

Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS

Gyslain Giguère

Université du Québec à Montréal

This paper aims at providing a quick and simple guide to using a multidimensional scaling procedure to analyze experimental data. First, the operations of data collection and preparation are described. Next, instructions for data analysis using the ALSCAL procedure (Takane, Young & DeLeeuw, 1977), found in SPSS, are detailed. Overall, a description of useful commands, measures and graphs is provided. Emphasis is made on experimental designs and program use, rather than the description of techniques in an algebraic or geometrical fashion.

In science, being able to synthesize data using a smaller number of descriptors constitutes the first step to understanding. Hence, when one must extract useful information from a complex situation implying many hypothetical variables and a huge database, it is convenient to be able to rely on statistical methods which help finding some sense by extracting hidden structures in the data (Kruskal & Wish, 1978). Torgerson (1952), among others, proposed such a method, called multidimensional scaling (MDS). At the time, he believed that while the use of psychophysical measures was appropriate for certain types of experimental situations in which comparing dimension values turned out to be fairly objective (Weber's law and the Just Noticeable Differences paradigm, for example), most of the situations encountered by experimental psychologists involved knowing neither beforehand the identity nor the number of psychologically relevant dimensions stemming from the data set.

In essence, MDS is a technique used to determine a n -dimensional space and corresponding coordinates for a set of objects, strictly using matrices of pairwise dissimilarities

between these objects. When using only one matrix of similarities, this is akin to Eigenvector or Singular value decomposition in linear algebra, and there is an exact solution space. When using several matrices, there is no unique solution, and the complexity of the model commands an algorithm based on numerical analysis. This algorithm finds a set of orthogonal vector dimensions in an iterative fashion, slowly transforming the space to reduce the discrepancies between the inter-object distances in the proposed space, and the corresponding scaled original pairwise dissimilarities between these objects.

A classic example, found in virtually all introductory books on multidimensional scaling (see for example Kruskal & Wish, 1978), fully illustrates the usefulness of MDS (see Figure 1). Estimating distances between a few pairs of U.S. cities could be executed quite easily by using a ruler and a map of the United States of America. But what if the opposite had to be done? What if an individual was given a matrix of distances between pairs of cities and had to draw the map using strictly these distances? This task would be quite tenuous. That is where MDS becomes useful.

In psychology, one rarely needs to use direct physical distances. However, measuring the similarity between objects similarity is an important concept in most areas of cognition. MDS is therefore mainly used to "help systematize data in areas where organizing concepts and underlying dimensions are not well-developed" (Schiffman, Reynolds & Young, 1981, p.3). It can be used to explore and discover the defining characteristics of unknown social and psychological structures, but also to confirm a priori

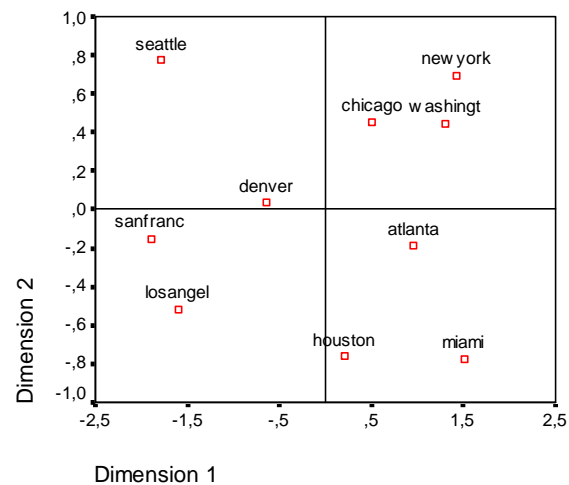
The author would like to thank Sébastien Hélie for comments on this paper, as well as the Fonds québécois pour la recherche sur la nature et les technologies (NATEQ) for financial support in the form of a scholarship. Correspondence concerning this article should be sent to: Gyslain Giguère, UQÀM, Département de psychologie, LÉINA, C.P. 8888, Succ. CV, H3C 3P8. E-mail: giguere.gyslain@courrier.uqam.ca.

Figure 1. Upper panel: data matrix containing intercity distances for 10 U.S. cities. Lower panel: optimal two-dimensional configuration computed by SPSS ALSCAL.

	atlanta	chicago	denver	houston	losangel	miami	newyork	sanfranc	seattle	washingt	cityname
1	Atlanta
2	587	Chicago
3	1212	920	Denver
4	701	940	879	Houston
5	1936	1745	831	1374	Los Angeles
6	604	1188	1726	968	2339	Miami
7	748	713	1631	1420	2451	1092	New York
8	2139	1858	949	1645	347	2594	2571	.	.	.	San Francisco
9	2182	1737	1021	1891	959	2734	2408	678	.	.	Seattle
10	543	597	1494	1220	2300	923	205	2442	2329	.	Washington, DC

Derived Stimulus Configuration

Euclidean distance model



hypotheses about these structures. Usually, MDS analysis starts from data representing similarity scores between objects, and tries to identify which dimensions could have been used for object comparison, for instance. MDS can also be used to test the validity of hypotheses about specific psychological measures used in differentiating objects (Broderson, 1968, in Borg & Groenen, 1997), and identify significant object groupings.

In the present paper, the basics of collecting and analyzing similarity data are described. The first section focuses on data collection methods and experimental design in general, while the second section concentrates on program use and output interpretation.

Experimental design

Collecting data

Practically any matrix of data, representing individual degrees of relation between items, can be used in MDS, those of interest for cognitive psychologists being primarily similarities and dissimilarities (of course), rank-orders, and confusion data. The exact MDS model to be used is influenced by the goal of the analysis, but is mainly determined by the data set's characteristics (which are defined later in the text).

