Multidimensional Scaling

Leland Wilkinson

Multidimensional Scaling (MDS) offers nonmetric multidimensional scaling of a similarity or dissimilarity matrix in one to five dimensions. Multidimensional scaling is a powerful data reduction procedure that can be used on a direct similarity or dissimilarity matrix or on one derived from rectangular data with Correlations. SYSTAT provides three MDS loss functions (Kruskal, Guttman, and Young) that produce results comparable to those from three of the major MDS packages (KYST, SSA, and ALSCAL). All three methods perform a similar function: to compute coordinates for a set of points in a space such that the distances between pairs of these points fit as closely as possible to measured dissimilarities between a corresponding set of objects.

The family of procedures called *principal components* or *factor analysis* is related to multidimensional scaling in function, but multidimensional scaling differs from this family in important respects. Usually, but not necessarily, multidimensional scaling can fit an appropriate model in fewer dimensions than can these other procedures. Furthermore, if it is implausible to assume a linear relationship between distances and dissimilarities, multidimensional scaling nevertheless provides a simple dimensional model.

MDS also computes the INDSCAL (individual differences multidimensional scaling) model (Carroll and Chang, 1970). The INDSCAL model fits dissimilarity or similarity matrices for multiple subjects into one common space, with jointly estimated weight parameters for each subject (that is, a dissimilarity matrix is input for each subject and separate (monotonic) regression functions are computed). MDS can fit the INDSCAL model using any of the three loss functions, although we recommend using Kruskal's STRESS for this purpose.

Finally, MDS can fit the nonmetric unfolding model. This allows one to analyze rank-order preference data.

Statistical Background

Multidimensional scaling (MDS) is a procedure for fitting a set of points in a space such that the distances between points correspond as closely as possible to a given set of dissimilarities between a set of objects. Dissimilarities may be measured directly, as in psychological judgments, or derived indirectly as in correlation matrices computed on rectangular data.

Assumptions

Because MDS, like cluster analysis, operates directly on dissimilarities, no statistical distribution assumptions are necessary. There are, however, other important assumptions. First, multidimensional scaling is a spatial model. To fit points in the kind of spaces that MDS covers, assume that your data satisfy *metric* conditions:

- The distance from an object to itself is 0,

- The distance from object A to object B is the same as that from B to A,

- The distance from object A to C is less than or equal to the distance from A to B plus B to C. This is sometimes called the **triangle inequality**.

You may think these conditions are obvious, but there are numerous counter-examples in psychological perception and elsewhere. For example, commuters often view the distance from home to the city as closer than the distance from the city to home because of traffic patterns, terrain, and psychological expectations related to time of day. Framing or context effects can also disrupt the metric axioms, as Amos Tversky has shown. For example, Miami is similar to Havana. Havana is similar to Moscow. Is Miami similar to Moscow? If your data (objects) are not consistent with these three axioms, do not use MDS.

Second, there are ways of deriving distances from rectangular data that do not satisfy the metric axioms. The ones available in Correlations do, but if you are thinking of using some other derived measure of similarity, check it carefully.

Finally, it is assumed that all your objects will fit in the same metric space. It is best if they diffuse somewhat evenly through this space as well. Do not expect to get

interpretable results for 25 nearly indistinguishable objects and one that is radically different.

Collecting Dissimilarity Data

You can collect dissimilarities directly or compute them indirectly.

Direct Methods

Examples of direct dissimilarities are:

Distances. Take distances between objects (for example, cities) directly off a map. If the scale is local, MDS will reproduce the map nicely. If the scale is global, you will need three dimensions for an MDS fit. Two or three dimensional spatial distances can be measured directly. Direct measures of social distance might include spatial propinquity or the number of times or amount of time one individual interacts with another.

Judgments. Ask subjects to give a numerical rating of the dissimilarity (for example, 0 to 10) between all pairs of objects.

Clusters. Ask people to sort objects into piles; or examine naturally occurring aggregates, such as paragraphs, communities, and associations. Record 0 if two objects occur in the same group and 1 if they do not. Sum these counts over replications or judges.

Triads. Ask subjects to compare three objects at a time and report which two are most similar (or which is the odd one out). Do this over all possible triads of objects. To compute dissimilarities, sum over all triads, as for the clustering method. There are usually many more triads than pairs of objects, so this method is more tedious. However, it allows you to independently assess possible violations of the triangle inequality.

Indirect Methods

Indirect dissimilarities are computed over a rectangular matrix whose columns are objects and rows are attributes. You can transpose this matrix if you want to scale rows instead. Possible indirect dissimilarities include:

Computed Euclidean distances. These are the square root of the sum-of-squared discrepancies between columns of the rectangular matrix.

Negatives of correlations. For standardized data (mean of 0 and standard deviation of 1), Pearson correlations are proportional to Euclidean distances. For unstandardized data, Pearson correlations are comparable to computing Euclidean distances after standardizing. MDS automatically negates correlations if you do not. Other types of correlations for example, Spearman and gamma are analogous to standardized distances, but only approximately. Also, be aware that large negative correlations will be treated as large distances and large positive correlations as small distances. Make sure that all variables are scored in the same direction before computing correlations. If you find that a whole row of a correlation matrix is negative, reverse the variable by multiplying by –1, and recompute the correlations.

Counts of discrepancies. Counting discrepancies between columns or using some of the binary association measures in Correlations is closely related to computing the Euclidean distance. These methods are also related to the clustering distance calculations mentioned above for direct distances.

Scaling Dissimilarities

Once you have dissimilarities (or similarities, correlations, etc., which MDS automatically transforms to dissimilarities), you may scale them. You do not need to know how the computer does the calculations in order to use the program intelligently as long as you pay attention to the following:

Stress and Iterations

Stress is the goodness-of-fit statistic that MDS tries to minimize. It consists of the square root of the normalized squared discrepancies between interpoint distances in the MDS plot and the smoothed distances predicted from the dissimilarities. Stress varies between 0 and 1, with values near 0 indicating better fit. It is printed for each *iteration*, which is one movement of all the points in the plot toward a better solution. Make sure that iterations proceed smoothly to a minimum. This is true for the examples in this chapter. If you find that the stress values increase or decrease in uneven steps, you should be suspicious.

The Shepard Diagram

The Shepard diagram is a scatterplot of the distances between points in the MDS plot against the observed dissimilarities (or similarities). The points in the plot should adhere cleanly to a curve or straight line (which would be the smoothed distances). In other words, you should look at a good Shepard plot and think it resembles the outcome of a well-designed experiment. For more information refer examples in the chapter.

