA. (Very readable and accurate brief introduction to MDS that should be read by everyone wanting to know more.)

[7] McGee. V. C. (1978). *Multivar. Behav. Res., 3.* 233-248.

[8] Messick. S. J. and Abelson. R. P. (1956). *Psychometrika.* 21, 1-17.

[9] Ramsay. J. 0. (1982). *J. Royal Statistical Society, A. vol. 145.* 285-312. (Foundations for one aspect of the current state of the art. Introduces hypothesis testing into the MDS framework, providing statistical tests to help decide on the appropriate dimensionality and model.)

[10] Ramsay, J. 0. *(1982). Multiscal 2 Manual.* Department of Psychology, McGill University, Montreal, Canada. (Very high-quality user's guide to the program based on the preceding reference.)

[11] Richardson, M. W. (1938). *Psychological Bulletin, 35,* 659-660.

[12] Roskam, E. E. *MINISSA Standard Version.* Nijmegen Mathematics-Psychology Department. University of Nijmegen. Nijmegen. Holland. (The MINISSA user's guide.)

[13] SAS Institute. (1980). *SAS Supplemental Librarv User's Guide.* SAS Institute. Cary, NC.

[14] Schiffman. S. S.. Reynolds, M. L., and Young. F. W. (1981). *Introduction to Multidimensional Scaling.* Academic Press, New York

[15] Shepard, R. N. (1962). *Psychometrika.* 27, 125140: 219-246. (Started the second major MDS breakthrough by proposing the first nonmetric algorithm. Intuitive arguments placed on firmer ground by Kruskal.)

[16] Stoop. L. and de Leeuw. J. (1982). *How to Use SMACOF-IB.* Department of Data T'heory. University of Leiden. The Netherlands. (A complete user's guide.)

[17] Takane, Y., Young. F. W. and de Leeuw, J, (1977). *Psychometrika.* 42. 7-67 (the third major MDS breakthrough. Combined all previous major MDS developments into a single unified algorithm.)

[18] Torgerson. W. S. (1952). *Psychometrika.* 17. 401-419. (The first major MDS breakthrough.)

[20] Young. F. W. (1981). *Psychometrika.* 46. 357-388. (A readable overview of nonmetric issues in the context of the general linear model and components and factor analysis.)

[21] Young. F. W. (1984). *Research Methods for Multimode Data Analvsis in the Behavioral Sciences.* H. G. Law, C. W. Snyder, J. Hattie, and R. P. MacDonald, eds. (An advanced treatment of the most general models in MDS. Geometrically oriented. Interesting political science example of a wide range of MDS models applied to one set of data.)

[22] Young. F. W. and Hamer. R. M. (1994). *Theorv and Applications of Multidimensional Scaling.* Eribaum Associates. Hillsdale, NJ. (The most complete theoretical treatment of MDS and the most wide-ranging collection of applications.)

[23] Young. F. W. and Lewyckyj, R. (1979). *ALSCAL-4 user's guide.* 2nd ed. Data Analysis and