Different psychological tasks can be used to collect similarity data. In the most common experimental task, namely *pairwise comparison* (as used in Shin & Nosofsky, 1992), participants are asked to judge the resemblance or

difference between two objects which are presented simultaneously or sequentially. They are often instructed to respond by moving a cursor to the desired position on a continuous visual scale defining a similarity continuum. This is called the *graphic rating method* (Davison, 1983). Another way of collecting these judgments is to ask participants to report the perceived similarity level using a choice of discrete numbers from a predefined scale. For example, in a pairwise comparison task, the number "1" could mean "highly similar", while the number "9" would mean "highly dissimilar", and all discrete numbers in between would represent various levels of similarity. Because the possible answers are limited in number and discrete, this is called the *category rating method* of collection (Davison, 1983).

In *magnitude estimation tasks* (Stevens, 1971), a certain stimulus pair is chosen as a standard on each trial. Each of the remaining pairs of stimuli is to be judged against the standard, in a relative way. For example, if the objects from a given pair look four times as dissimilar as the standard, the participant would give "4" as an answer, and if another pair looks half as dissimilar as the standard pair, the participant would give "1/2" as an answer. The estimated dissimilarity of a specific pair is equal to the geometric mean of all the judgments assigned to it by different participants. Therefore, only one matrix of dissimilarities is produced, whatever the number of participants.

Another variant which uses a standard is the *conditional rank-ordering* task (see Schiffman et al., 1981; also called the *anchor stimulus method*, Borg & Groenen, 1997): for each round of trials, all stimuli are presented simultaneously, and a single stimulus is chosen as a standard. The participant is asked to determine which other stimulus from the set is most similar to the standard. This stimulus is given the highest rank, and is removed from the set. In an iterative fashion, the participant must decide which remaining stimulus is now most similar to the standard, until all stimuli have been ranked. The standard, which is chosen randomly, must be different for each round of trials, so that at the end of the experiment, every stimulus has played that role. Ranking can also be constraint-free. For that purpose, each object pair is typically presented on a card. The participant is then asked to sort these cards so that the most similar object is on top of the card stack and the most dissimilar one at the bottom.

In a *free sorting* task, a participant is presented with all stimuli simultaneously, and is asked to divide the set in an undefined number of subsets containing objects that appear similar in some sense. At the end of the task, two objects from the same group are given a similarity of "1", as opposed to stimuli from different groups, who are given a

score of zero. One has to be careful with this tasks, since the well-known fact that participants naturally tend to judge inter-object similarity using very few attributes could lead to a very low number of subsets (see Ahn & Medin, 1992; Regehr & Brooks, 1995).

A related task is the *category sorting* task (Davison, 1983). Here, each possible pair of stimuli is printed on a separate card. Participants must classify pairs in a specified number of groups each representing a specific level of similarity. Once the task has been achieved, each pair is given a score according to its similarity category membership. For example, pairs from the "highly similar" category are given the lowest ranking score, "1", while pairs from the "highly dissimilar" group are given a rank equal to the number of predefined categories. The experimenter can decide to constrain the number of cards per group, and ask that all the groups contain at least one card to avoid use of too few similarity levels.

Finally, when few stimuli are used, discrimination and identification tasks can provide indirect similarity measures. The logic behind these two tasks is that as two items are more and more similar, they should be more and more difficult to discriminate. Hence, stimulus confusability can be used as a measure. Stimulus-stimulus confusions in discrimination tasks occur when a participant is presented with a pair of different stimuli and asked if the two stimuli are the same or different. The number of "same" responses to "different" pairs is the measure of confusability (and indirectly, of similarity) for a given pair of stimuli.

Experimental design issues

Similarity or dissimilarity?

For technical reasons, most authors (such as Young & Harris, 2004) encourage the use of dissimilarities as input to the MDS program, because their relationship to distances is direct and positive (that is, the higher the dissimilarity, the larger the perceived psychological distance). If similarities have been collected, Kruskal & Wish (1978) recommend that they be transformed by subtracting the original data values from a constant which is higher than all collected scores.

Trial ordering

In a MDS task, stimulus ordering is subject to two particular problems, namely position and timing effects. Position effects occur when an item is too often in the same position of presentation within a pair (for instance, if items are presented simultaneously, it appears too often on the left part of the screen). Timing effects for a given stimulus occur when the pairs in which that stimulus appears are not equally spaced throughout the trial list. To avoid these

effects, the scientist may choose to use Ross ordering (Ross, 1934), a technique used to balance position and time effects by explicit planning when the number of items to be compared is low. If this is not the case, random ordering should then be used: there is however no guarantee that position and timing effects are avoided, but they are kept to a minimum over replications.

Reducing the number of necessary judgments

In multidimensional scaling, the more judgments are collected for each stimulus pairs, the more points can be fit in an n -dimensional space. An analysis with more points provides a more robust and precise stimulus space. That is why researchers usually prefer obtaining complete judgment matrices from many participants. However, with n items, a complete square similarity matrix is composed of $n(n-1)$ possible cells or comparisons (when excluding identical pairs), and this number grows rapidly when adding more stimuli. Because the number of produced stimulus pairs in a design may be too high to be judged by a single participant, there are a few ways to reduce this number that have been proposed. First, if theoretically supported, one may assume that judgments are symmetric: this reduces by one half the number of required trials. This assumption is actually taken for granted in most psychological experiments, even if it was never proven (Schiffman et al., 1981). To avoid unknown asymmetry effects and respect acceptable ordering characteristics, the experimental program may be made to “flip the coin” before each trial to randomly determine the position of each stimulus in the presentation (left or right if the items are presented simultaneously, first or second if they are presented sequentially).

Second, pairs may be randomly distributed over participants, with the set of pairs judged by different participants being either completely independent or overlapping (in both cases, no pair should be excluded). With a large number of participants, these subsets can be created in a random fashion (Spence & Domoney, 1974). This produces a robust result when using Classical MDS (CMDS), mainly because judgments are generally averaged over participants, and this produces a complete matrix, but is not recommended when using models with replications or weights (RMDS and WMDS). Missing data also leads to this robustness difference, and should be avoided at all costs when not using CMDS.

In all cases, the number J of recommended judgments per pair of stimuli used in the MDS analysis should be equal to:

$$J = \frac{40D}{(I-1)}$$

where D equals the maximal anticipated number of dimensions, and I represents the number of items used in the experiment (MacCallum, 1979).