If the Shepard diagram resembles a stepwise or *L*-shaped function, beware, you may have achieved a degenerate solution. Publish it and you will be excoriated by the clergy.

The MDS Plot

The plot of points is what you seek. The points should be scattered fairly evenly through the space. The orientation of axes is arbitrary—remember we are scaling distances, not axes. Feel free to reverse axes or rotate the solution. MDS rotates it to the largest dimensions of variation, but these do not necessarily mean anything for your data.

You may interpret the axes as in principal components or factor analysis. More often, however, you should look for clusters of objects or regular patterns among the objects, such as circles, curved manifolds, and other structures. See the Guttman loss function example for a good view of a circle.

For more information, see Borg and Lingoes (1981, 1987), Carroll and Arabie (1980), Davison (1983), Green and Rao (1972), Kruskal, Wish and Uslaner (2006), Schiffman, Reynolds, and Young (1981), and Shepard, Romney and Nerlove (1972).

Multidimensional Scaling in SYSTAT

Multidimensional Scaling Dialog Box

To open the Multidimensional Scaling dialog box, from the menus choose:

Advanced
  Multidimensional Scaling...
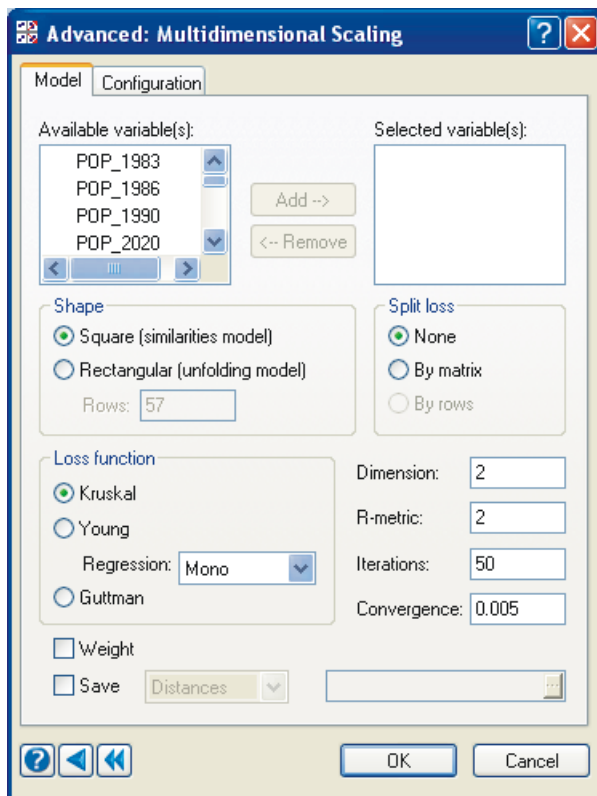
The following options are available:

Selected variable(s). Select the variables that contain the matrix of data to be analyzed.

Shape. Specify the type of matrix input. For a similarities model, select Square. For an unfolding model, select Rectangular and enter the number of rows in your matrix.

Loss function. MDS scales similarity and dissimilarity matrices using three loss functions:

- **Kruskal.** Uses Kruskal's STRESS formula 1 scaling method.
- **Young.** Uses Young's S-STRESS scaling method, which allows you to scale using the loss function featured in ALSCAL.
- **Guttman.** Uses Guttman's coefficient of alienation scaling method.

Note: Iterations with Kruskal's method are faster but usually take longer to converge to a minimum value than those with the Guttman method. The procedure used in the

latter has been found in simulations to be less susceptible to local minima than that used in the Kruskal method (Lingoes and Roskam, 1973). We do not recommend Young's S-STRESS loss function. Because it weights squares of distances, large distances have more influence than smaller ones. Weinberg and Menil (1993) summarized why this is a problem: "…error variances of dissimilarities tend to be positively correlated with their means. If this is the case, large distances should be, if anything, *down*-weighted relative to small distances."

When using the Kruskal or Young loss functions, choose the form of the function relating distances to similarities (or dissimilarities):

- **Mono.** Specifies nonmetric scaling.
- **Linear.** Specifies metric scaling.
- **Log.** Specifies a log function, allowing a smooth curvilinear relation between dissimilarities and distances.
- **Power.** Specifies a power function. (This option is available only with Kruskal loss function.)

By default, SYSTAT takes it as Kruskal MONOTONIC loss function.

Note: If you use the Kruskal loss function, you can fit a MONOTONIC, LINEAR, or LOG function of distances onto input dissimilarities. The standard option is MONOTONIC multidimensional scaling. To avoid degenerate solutions, however, log or linear scaling is sometimes handy. Log scaling is recommended for this purpose because it allows a smooth curvilinear relation between dissimilarities and distances.

Split loss. For an individual differences of unfolding model, split the calculation of the loss function by rows of the matrix or by matrices. Splitting by rows is possible only for a rectangular matrix.

Dimension. Number of dimensions in which to scale. The number of dimensions must be a positive integer less than or equal to the number of variables that you scale and 5. The default value is 2.

R-metric. Constant for the Minkowski power metric for computing distances. For ordinary Euclidean distance, enter 2. For city-block distance, enter 1. For values other than 1 or 2, computation is slower because logarithms and exponentials are used. The default value is 2.

The general formula for calculating distances is:

$$d_{jk} = \left[\sum_{i=1}^{p} |x_{ij} - x_{ik}|^r \right]^{\frac{1}{r}}$$

where *r* is the specified power and *p* is the number of dimensions.

Iterations. Limit for the number of iterations.

Convergence. Iterations terminate when the maximum absolute difference between any coordinate in the solution at iteration *i* versus iteration $i - 1$ is less than the specified convergence criterion. Because the configuration is standardized to unit variance on every iteration, iteration stops when no coordinate moves more than the specified convergence criterion (0.005 by default) from its value on the previous iteration.

Most MDS programs terminate when stress reaches a predetermined value or changes by less than a small amount. These programs can terminate prematurely, however, because comparable stress values can result from different configurations. The SYSTAT convergence criterion allows you to stop iterating when the configuration ceases to change.

Weight. Adds weights for each dimension and each matrix (subject) into the calculation of separate distances that are used in the minimization. For an individual differences model, select Weight.

**Save.** You can save three sets of output to a data file:

- **Configuration.** Saves the final configuration.
- **Distances.** Saves the matrix of distances between points in the final scaled configuration.
- **Residuals.** Saves the data, distances, estimated distances, residuals, and the row and column number of the original distance in the rectangular SYSTAT file.

With the residuals, MDS displays the root-mean-squared residuals for each point in its output. Because STRESS is a function of the sum-of-squared residuals, the root-mean-squared residuals are a measure of the influence of each point on the STRESS statistic. This can help you identify ill-fitting points.

Multidimensional Scaling Configuration

SYSTAT offers several alternative initial configurations.
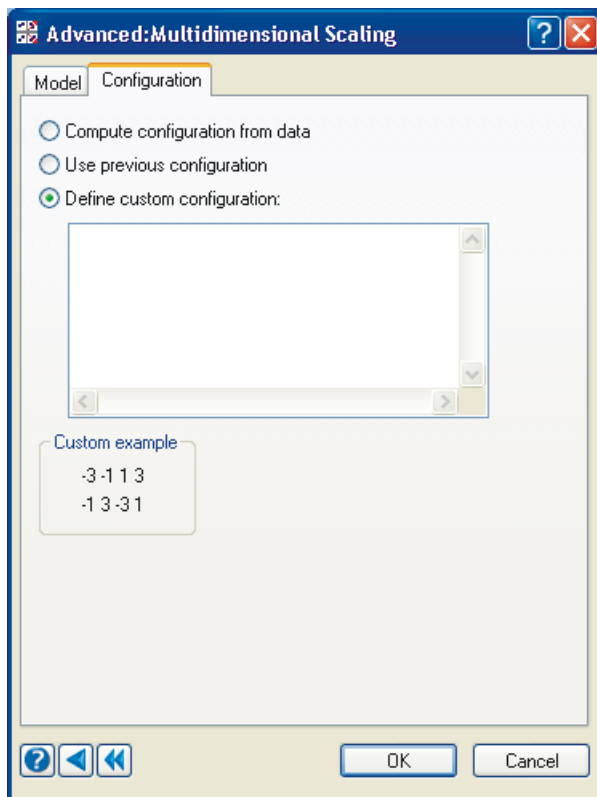
Compute configuration from data. By default, the configuration is computed from the data. The method used depends on the loss function.

Use previous configuration. Uses the configuration from the previous scaling.

Define custom configuration. You can specify a custom starting configuration for the scaling. There must be as many rows as items and columns as dimensions. When you type a matrix, SYSTAT reads as many numbers in each row as you specify. It reads as many rows as there are points to scale.

You can specify a configuration for confirmatory analysis. Enter a hypothesized configuration and let the program iterate only once. Then look at the stress.

Using Commands

First, specify your data with USE *filename*. Continue with:

```
MDS
     MODEL varlist / ROWS=n SHAPE=SQUARE or RECT
     CONFIG LAST
          or
     CONFIG [matrix]
     ESTIMATE / DIM=n  R=n  ITER=n  WEIGHT  CONVERGE=n ,
                 LOSS=GUTTMAN or KRUSKAL or YOUNG ,
                 REGRESS=MONO or LINEAR or LOG or POWER ,
                 SPLIT=ROW or MATRIX
     SAVE filename / CONFIG or DIST or RESID
```

Usage Considerations

Types of data. MDS uses a data file that contains an SSCP, covariance, correlation, or dissimilarity matrix. When you open the data file, MDS automatically recognizes its type.

Print options. The output is standard for all PLENGTH options.

Quick Graphs. MDS produces a Shepard diagram for each matrix analyzed and a plot of the final configuration. For solutions containing four or more dimensions, the final configuration appears as a scatterplot matrix of all dimension pairs.

Saving files. You can save the final configuration, matrix of distances between points in the final scaled configuration, distances, estimated distances, residuals, and the row and column number of the original distance in SYSTAT data files.

BY groups. MDS produces a separate analysis for each level of a BY variable.

Case frequencies. FREQ is not available in MDS.

Case weights. WEIGHT is not available in MDS.

Examples

Example 1
Kruskal Method

The data in the *ROTHKOPF* file are adapted from an experiment by Rothkopf (1957). They were originally obtained from 598 subjects who judged whether or not pairs of Morse code signals presented in succession were the same. Morse code signals for letters and digits were used in the experiment, and all pairs were tested in each of two possible sequences. For multidimensional scaling, the data for letter signals have been averaged across sequence, and the diagonal (pairs of the same signal) has been omitted. The data in this form were first scaled by Shepard.

The input is:

```
MDS
    USE ROTHKPF1
    MODEL a .. z
    IDVAR code$
    ESTIMATE / LOSS=KRUSKAL
```

Use the shortcut notation (..) in MODEL for listing consecutive variables in the file (otherwise, simply list each variable name separated by a space).

The program begins by generating an initial configuration of points whose interpoint distances are a linear function of the input data. For this estimation, MDS uses a metric multidimensional scaling. To do this, missing values in the input matrix are replaced by mean values for the whole matrix. Then the values are converted to distances by adding a constant.

The output is:

```
Monotonic Multidimensional Scaling

Kruskal Method
The data are analyzed as similarities
Minimizing Kruskal STRESS (form 1) in 2 dimensions

Iteration History

  Iteration     STRESS
  --------------------
        0    0.263539
        1    0.237909
        2    0.218820
        3    0.202184
        4    0.190513
        5    0.184341
        6    0.181174
        7    0.179394
        8    0.178269
```

```
Stress of Final Configuration     : 0.178269
Proportion of Variance (RSQ)      : 0.845020


Coordinates in 2 Dimensions

  Variable          Dimension
                      1          2
  .-             -1.211291   -0.310037
  -...            0.587818   -0.449746
  -.-.            0.667949    0.050103
  -..             0.061532   -0.439883
  .              -1.542846    0.893490
  ..-.            0.475856   -0.571910
  --.             0.224256    0.645882
  ....            0.032423   -1.047075
  ..             -1.447269   -0.381961
  .---            0.776074    0.765947
  -.-             0.224747    0.024567
  .-..            0.603292   -0.269646
  --             -0.621882    0.757884
  -.             -1.153966   -0.042454
  ---             0.468887    1.024640
  .--.            0.629749    0.305905
  --.-            0.897228    0.555671
  .-.            -0.283513   -0.343725
  ...            -0.655589   -1.038669
  -              -1.469059    0.948010
  ..-            -0.310876   -0.750825
  ...-            0.365593   -0.869607
  .--             0.041743    0.131315
  -..-            0.832711   -0.148606
  -.--            0.870719    0.381966
  --..            0.935717    0.178765
```
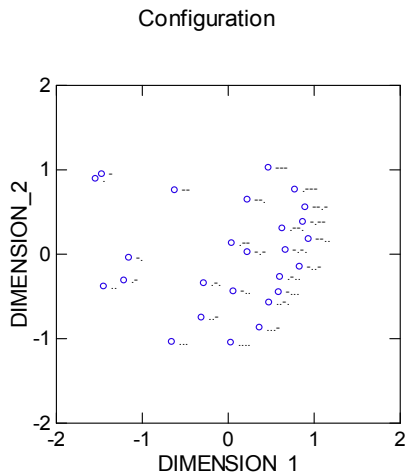
Shepard Diagram