Analyzing dissimilarity data

Data and measurement characteristics

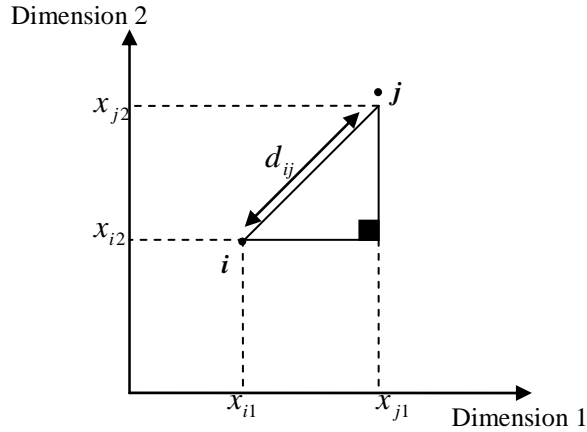
Data Levels

With MDS, data matrices can be defined by considering many of their characteristics, namely their measurement level, shape, and conditionality. According to Coombs' (1964) data theory, generally speaking, there are four levels of data measurement, which are ordered from the weakest to the most stringent. The first one is the nominal (or categorical) level, where objects are simply sorted into a limited number of groups. This level of data is not recommended for use unless the number of categories is quite large. With the ordinal level, objects are arranged in order of magnitude, but the only available information is their comparative ranking. No numerical relationship holds between objects. When using the interval or ratio levels, objects are placed on a scale such that the magnitude of the differences between objects is shown by the scale. The difference between these levels is that while the ratio level can lead to relative differences (as in “object x is twice or three times as large or fast as object y ”), in the interval level, there is no absolute zero, which prevents this kind of conclusion. In both types of measurement levels, however, a precise difference between values is always the same, wherever it is situated on the scale (e.g. the difference between 20 and 50 is the same as the difference between 70 and 100).

Data shapes

Data shapes are twofold. Square data occurs when the same objects are represented by the rows and columns of the matrix. This generally happens when all objects are compared to each other. Hence, the number of rows and columns are identical. When the order of presentation within a trial has no effect, that is if the similarity between objects a and b is the same whichever object is presented first, then the data is said square symmetric. In the opposite case, if the order of presentation affects the value of similarity, the data is square asymmetric. Rectangular data usually occurs when the objects represented by the rows are different than the ones represented by the columns. An example would be if the rows represented different individuals, and the columns different psychological test scores¹. By definition, rectangular data is asymmetric. In this paper, emphasis is put on square data.

Figure 2. Two-dimensional Euclidian space. The Euclidian distance between points i and j is the hypotenuse of the hypothetical right triangle.



Measurement conditionality

The result of data collection is that a certain number of square or rectangular dissimilarity matrices equal to the number of participants are obtained. Depending on the role played by individual differences and the shape of the data, different conditionality statuses define measurement.

Data is said to be *matrix conditional* if there are hypothesized individual differences. This means that data from a specific data matrix can meaningfully be compared to each other, but cannot be compared to data from other matrices. Data from direct similarity rating usually falls in this category. *Unconditional* data matrices can be meaningfully compared to each other. For example, if one were to measure response times from confusion errors, this objective measure could be compared across participants.

The Euclidian model

MDS algorithms such as SPSS ALSCAL use the Euclidian model as a basis to compute optimal distances between objects in an n -dimensional stimulus space. The related distance function, Euclidian distance, corresponds to our everyday experience with objects (Schiffman et al., 1981). It is derived from the Pythagorean Theorem, and is defined as the length of the hypotenuse linking two points in an hypothetical right triangle (Figure 2). The distance function for a Euclidian stimulus space is given by:

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

where d_{ij}^2 is the squared Euclidean distance between points i and j , and x_{ia} and x_{ja} are the respective coordinates of points i and j on axis a .

If perceptual/cognitive differences in the use of psychological dimensions are assumed, the distance

calculation becomes:

$$d_{ijk}^2 = \sum_a w_{ka} (x_{ia} - x_{ja})^2$$

where d_{ijk}^2 is the squared Euclidean distance between points i and j for participant k , x_{ia} and x_{ja} are the respective coordinates of points i and j on coordinate a , and w_{ka} ($0 \leq w_{ka} \leq 1$) represents the weight given to dimension a by participant k . A higher weight on a given dimension has the effect of stretching the stimulus space on that particular dimension, while a lower weight has the opposite effect, namely shrinking.

The metric vs. nonmetric distinction

To create an n -dimensional map of distances, MDS algorithms must use a function to scale the original dissimilarities into "disparities", which are directly comparable to the obtained distances from the analysis. For this purpose, two types of function may be used (Figure 3). Torgerson (1952) proposed the use of a linear function to map the original data onto "disparities":

$$\delta_{ij} = f(s_{ij}) = as_{ij} + b$$

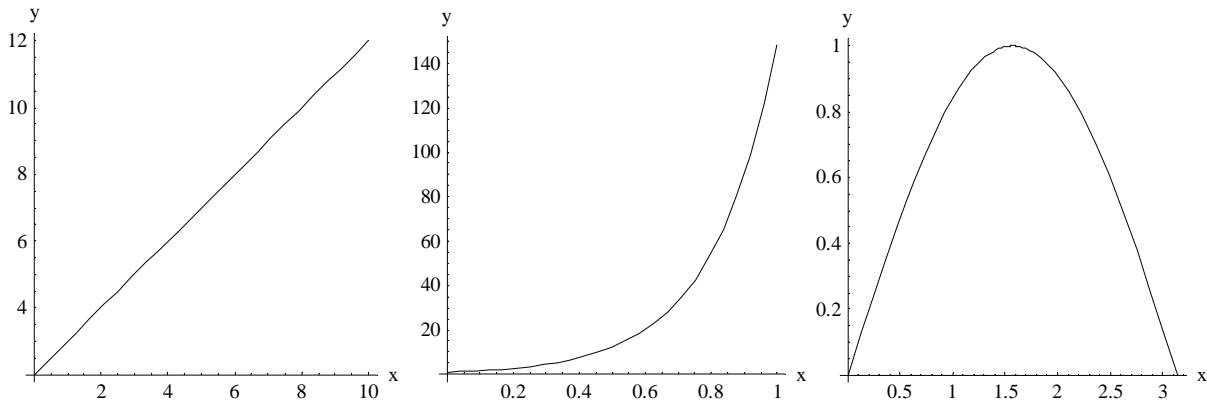
where δ_{ij} is the calculated disparity between objects i and j , s_{ij} is the original dissimilarity score for this pair of objects, and a and b are the slope and intercept of the linear function ($a \geq 0$). An analysis using this transformation function is called *metric* MDS.