Configuration
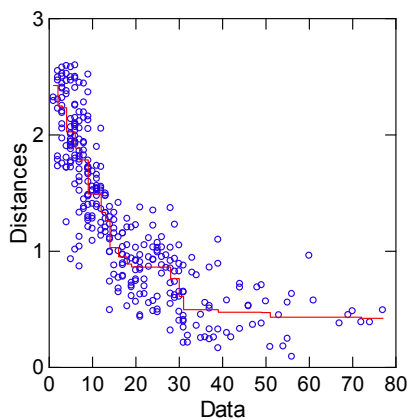
The solution required eight iterations. Notice that STRESS reduces at each iteration. Final STRESS values near zero may indicate the presence of a degenerate solution.

The Shepard diagram is a scatterplot of distances between points in the MDS plot against the observed dissimilarities or similarities. In monotonic scaling, the regression function has steps at various points. For most solutions, the function in this plot should be relatively smooth (without large steps). If the function looks like one or two large steps, you should consider setting REGRESSION to LOG or LINEAR under ESTIMATE.

Notice that large values of the data tend to have small distances in the configuration. The diagram displays an overall decreasing trend because we are using similarities (large data values indicate similar objects). For dissimilarities, the Shepard diagram displays an increasing trend.

In the configuration plot, the points should be scattered fairly evenly through the space. If you are scaling in more than two dimensions, you should examine plots of pairs of axes or rotate the solution in three dimensions. The solution has been rotated to principal axes (that is, the major variation is on the first dimension). This rotation is not performed unless the scaling is in Euclidean space, as in the present example.

The two-dimensional solution clearly distinguishes short signals from long and dots from dashes. Dashes tend to appear in the upper right and dots in the lower left. Long codes tend to appear in the lower right and short in the upper left.

Example 2
Guttman Loss Function

To illustrate the Guttman loss function, this example uses judged similarities among 14 spectral colors (from Ekman, 1954). Nanometer wavelengths (W434, …, W674) are used to name the variables for each color. Blue-violets are in the 400's; reds are in the 600's. The judgments are averaged across 31 subjects; the larger the number for a pair of colors, the more similar the two colors are. The file (*EKMAN*) has no diagonal elements, and its type is SIMILARITY.

The Guttman method is used to scale these judgments in two dimensions to determine whether the data fit a perceptual color wheel. The Kruskal loss function will give you a similar result.

The input is:

```
MDS
    USE EKMAN
    MODEL w434 .. w674
    ESTIMATE / LOSS=GUTTMAN
```

The output is:

```
Monotonic Multidimensional Scaling

Guttman loss function
The data are analyzed as similarities
Minimizing Guttman/Lingoes Coefficient of Alienation in 2 dimensions

Iteration History

  Iteration   Alienation
  ----------------------
         0      0.070826
         1      0.042072
         2      0.037764
         3      0.036151
         4      0.035074

Alienation of Final Configuration  : 0.035074
Proportion of Variance (RSQ)       : 0.996227

Coordinates in 2 Dimensions

  Variable         Dimension
                      1           2
  W434        0.311713   -0.905203
  W445        0.400413   -0.840312
  W465        0.893585   -0.574320
  W472        0.952088   -0.484501
```
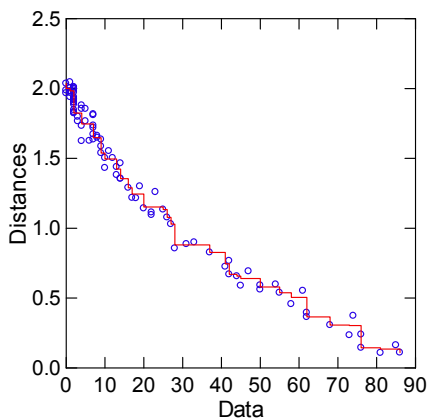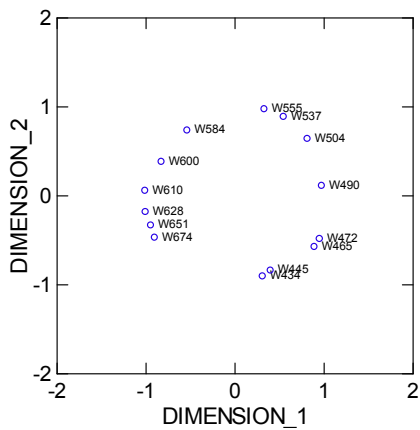
```
W490         0.975491      0.112340
W504         0.814841      0.640540
W537         0.547614      0.888347
W555         0.329882      0.974307
W584        -0.536487      0.734375
W600        -0.826975      0.381875
W610        -1.010004      0.056985
W628        -1.005072     -0.181708
W651        -0.944729     -0.332423
W674        -0.902358     -0.470305
```

Shepard Diagram

Configuration

The fit of configuration distances to original data is extremely close, as evidenced by the low coefficient of alienation and clean Shepard diagram.

The resulting configuration is almost circular, denoting a "circumplex" by Guttman (1954). There is a large gap at the bottom of the figure, however, because the perceptual color between deep red and dark purple is not a spectral color.

Example 3
Individual Differences Multidimensional Scaling

The data in the *COLAS* file are taken from Schiffman, Reynolds, and Young (1981). The data in this file have an unusual structure. The file consists of 10 dissimilarity matrices stacked on top of each other. They are judgments by 10 subjects of the dissimilarity (0–100) between pairs of colas. The example will fit the INDSCAL (individual differences scaling) model to these data, seeking a common group space for the 10 different colas and a parallel weight space for the 10 different judges.

The input is:

```
MDS
    USE COLAS
    MODEL dietpeps .. dietrite
    ESTIMATE / LOSS=KRUSKAL  WEIGHT  SPLIT=MATRIX  DIM=3
```

The WEIGHT option tells SYSTAT to weight each matrix separately. Without this option, all matrices would be weighted equally, and you would have a single pooled solution. You want to use weighting so that you can see which subjects favor one dimension over the others in their judgments. The MATRIX option of SPLIT tells SYSTAT to compute separate (monotonic) regression functions for each subject (matrix). Finally, scale the result in three dimensions, as did Schiffman et al. (1981).

The output is:

```
Monotonic Multidimensional Scaling
Kruskal Method
The data are analyzed as dissimilarities
There are 10 replicated data matrices
Dimensions are weighted separately for each matrix
Fitting is split between data matrices
Minimizing Kruskal STRESS (form 1) in 3 dimensions

Iteration History
```

```
   Iteration    STRESS
   --------------------
          0   0.220898
          1   0.184422
          0   0.221309
          1   0.184508

Stress of Final Configuration   : 0.184508
Proportion of Variance (RSQ)    : 0.535014

Coordinates in 3 Dimensions

  Variable              Dimension
                  1          2          3
  DIETPEPS   -0.608199   0.195575   0.777055
  RC          0.521748   0.052353   0.756390
  YUKON       0.415860  -0.089042  -0.867859
  PEPPER      0.271872  -1.265870   0.059119
  SHASTA      0.797845   0.024902  -0.143788
  COKE        0.390732   0.836586  -0.347338
  DIETPEPR   -0.747107  -0.842914  -0.173399
  TAB        -0.790969   0.438430  -0.609165
  PEPSI       0.570666   0.221001   0.381030
  DIETRITE   -0.822448   0.428980   0.167955

  Matrix ¦   Stress      RSQ           Dimension
         ¦                          1         2         3
  -------+------------------------------------------------
       1 ¦ 0.188374   0.547755   0.697761   0.433686   0.526788
       2 ¦ 0.199808   0.416268   0.452105   0.465357   0.721233
       3 ¦ 0.196430   0.467821   0.347855   0.522989   0.739323
       4 ¦ 0.170677   0.564314   0.591235   0.492475   0.608234
       5 ¦ 0.178156   0.594393   0.704482   0.370109   0.563840
       6 ¦ 0.171913   0.621371   0.704169   0.367609   0.570188
       7 ¦ 0.181071   0.551692   0.419485   0.582263   0.659122
       8 ¦ 0.180465   0.559729   0.483517   0.597254   0.608641
       9 ¦ 0.163263   0.624525   0.562688   0.495564   0.625805
      10 ¦ 0.211658   0.402270   0.435248   0.609372   0.617438
```
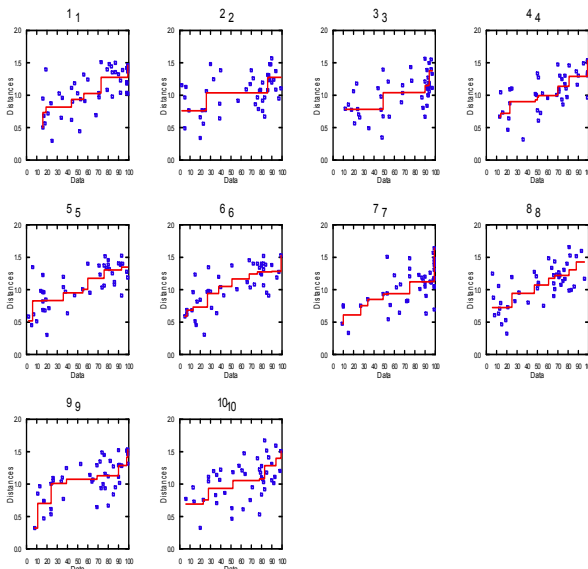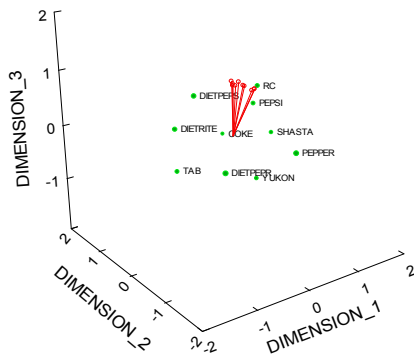
Shepard Diagram

Configuration

The solution required four iterations. Notice that the second two iterations appear to be a restart. This is true, because the fourth matrix has a missing value. SYSTAT uses the EM algorithm to reestimate this value, compute a new metric solution, and iterate two more times until convergence. This extra set of iterations did not do much for you in

this example because the stress is insignificantly higher than it would have been had you stopped at only two iterations. With many missing values, however, the EM algorithm will improve MDS solutions substantially.

For the INDSCAL model, you have a set of coordinates for the colas and one for the subjects. In the three-dimensional graph of the coordinates, the colas are represented by symbols and the subjects by vectors. The first dimension separates the diet colas from the others. The second dimension differentiates between Dr. Pepper/diet Dr. Pepper and the remaining colas.

For each subject, you have a contribution to overall stress and a separate squared correlation (RSQ) between the predicted and obtained distances in the configuration. Notice that subject 10 is fit worst (STRESS = 0.211658) and subject 9 best (STRESS = 0.163263). Furthermore, subjects 1, 5, and 6 have a high loading on the first dimension, indicating that they place a higher emphasis on diet/nondiet differences than on cherry cola/cola differences. Subjects 7, 8, and 10, on the other hand, emphasize the second dimension more.

Example 4
Nonmetric Unfolding

The *COLRPREF* data set contains color preferences among 15 SYSTAT employees for five primary colors. This example uses the MDS unfolding model to scale the people and the colors in two dimensions, such that each person's coordinate is near his or her favorite color's coordinate and far from his or her least favorite color's coordinate. For this example, use ROWS to specify the number of rows for a rectangular matrix and SHAPE to specify the type of matrix input to use. When you enter these data for the first time, you must remember to specify their type as DISSIMILARITY so that small numbers are understood as meaning most similar (preferred).

To scale these with the unfolding model, specify:

```
MDS
    USE COLRPREF
    MODEL red .. blue / SHAPE=RECT
    IDVAR name$
    ESTIMATE / SPLIT=ROWS
```

Notice that you are using the Kruskal loss function as the default.