Shepard (1962a, 1962b) later discovered that metric information could be recovered even with weaker, non-metric data. He found that with rank-order data, the choice of a linear function was too stringent, and proposed the use of any positive monotonic function (not necessarily a linear one) as sufficient to achieve the analysis. A positive monotone function is defined as a transformation which respects the rank order of the data, or more precisely where the following relationship is respected:

$$(s_{ij} \leq s_{ik}) \Rightarrow f(s_{ij}) \leq f(s_{ik})$$

where s_{ij} is the original dissimilarity measured between objects i and j , s_{ik} is the original dissimilarity measured between objects i and k , f is a positive monotonic function, $f(s_{ij})$ is equal to δ_{ij} , the disparity between objects i and j , and $f(s_{ik})$ is equal to δ_{ik} , the disparity between objects i and k . An analysis using this type of function is called *nonmetric* MDS. When using nonmetric Classical MDS, a single transformation is used, while with nonmetric Replicated MDS, a different function is used for each different data matrix. All models discussed in this paper can be used with both function types.

Figure 3. Left panel: example of a positive linear function. Middle panel: example of a positive monotonic function which is not linear, namely the exponential functions. Right panel: example of a non-monotonic function.



Defining specific MDS models²

Classical MDS (CMDS) is a model which uses only one matrix of raw or averaged data, which is matrix-conditional. When using this model, the algorithm produces a hypothetical Euclidian stimulus space which matches the original data as much as possible. The first step is to transform the original dissimilarities into disparities using a linear ($l(S)$ – for metric MDS) or monotonic ($m(S)$ – for nonmetric MDS) function.

The model equation to be fit by ALSCAL is then:

$$T(S) = D^2 + SSE$$

where S is the original dissimilarities matrix, $T(S)$ is a disparity matrix stemming from the transformation T , equal to $l(S)$ or $m(S)$ depending on whether the data are interval/ratio (metric MDS) or ordinal (nonmetric MDS), D^2 represents the squared Euclidian distances fit by ALSCAL, and SSE is the sum of squared errors between the distances and disparities. CMDS is the less robust analysis, because the algorithm only gets to fit a number of points equal to the number of stimulus pairs (or less if the matrix is symmetric or if an incomplete data scheme has been used).

In Replicated MDS (RMDS), several matrices of data are used. These data are usually defined as matrix conditional. Once again, only one stimulus space is produced. Because systematic response bias differences between participants are included in the analysis (reflecting the different ways they use the response scale), the first step is to determine several transformation functions, one for each dissimilarity matrix. Once again, these transformations can be either all linear or all monotonic, but within these constraints, can all be different. All matrices are judged to be linearly or monotonically related, except for error. The model equation to be fit by ALSCAL is then:

$$T_k(S_k) = D^2 + SSE_k$$

where S_k is the original dissimilarities matrix for participant k , $T_k(S_k)$ is an individual disparities matrix for participant k stemming from a unique transformation T_k , once again equal to $l_k(S_k)$ or $m_k(S_k)$ depending if the analysis is metric or non metric. D^2 represents the squared Euclidian distances fit by ALSCAL for the common stimulus space, and SSE_k is the sum of squared errors between the distances and disparities for participant k . RMDS is far more robust than CMDS, because the algorithm can use an increased number of points, stemming from the fact that all data from all matrices are used in the analysis.

The last type of MDS explored in this paper is Weighted MDS (WMDS – also known as Individual differences scaling or INDSCAL). In this type of analysis, SPSS ALSCAL provides the usual stimulus space, but also a participant space which indicated the differential weighting given to dimensions in the common stimulus space by each participant, and the models fit to each participant's data. In this model, several matrices of either matrix-conditional or unconditional data are used. In WMDS, any two participants' personal distances need not be related by any linear or monotonic function. The model equation to be fit by ALSCAL is then:

$$T_k(S_k) = D_k^2 + SSE_k$$

where S is the original dissimilarities matrix for participant k , $T_k(S_k)$ is an individual disparities matrix for participant k stemming from a unique transformation T_k , once again equal to $l_k(S_k)$ or $m_k(S_k)$ depending if the analysis is metric or non metric. D_k^2 represents the squared Euclidian distances fit by ALSCAL for participant k , and SSE_k is the sum of squared errors between the distances and disparities for participant k . The distances are found in participant k 's personal stimulus space, which can be recovered by applying the personal weights to the common distance space, as in:

$$X_k = XW_k^{1/2}$$

Table 1. Decision table relating data characteristics to their appropriate MDS model

Shape of data?	Square					
	One		Several			
Number of matrices?	One		Several			
Perceptual/cognitive differences assumed?	N/A		No		Yes	
Measurement Conditionality?	Unconditional		Matrix conditional		Matrix conditional	
Data level?	Interval or Ratio	Ordinal or Nominal	Interval or Ratio	Ordinal or Nominal	Interval or Ratio	Ordinal or Nominal
MDS Model	Metric CMDS	Nonmetric CMDS	Metric RMDS	Nonmetric RMDS	Metric WMDS	Nonmetric WMDS

Note. To find the appropriate model, one must answer all questions sequentially, from top to bottom. For example, if working with several square matrices of interval data, where no perceptual/cognitive differences are assumed, the user's path would be "Square→Several→No→Matrix conditional→Interval or Ratio, and the reader should conclude that the use of metric RMDS is appropriate.

where X_k is the coordinate matrix representing participant k 's stimulus space, X is the coordinate matrix from the common stimulus space, and W_k is the weight matrix for participant k . WMDS possesses the robustness of RMDS, but also provides some flexibility, because the stimulus space does not have to "directly" fit every matrix of data.

The reader is encouraged to use Table 1 to determine the exact MDS model needed for the analysis, depending on the shape and level of the data, the number of similarity matrices used in the analysis, the measurement conditionality, as well as the decision to take into account psychological differences between participants.

Syntax use

To achieve an MDS analysis, the data must first be entered in matrix fashion in an SPSS data file. Figure 4 shows an example file with three hypothetical data matrices. Each matrix represents the similarity judgments collected for one participant (*subject*). For identification purposes, the subject number is put in the first column, along each line of this participant's data matrix. However, when entering many square matrices, no identifier has to be provided for each matrix; ALSCAL achieves the file separation by itself. Each column's variable name (*object1*, *object2*, *object3*, *object4*) is attributed to the compared object's identity. These variable names will be used as identifiers in the output and graphs. As can be seen, when considering that the data are symmetric (as is the case here), one does not need to enter data above the matrix diagonal (nor on the diagonal for that matter, because any object is always minimally dissimilar to itself). When

entering many matrices, no identifier has to be provided for each matrix; ALSCAL achieves the file separation by itself. Using this file, the user can produce different RMDS or WMDS, assuming that the data are either

To use the ALSCAL procedure in SPSS syntax, the user must enter the ALSCAL command, followed by the list of variables (names of the data columns), in the following fashion:

ALSCAL VARIABLES = v1 to vn

where v_1 represents the first column of data, and v_n represents the last one. Following this command, many

Figure 4. SPSS file representing three data matrices. Each matrix contains inter-object dissimilarity judgments. The reader is encouraged to analyze the example data using the corresponding file found on the journal's website (www.tqmp.org), and the code examples from Table 2.

subject	object1	object2	object3	object4	object
1	1	.	.	.	Object 1
2	1	6	.	.	Object 2
3	1	1	4	.	Object 3
4	1	3	5	2	Object 4
5	2	.	.	.	Object 1
6	2	5	.	.	Object 2
7	2	1	4	.	Object 3
8	2	3	6	2	Object 4
9	3	.	.	.	Object 1
10	3	6	.	.	Object 2
11	3	1	4	.	Object 3
12	3	3	5	2	Object 4

Table 2. Default values and typical examples of syntax for different MDS models

Default values
<code>/SHAPE=SYMMETRIC</code>
<code>/LEVEL=ORDINAL</code>
<code>/CONDITION=MATRIX</code>
<code>/MODEL=EUCLID</code>
<code>/CRITERIA=CUTOFF(0) CONVERGE(.001)</code> <code>ITER(30) STRESSMIN(.005)DIMENS(2)</code>
<code>/PRINT=DATA</code>
<code>/PLOT=DEFAULT</code>
Metric CMDS or RMDS
<code>ALSCAL VARIABLES = v1 to vn</code>
<code>/LEVEL=INTERVAL or RATIO</code>
<code>/CRITERIA= DIMENS(1,6)</code>
<code>/PRINT=DATA HEADER</code>
<code>/PLOT=DEFAULT ALL.</code>
Metric WMDS
<code>ALSCAL VARIABLES = v1 to vn</code>
<code>/LEVEL=INTERVAL or RATIO</code>
<code>/MODEL=INDSCAL</code>
<code>/CRITERIA= DIMENS(2,6)</code>
<code>/PRINT=DATA HEADER</code>
<code>/PLOT=DEFAULT ALL.</code>

Note. When achieving a nonmetric analysis, the LEVEL subcommand values are ORDINAL or NOMINAL. In WMDS, note that the minimal dimensionality is always 2. Any block of syntax in SPSS must end with a period; else the program is not executed.

subcommands can be entered to specify the type of analysis needed, and other necessary criteria. Each subcommand must be preceded by a backslash. Examples of typical syntax blocks for different MDS types are detailed in Table 2.

The first subcommand is SHAPE, which describes the shape of the data. The eligible values are SYMMETRIC (for square symmetric data), ASYMMETRIC (for square asymmetric data), and RECTANGULAR. The following subcommand concerns the LEVEL of the data. It determines if the algorithm should use a metric or nonmetric transformation function. The values which can be entered are ORDINAL, INTERVAL, RATIO and NOMINAL. Measurement conditionality is defined by the CONDITION subcommand, which takes the values MATRIX (for matrix-conditional data), ROW (for row-conditional data), or UNCONDITIONAL. Next is the MODEL subcommand. Five options are available, but the user usually limits himself to the basic Euclidian model (represented by the value EUCLID) or the Individual differences scaling model (for which the appropriate value is INDSCAL).³

The CRITERIA subcommand is special, because it is composed of many parameters, for which different values

must be provided. For each parameter, the wanted values must be inserted between parentheses. The first parameter is the CUTOFF, which specifies the lower bound for the scores found in the data matrices. By default, this is left at zero ("0"), so that only negative scores are eliminated. The three next parameters provide a disjunctive stopping rule for the algorithm (i.e. as soon as the algorithm reaches a critical value for one of the parameters, the fitting procedure ends). CONVERGE represents the minimal S-STRESS (a measure defined in the next section) improvement needed for a supplemental iteration. ITER defines the maximal number of iterations for the analysis. Finally, STRESSMIN defines a minimal cutoff for the S-STRESS value. For a given iteration, if the value is equal to or lower than the cutoff, the program stops.

One last parameter is the number of dimensions of the n-dimensional space. If a single value is put in the parentheses, only one solution is computed. If a pair of values (min, max) are inserted, ALSCAL computes different solutions for all values included in the specified interval. If the researcher has an hypothesis about the number of necessary dimensions, it is then recommended to repeat the analysis using from $d-3$ to $d+3$ dimensions, d being the number of hypothesized dimensions. ALSCAL can produce outputs using from one to six dimensions.