The output is:

```
Monotonic Multidimensional Scaling
Kruskal Method
The data are analyzed as dissimilarities
The data are rectangular (lower corner matrix)
Fitting is split between rows of data matrix
Minimizing Kruskal STRESS (form 1) in 2 dimensions

Iteration History

   Iteration     STRESS
   ---------------------
          0    0.148373
          1    0.135423
          2    0.125152
          3    0.117255
          4    0.111131
          5    0.106394
          6    0.102622
          7    0.099539
          8    0.096883
          9    0.094498
          0    0.107455
          1    0.100496
          2    0.096037
          3    0.092747
          4    0.090087

Stress of Final Configuration   : 0.090087
Proportion of Variance (RSQ)    : 0.940008

Coordinates in 2 Dimensions

   Variable          Dimension
                       1          2
   RED            0.252839   -0.486827
   ORANGE         0.530030   -1.697840
   YELLOW        -1.312679   -0.563914
   GREEN          1.388778    0.255362
   BLUE          -0.548163    0.785062
   Patrick        0.560619    0.782517
   Laszlo        -0.728868   -0.132010
   Mary          -1.005806    0.113803
   Jenna          0.194159   -0.247226
   Julie         -0.702923   -0.219116
   Steve          1.176419   -0.756052
   Phil           0.612587    0.614672
   Mike          -0.802781   -0.017760
   Keith          0.273582    0.758853
   Kathy          0.048997    0.756548
   Leah          -0.718963    0.004649
   Stephanie      0.498464    0.577649
   Lisa           0.784008    0.209336
   Mark          -0.565003    0.500289
   John           0.064703   -1.237996

Row Fit Measures

   Row      ¦   Stress      RSQ
   ----------+--------------------
   Patrick  ¦ 0.000000   1.000000
   Laszlo   ¦ 0.068318   0.969913
   Mary     ¦ 0.004396   0.999893
   Jenna    ¦ 0.048405   0.983337
   Julie    ¦ 0.271710   0.508263
```
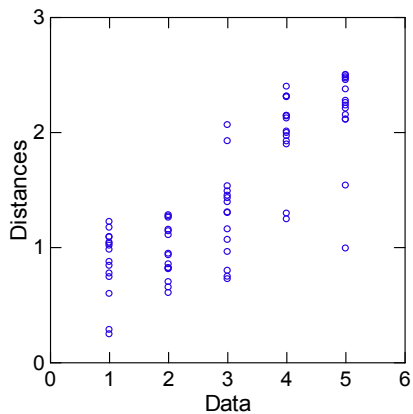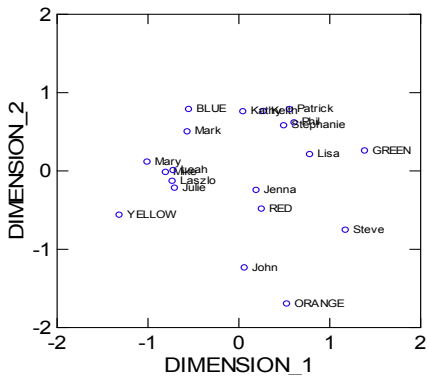
```
Steve      ¦ 0.033042   0.992776
Phil       ¦ 0.061234   0.972002
Mike       ¦ 0.083004   0.958462
Keith      ¦ 0.171898   0.773657
Kathy      ¦ 0.000000   1.000000
Leah       ¦ 0.067386   0.971396
Stephanie  ¦ 0.028564   0.993661
Lisa       ¦ 0.055084   0.980702
Mark       ¦ 0.000000   1.000000
John       ¦ 0.024703   0.996053
```

Shepard Diagram

Configuration

Nonmetric Unfolding and the EM Algorithm

The nonmetric unfolding model has often presented problems to MDS programs because so much data are missing. If you think of the unfolding matrix as the lower corner matrix in a larger triangular matrix of subjects + objects, you can visualize how much data (namely, all of the subject-object comparisons) are missing. Since SYSTAT uses the EM algorithm for missing values, unfolding models do not degenerate as frequently. SYSTAT does a complete MDS using all available data and then estimates missing dissimilarities/similarities using the distances in the solution. These estimated values are then used to get a starting configuration for another complete iteration cycle. This process continues until there are no changes between EM cycles.

The following example, from Borg and Lingoes (1987) adapted from Green and Carmone (1970), shows how this works. This unfolding data set contains dissimilarities only between the points delineating *A* and *M*, and these dissimilarities are treated only as rank orders. Borg and Lingoes discuss the problems in fitting an unfolding model to these data.

The input is:

```
MDS
    USE AM
    IDVAR row$
    MODEL / SHAPE=RECT
    ESTIMATE / LOSS=GUTTMAN  SPLIT=ROWS
```

Notice that the example uses the Guttman loss function, but the others provide similar results.