The last relevant subcommand pertains to the content of SPSS's output. For the PRINT subcommand, putting the DATA value cause ALSCAL to print all matrices of original and transformed data, while adding the HEADER value produces a summary of all options. It is recommended to include both values.

Interpreting the output

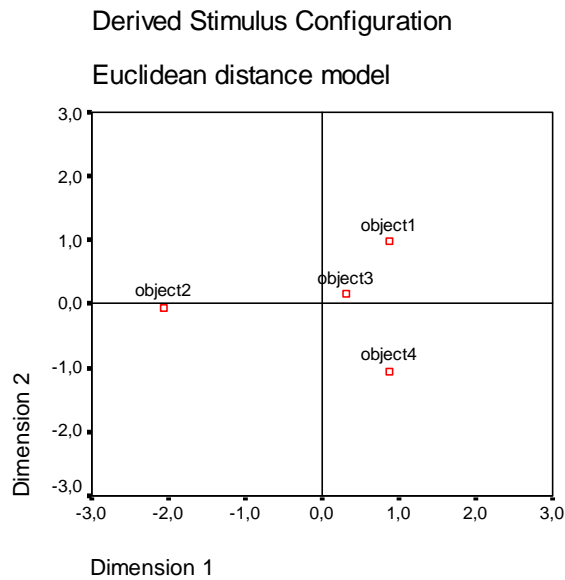
Measures of fit

To determine the badness-of-fit between the hypothesized structure and the original data, SPSS ALSCAL uses a loss function called S-STRESS, which is derived from the STRESS⁴ measure proposed by Kruskal (1964). When analyzing similarity data, S-STRESS Formula 1 (SS1) should be chosen to fit the model:

$$SS1 = \left[\frac{\sum_{(i,j)} (\delta_{ij}^2 - d_{ij}^2)^2}{\sum_{(i,j)} (d_{ij}^2)^2} \right]^{1/2}$$

where δ_{ij}^2 is the squared disparity between items i and j , d_{ij}^2 is the related squared distance, I is the number of rows and J is the number of columns in the matrix. When data are symmetric, the sums can be constrained to the cases where $1 = i < j = n$, where n represents the dimension of the square

Figure 5. Optimal two-dimensional configuration computed by SPSS ALSCAL for the data presented in Figure 4.



matrix, i represents the row number, and j represents the column number. In other words, only data from the lower triangular part of the matrix should be used. A value of zero means a perfect fit.

When using CMDS, RMDS or WMDS, the user is provided with the S-STRESS measure computed after each complete program iteration. This is called the “iteration history”. SPSS ALSCAL also gives the STRESS measure for the last iteration, as well as the R-squared (r^2), which represents the level of variance in the data which is explained by the hypothesized n-dimensional configuration.⁵

If using RMDS or WMDS, the S-STRESS measure given in the iteration history is calculated differently; for example, SS1 is calculated using the following formula:

$$SS1 = \left[\frac{1}{m} \sum_k (SS1_k)^2 \right]^{1/2}$$

where $SS1_k$ is the corresponding S-STRESS measure calculated for participant k , and m is the number of data matrices (or participants) entered in the analysis. In RMDS, STRESS and r^2 are then displayed for each stimulus in every data matrix, as well as averaged for each matrix. In WMDS, only the matrix averages are displayed.

There are unfortunately no guidelines to interpret the level of fit from S-STRESS. However, some hints have been given for the original STRESS measure, which is also given by SPSS ALSCAL after completion of the analysis. It should be kept in mind that all fitting is done using S-STRESS however, so the final STRESS score may actually not be

optimal. For STRESS, Kruskal & Wish (1978) have proposed meanings using the following levels: STRESS > .20: Poor; $.10 \leq \text{STRESS} \leq .20$: Fair; $.05 \leq \text{STRESS} \leq .10$: Good; $.025 \leq \text{STRESS} \leq .05$: Excellent; .00: Perfect. Of course, one has to be careful with such straightforward interpretations, since STRESS is known to vary according to many other factors. It is higher when using a metric distance function, as well as a higher number of stimulus pairs or data matrices, or if there is a high level of error in the data. It decreases with higher space dimensionality, missing data, and when using a nonmetric distance function.

Configuration

Following the fit measures, the optimal configuration derived in the number of required dimensions, is displayed as a list of n-dimensional coordinates for all stimuli. Also, the matrix of disparities calculated during the last iteration is printed. This information is paired with a n-dimensional graph of the configuration ($n \leq 3$) which can be hand rotated (stimulus points and dimensions simultaneously) using SPSS's graph options to facilitate dimension interpretation (Figure 5).

If WMDS is used, the output also includes information related to the participant space. The first supplemental information is the matrix of subject (participant) weights and weirdness indexes. These weights have the effect of

Figure 6. Example of a subject weight space for three participants. The points represent the endpoints of the corresponding weight vectors. For participant 1, dimension 1 is mostly used, while for participant 3, it is dimension 2 that is mostly used in similarity judgements. For participant 2, dimension 1 is almost twice as important as dimension 2 when judging similarity.