The output is:

```
Monotonic Multidimensional Scaling
Guttman loss function
The data are analyzed as dissimilarities
The data are rectangular (lower corner matrix)
Fitting is split between rows of data matrix
Minimizing Guttman/Lingoes Coefficient of Alienation in 2 dimensions

Iteration History

   Iteration   Alienation
   ----------------------
          0     0.076137
          1     0.037826
          2     0.023535
          3     0.017735
          4     0.013271
          5     0.009960
```

```
Alienation of Final Configuration   : 0.009960
Proportion of Variance (RSQ)        : 0.999247
Coordinates in 2 Dimensions

  Variable        Dimension
                    1           2
  A1          -0.938673   -1.018145
  A2          -0.892414   -0.975977
  A3          -1.090552   -0.414280
  A4          -1.066410   -0.398294
  A5          -1.187946    0.146240
  A6          -1.227090    0.337007
  A7          -1.543054    0.668773
  A8          -0.997198    0.552347
  A9          -0.694101    0.467134
  A10         -0.305124    0.356277
  A11          0.014600    0.102324
  A12          0.104769    0.102859
  A13          0.130734    0.092203
  A14         -0.845901    0.094247
  A15         -0.739913    0.136811
  A16         -0.569064    0.128649
  M1           0.735047   -1.080081
  M2           0.430679   -0.524410
  M3           0.201071   -0.564505
  M4           0.013212   -0.431126
  M5          -0.154900   -0.326271
  M6          -0.205833   -0.180667
  M7          -0.172336    0.121768
  M8          -0.056279    0.224731
  M9           0.175900    0.267054
  M10          0.560531    0.243136
  M11          0.588937    0.218047
  M12          0.588937    0.218047
  M13          0.831710    0.871193
  M14          0.890298    0.660027
  M15          1.041142    0.212429
  M16          1.238422    0.156627
  M17          1.498853    0.231883
  M18          1.701128   -0.210182
  M19          1.940814   -0.485875

Row Fit Measures

   Row ¦   Stress         RSQ
  -----+-------------------
   M1  ¦ 0.000000     1.000000
   M2  ¦ 0.000000     1.000000
   M3  ¦ 0.000000     1.000000
   M4  ¦ 0.000463     0.999998
   M5  ¦ 0.027442     0.993181
   M6  ¦ 0.022243     0.996393
   M7  ¦ 0.024279     0.997286
   M8  ¦ 0.016153     0.998870
   M9  ¦ 0.000250     1.000000
   M10 ¦ 0.000000     1.000000
   M11 ¦ 0.000000     1.000000
   M12 ¦ 0.000000     1.000000
   M13 ¦ 0.001745     0.999957
   M14 ¦ 0.000000     1.000000
   M15 ¦ 0.000000     1.000000
   M16 ¦ 0.000000     1.000000
   M17 ¦ 0.000000     1.000000
   M18 ¦ 0.000000     1.000000
   M19 ¦ 0.000000     1.000000
```
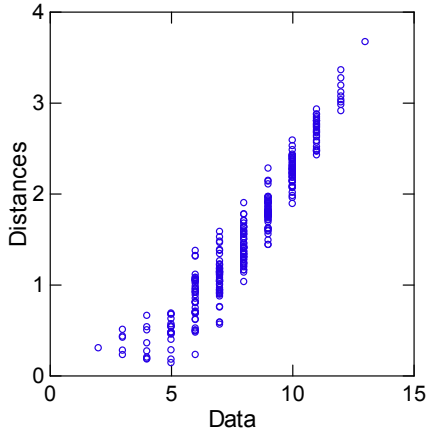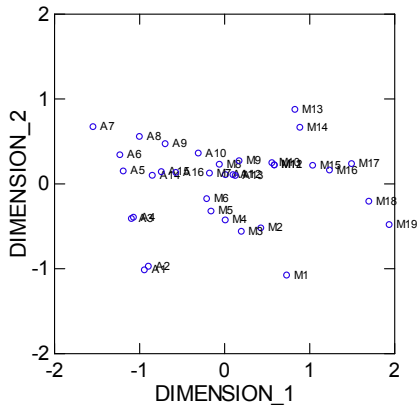
Shepard Diagram

Configuration

Example 5
Power Scaling Ratio Data

As similarities or dissimilarities are often collected as rank-order data, the nonmetric MDS model has to work "backward" in order to solve for a configuration fitting the data. As J. D. Carroll has pointed out, the MDS model should really express observed data as a function of distances between points in a configuration rather than the other

way around. If your data are direct or derived distances, however, you should try setting REGRESSION = POWER with LOSS = KRUSKAL. This way, you can fit a Stevens power function to the data using distances between points in the configuration. The results may not always differ much from nonmetric or linear or log MDS, but SYSTAT will also tell you the exponent of the power function in the Shepard diagram. Notice, with this model, that the data and distances are transposed in the Shepard diagram because loss is being computed from errors in the data rather than the distances. SYSTAT calls the loss for the power model PSTRESS to distinguish it from Kruskal's STRESS. In PSTRESS, you use DATA and its DHAT instead of DIST to compute the loss.

The *HELM* data set contains highly accurate estimates of distance between color pairs by one experimental subject (CB). These are from Helm (1959) and reprinted by Borg and Lingoes (1987).

To scale these with power model, specify:

```
MDS
    USE HELM
    MODEL a .. s
    ESTIMATE / REGRESS=POWER
```

The output is:

```
Power regression function, where Dissimilarities=a*Distances^p
Kruskal Method
The data are analyzed as dissimilarities
Minimizing PSTRESS (STRESS with DIST and DATA exchanged) in 2 dimensions

Iteration History

  Iteration    PSTRESS
 --------------------
        0    0.142062
        1    0.131426
        2    0.127137
        3    0.125205

Stress of Final Configuration              : 0.125205
Estimated Exponent for Power Regression    : 0.851539
Proportion of Variance (RSQ)               : 0.910392

Coordinates in 2 Dimensions

  Variable         Dimension
                   1          2
   A          -0.828615  -0.792411
   C           0.396618  -1.087634
   E           1.134571  -0.503104
   G           0.977829   0.101019
   I           0.785506   0.483283
   K           0.331216   0.683545
   M          -0.205344   0.804234
   O          -0.725019   0.581419
   Q          -0.999584   0.052736
   S          -0.867177  -0.323088
```
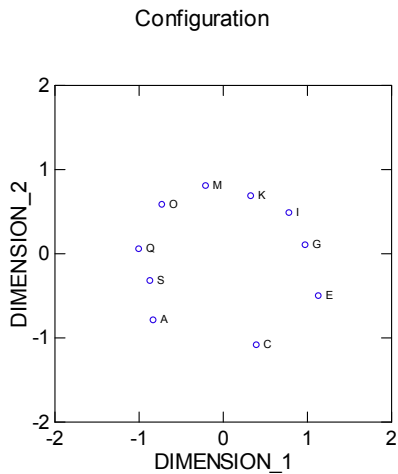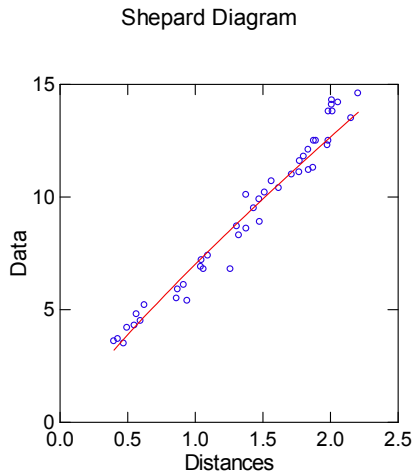
Shepard Diagram

Configuration

SYSTAT estimated the power exponent for the function, fitting distances to dissimilarities as 0.85. Color and many other visual judgments show similar power exponents less than 1.0.

Computation

This section summarizes algorithms separately for the Kruskal and Guttman methods. The algorithms in these options substantially follow those of Kruskal (1964a, 1964b)

and Guttman (1968). MDS output should agree with other nonmetric multidimensional scaling except for rotation, dilation, and translation of the configuration. Secondary documentation can be found in Schiffman, Reynolds, and Young (1981) and the other multidimensional scaling references. The summary assumes that dissimilarities are input. If similarities are input, MDS inverts them.