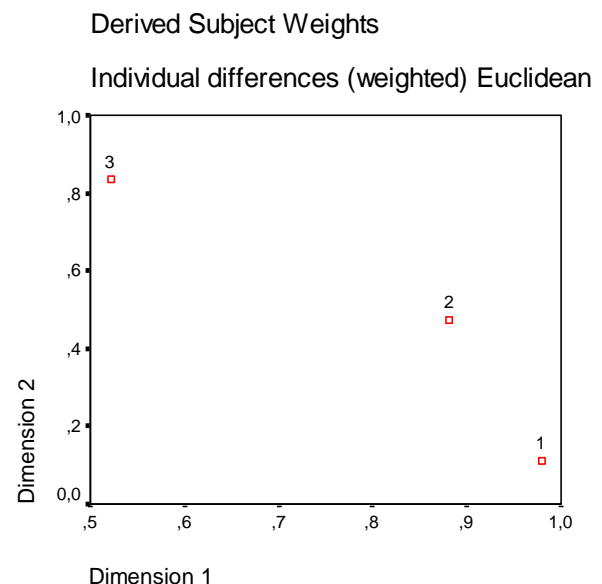
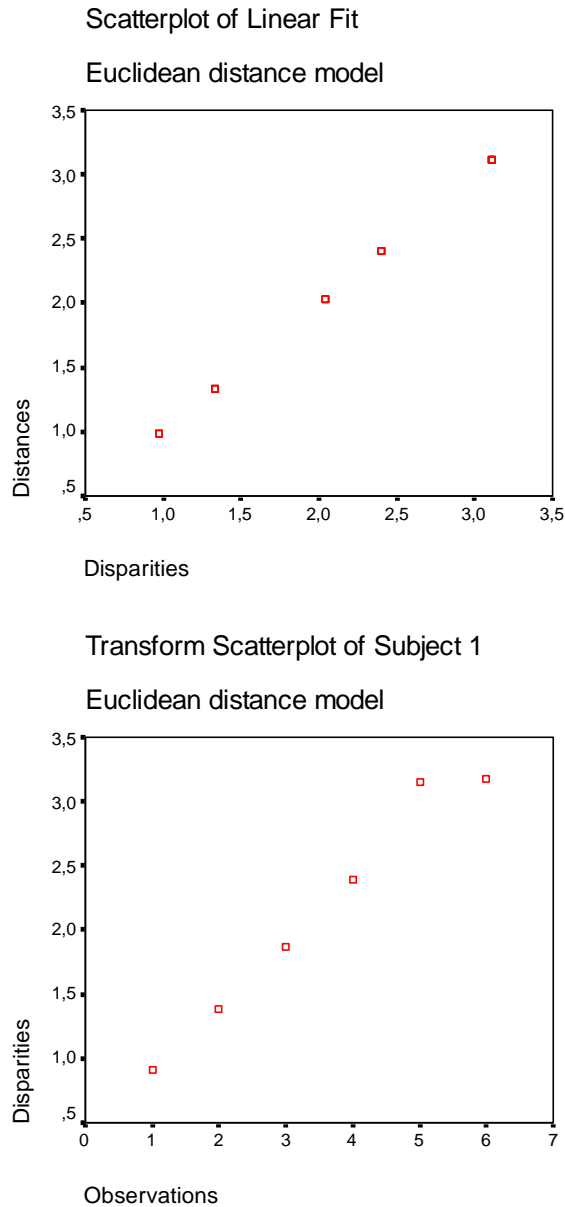


Figure 7. Upper panel (a): Linear fit scatterplot for the example exposed in Figure 4. Lower panel (b): Example of a transformation scatterplot. This plot is only produced when nonmetric MDS is used.



“morphing” the common stimulus space to fit each participant’s data. For each participant, the coordinates for his weight vector are given. The direction of this vector from the origin of the configuration represents the relative weighting of each dimension. The Euclidian norm $\|\mathbf{w}_k\|$ for participant k ’s weight vector \mathbf{w}_k is calculated using the following formula:

$$\|\mathbf{w}_k\| = \sqrt{\sum_a w_{ak}^2}$$

where w_a^2 represents the square of the coordinate on

dimension a for participant k , and it is also equal to r^2 or the level of variance from the data explained by the weight vector for this participant.

The weirdness index is a measure of the fit of the general configuration to each data matrix. If, from a participant’s weight vector, it can be seen that some of the weights are quite higher than the others, then the optimal configuration fits this participant’s data very poorly. Weirdness indexes range from zero to one. One should be careful about including data from a participant showing an index higher than 0.5.

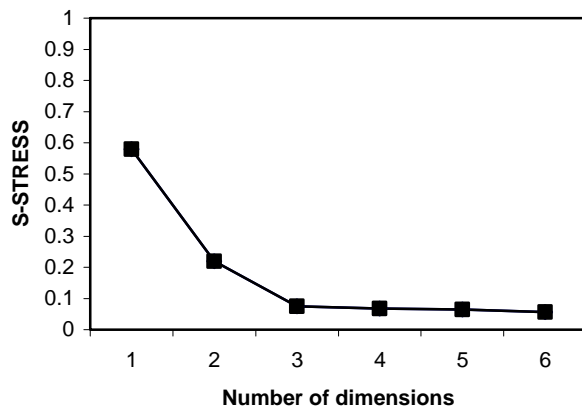
Second, the user is also given a graph containing the subject space (Figure 6). The subject weights are represented by points, which are the endpoints of the weight vectors.

Goodness-of-fit graphs

With CMDS, apart from the stimulus configuration, the user is given only one graph by default. It is the “linear fit scatterplot” (Figure 7a), where disparities are plotted against distances. Because there is only one matrix of data, there are usually very few points in the graph. Thus, it can be advantageous to look at this graph, since it can reveal problems which have occurred during the analysis such as a lack of convergence for the optimal solution. This phenomenon occurs when the program has stopped before the necessary number of iterations, and is shown by clear non-linear or negative patterns in the graph. It can be corrected by raising the number of iterations (non-linear patterns) and verifying if the data have been transformed to dissimilarities (negative patterns). In RMDS & WMDS, this lack of linear fit could be much harder to detect, because data points from all matrices are used. The key here is to look for an obvious linear trend. If unsure about the fit, one could use the given disparities and calculate the distances from the configuration coordinates to use linear regression.

If using nonmetric MDS, another graph plotting the original data (observations) against the disparities calculated by ALSCAL, called the “transformation scatterplot”, is provided (Figure 7b). If the plot does not show a positive monotonic relationship between the variables, it could once again mean that the program was not given a sufficient number of iterations to find a suitable transformation function, or that the data input to the program have not been pre-transformed. Transformations in the shape of a few horizontal steps could mean that a degenerate solution has been found. It is recommended to retry the analysis with the same number of dimensions if this occurs. In nonmetric RMDS & WMDS, a transformation scatterplot is printed separately for each data matrix. If only a few participants’ graphs do not show a positive monotonic trend, they could then be considered as outliers and removed from the

Figure 8. Example of a scree plot, with the S-STRESS measure for each number of dimensions. Here, there is an elbow at the three dimension mark, which usually means that a three-dimensional configuration should be chosen.



analysis. If too many participants show this trend, then it can be a sign of an incomplete or degenerate solution.

Post-analysis issues

Choosing the right dimensionality using S-STRESS

It is recommended that any MDS analysis be repeated using different dimensionalities. SPSS ALSCAL offers from 1 to 6 dimensions for model fitting. Once all the analyses have been made, S-STRESS becomes a useful measure to determine the optimal dimensionality of the stimulus space. By the use of "scree plots"⁶ (Figure 8), one can determine if adding an extra dimension significantly decreased the badness-of-fit. To achieve this, one should look for an "elbow" on the scree plot, which should appear in the graph one dimension beyond D , the correct number of dimensions. However, this is not an infallible solution. Davidson (1983), among others, proposed that one should generally choose the most interpretable dimensionality level; if adding a dimension does not help the interpretation, but removing one affects it, then the right level has been chosen.

Dimension interpretation

Once the n -dimensional map of distances and the coordinates matrix are produced by SPSS, the user may try to identify the meaning of the dimensions. Sometimes, interpreting dimensions can be quite straightforward (such as in Figure 1, where the dimensions are clearly the East-West and North-South axes). In other cases, the task can be simplified by using some transformations that can be applied to the dimension set or the coordinate matrix. These are called similarity transformations, since they preserve the ratio of distances between the points in the space. Apart

from orthogonal rotation, which can be done by hand using SPSS output's graph options, the other operations such as permutation, reflection, central dilation and translation (Young & Hamer, 1987) can be done using mathematical software such as Mathematica or Matlab. Using the matrix of coordinates from the n -dimensional stimulus space, it is possible to modify the basis by postmultiplying the matrix by:

- an n -dimensional, square matrix with a single "1" in each row and column. This matrix has the effect of permuting the order of the dimensions;
- an n -dimensional, square, orthogonal matrix with ± 1 s on its diagonal. These properties produce a reflection of the whole space on the dimensions to which the number "-1" has been applied;
- a diagonal, n -dimensional matrix containing equal positive values. If the values are larger than 1, the postmultiplication has the effect of stretching the dimensional space. If they are smaller, then the space is shrunk;
- adding a matrix which is the product by of a n -dimensional column vector of 1s and an n -dimensional row vector of constants, produces a shift of the set of points from one spot to the other in the stimulus space).

A combination of these transformations can be applied to the same space in a sequential way. Note that one has to be careful when using transformations with a WMDS model. It has been shown, among others, by Young & Hamer (1987), that stimulus spaces can not be rotated in that particular case..

Summary

This paper was intended as a hands-on guide to MDS-based experiments, mainly using SPSS. The reader is encouraged to reproduce the simple examples found in the paper. The files for these examples can be found on the journal's website, at www.tqmp.org.

While MDS seems like a very practical and straightforward way of exploring stimuli spaces, one should always be careful about giving strong interpretations about this type of analysis. There are no convincing statistical tests for validating common space interpretations, nor are there any ways to test the inter-participant differential weighting in WMDS. MDS remains to this day an exploratory method, which can get us "out of the dark", but does not provide certainty about the conclusions. Use with care.

References

- Ahn, W.K., & Medin, D.L. (1992). A two-stage model of category construction. *Cognitive Science*, 16, 81-121.
- Borg, I., Groenen, P. (1997). *Modern multidimensional scaling: theory and applications*. New York: Springer-

- Verlag.
- Coombs, C.H. (1964). A theory of data. New York: Wiley.
- Davison, M.L. (1983). Multidimensional scaling. New York: John Wiley & Sons.
- Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-28, 115-129.
- Kruskal, J.B., Wish, M. (1978). Multidimensional scaling. Beverly Hills, CA: Sage Publications.
- MacCallum, R.C. (1979). 3-MODE-MDS: A computer program to execute Tucker's three-mode multidimensional scaling. *Applied Psychological Measurement*, 3, 24
- Regehr, G., & Brooks, L.R. (1995). Category organization in free classification: the organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 347-363.
- Ross, R.T. (1934). Optimum orders for presentation of pairs in paired comparison. *Journal of Educational Psychology*, 25, 375-382
- Schiffman, S.S., Reynolds, M.L., Young, F.W. (1981). Introduction to multidimensional scaling: theory, methods, and applications. New York: Academic Press.
- Shepard, R.N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27, 125-140.
- Shepard, R.N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27, 219-246.
- Shin, H.J., & Nosofsky, R.M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, 121, 278-304.
- Spence, I., Domoney, D.W. (1974). Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika*, 39, 469-490.
- Stevens, S.S. (1971). Issues in psychophysical measurement. *Psychological Review*, 78, 426-450
- Takane, Y., Young, F. W., DeLeeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7-67.
- Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17, 401-419.
- Young, F.W., Hamer, R.B. (1987). Multidimensional scaling: history, theory and applications. Hillsdale, NJ: Laurence Erlbaum Associates.
- Young, F.W., Harris, D.F. (2004). Multidimensional scaling. In Norusis, M.J. (Ed.), *SPSS 13.0 Advanced Statistical Procedures Companion*, pp.287-354. Upper Saddle River, NJ: Prentice-Hall.

¹ This classification is weakly linked to the actual shape of the matrix: a matrix with an equal number of rows and columns is still considered rectangular if its rows and columns represent different objects.

² In the following section, the notation and model definitions proposed in Young & Harris (2004) are used.

³ For a more detailed discussion of available options, please refer to the SPSS Syntax Guide.

⁴ The formula for STRESS-1 can be obtained by replacing each squared disparity and distance in the S-STRESS1 formula by their positive square root.

⁵ These last two measures should be interpreted with care, because they are probably sub-optimal; the program fits the model using S-STRESS.

⁶ The name comes from the resemblance between the plot and the side of a mountain; *scree* refers to the debris fallen from a mountain and lying at its base.

Submitted 18 April 2006

Accepted 27 April 2006