Algorithms

Kruskal Method

The program begins by generating a configuration of points whose interpoint distances are a linear function of the input data. For this estimation, MDS uses a metric multidimensional scaling. Missing values in the input dissimilarities matrix are replaced by mean values for the whole matrix. Then the values are converted to distances by adding a constant. A scalar products matrix **B** is then calculated following the procedures described in Torgerson (1958). The initial configuration matrix **X** in *p* dimensions is computed from the first *p* eigenvectors of **B** using the Young-Householder procedure (Torgerson, 1958)

After an initial configuration is computed by the metric method, nonmetric optimization begins (there are no metric pre-iterations). At the beginning of each iteration, the configuration is normalized to have zero centroid and unit dispersion. Next, Kruskal's DHAT (fitted) distance values are computed by a monotonic regression of distances onto data. Tied data values are ordered according to their corresponding distances in the configuration.

Stress (formula 1) is calculated from fitted distances, observed distances, and input data values. If the stress is less than 0.001, or has decreased less than 0.001 per iteration in the last five iterations, or the number of iterations equals the number specified by the user (default is 50), iterations terminate (that is, go to the next paragraph). Otherwise, the negative gradient is computed for each point in the configuration by taking the partial derivatives of stress with respect to each dimension. Points in the configuration are moved along their gradients with a step size chosen as a function of the rate of descent; the steeper the descent, the smaller the step size. This completes an iteration.

After the last iteration, the configuration is shifted so that the origin lies in the centroid. Thus, the point coordinates sum to 0 on each dimension. Moreover, the configuration is normalized to unit size so that the sum of squares of its coordinates is 1. If the Minkowski constant is 2 (Euclidean scaling, which is the standard option), the final configuration is rotated to its principal axis.

Guttman Method

The initial configuration for the Guttman option is computed according to Lingoes and Roskam (1973). Principal components are computed on a matrix **C**,

$$c_{ij} = 1 - \frac{r_{ij}}{\dfrac{n(n-1)}{2}}$$

where r_{ij} are the ranks of the input dissimilarities (smallest rank corresponding to smallest dissimilarity), and n is the number of points. The diagonal elements of **C** are

$$c_{jj} = 1 - \Sigma r_{ij}$$

where the sum is taken over the entire row of the dissimilarity matrix.

For the iteration stage, the initial configuration is normalized as in the Kruskal method. Then rank images corresponding to each distance in the configuration are computed by permuting the configuration distances so that they mirror the rank order of the original input dissimilarities. Ties in the data are handled as in the Kruskal method. These rank images are used to compute the Guttman/Lingoes coefficient of alienation. Iterations are terminated if this coefficient becomes arbitrarily small, if the number of iterations exceeds the maximum, or if the change in its value becomes small. Otherwise, the points in the configuration are moved five times using the same rank images but different interpoint distances each time to compute a new negative gradient. These five cycles within each iteration are what lengthens the calculations in the Guttman method. This completes an iteration.

The final configuration is rotated and scaled as with the Kruskal method. Guttman/Lingoes programs normalize the extreme values of the configuration to unity and thus do not plot the configuration with a zero centroid, so MDS output corresponds to their output within rigid motion and configuration size.

Missing Data

Missing values in a similarity/dissimilarity matrix are ignored in the computation of the loss function that determines how points in the configuration are moved. For information on how this function is computed, see the discussion of algorithms.

If you compute a similarity matrix with Correlations for input to MDS, the matrix will have no missing values unless all of your cases in the raw data have a constant or missing value on one or more variables.

References

* Borg, I. and Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications.* 2nd ed. New York: Springer-Verlag.

Borg, I. and Lingoes, J. (1981). *Multidimensional data representations: When and why?* Ann Arbor: Mathesis Press.

Borg, I. and Lingoes, J. (1987). *Multidimensional similarity structure analysis*. New York: Springer Verlag.

Carroll, J. D. and Arabie, P. (1980). Multidimensional scaling. M. R. Rosenzweig and L. W. Porter, eds. *Annual Review of Psychology*, 31, 607–649.

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

* Carroll, J. D. and Wish, M. (1974). Models and methods for three-way multidimensional scaling. D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes, eds. *Contemporary Developments in Mathematical Psychology, Vol. II: Measurement, Psychophysics, and Neural Information Processing.* San Francisco: W. H. Freeman and Company.

* Coombs, C. H. (1964). *A theory of data.* New York: John Wiley & Sons.

Davison, M. L. (1983). *Multidimensional scaling.* New York: John Wiley & Sons.

Ekman, G. (1954). Dimensions of color vision. *Journal of Psychology*, 38, 467–474.

Green, P. E. and Carmone, F. J. (1970). *Multidimensional scaling and related techniques*. Boston: Allyn and Bacon.

Green, P. E. and Rao, V. R. (1972). *Applied multidimensional scaling*. New York: Holt, Rinehart, and Winston.

Guttman, L. (1954). A new approach to factor analysis: The radex. P. F. Lazarsfeld, ed. *Mathematical Thinking in the Social Sciences.* New York: Free Press.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469–506.

Helm, C. E. (1959). A multidimensional ratio scaling analysis of color relations. *Technical Report*, Princeton University and Educational Testing Service, June 1959.

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

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115–129.

Kruskal, J. B.,Wish, M and Uslaner, E.M. (2006). *Multidimensional scaling.* Beverly Hills, Calif.: Sage Publications.

Lingoes, J. C. and Roskam, E. E. (1973). A mathematical and empirical study of two multidimensional scaling algorithms. *Psychometrika Monograph Supplement*, 19.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, 53, 94–101.

Schiffman, S. S., Reynolds, M. L., and Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications*. New York: Academic Press.

Shepard, R. N. (1963). Analysis of proximities as a study of information processing in man. *Human Factors*, 5, 33–48.

Shepard, R. N., Romney, A. K., and Nerlove, S., eds. (1972). *Multidimensional scaling: Theory and application in the behavioral sciences.* New York: Academic Press.

* Takane, Y., Young, F. W., and de Leeuw, J. (1977). Nonmetric individual differences scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 42, 3–27.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley & Sons.

Weinberg, S. L. and Menil, V. C. (1993). The recovery of structure in linear and ordinal data: INDSCAL and ALSCAL. *Multivariate Behavioral Research*, 28:2, 215–233.

(* indicates additional references.)