

[www.statsoft.com](http://www.statsoft.com)

- [Products](#)
- [Solutions](#)
- [Buy](#)
- [Trials](#)
- [Support](#)






























TextbookMultidimensional Scaling

What can we help you find?

[Search](#)

## Looking for info about statistics?

We wrote the book on it.  
And you can read it for free!

-  [Elementary Concepts](#)
-  [Statistics Glossary](#)
-  [Basic Statistics](#)
-  [ANOVA / MANOVA](#)
-  [Association Rules](#)
-  [Boosting Trees](#)
-  [Canonical Analysis](#)
-  [CHAID Analysis](#)
-  [C & R Trees](#)
-  [Classification Trees](#)
-  [Cluster Analysis](#)
-  [Correspondence Analysis](#)
-  [Data Mining Techniques](#)
-  [Discriminant Analysis](#)
-  [Distribution Fitting](#)
-  [Experimental Design](#)
-  [Factor Analysis](#)
-  [General Discrim. Analysis](#)
-  [General Linear Models](#)
-  [Generalized Additive Mod.](#)
-  [Generalized Linear Mod.](#)
-  [General Regression Mod.](#)
-  [Graphical Techniques](#)
-  [Ind.Components Analysis](#)
-  [Linear Regression](#)
-  [Log-Linear Analysis](#)
-  [MARSplines](#)
-  [Machine Learning](#)
-  [Multidimensional Scaling](#)
-  [Neural Networks](#)
-  [Nonlinear Estimation](#)
-  [Nonparametric Statistics](#)
-  [Partial Least Squares](#)
-  [Power Analysis](#)
-  [Process Analysis](#)
-  [Quality Control Charts](#)
-  [Reliability / Item Analysis](#)
-  [SEPATH \(Structural eq.\)](#)
-  [Survival Analysis](#)
-  [Text Mining](#)
-  [Time Series / Forecasting](#)

# Multidimensional Scaling

- [General Purpose](#)
- [Logic of MDS](#)
- [Computational Approach](#)
- [How many dimensions to specify?](#)
- [Interpreting the Dimensions](#)
- [Applications](#)
- [MDS and Factor Analysis](#)

## General Purpose









Multidimensional scaling (*MDS*) can be considered to be an alternative to factor analysis (see [Factor Analysis](#)). In general, the goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects. In factor analysis, the similarities between objects (e.g., variables) are expressed in the correlation matrix. With MDS, you can analyze any kind of similarity or dissimilarity matrix, in addition to correlation matrices.

## Logic of MDS

The following simple example may demonstrate the logic of an MDS analysis. Suppose we take a matrix of distances between major US cities from a map. We then analyze this matrix, specifying that we want to reproduce the distances based on two dimensions. As a result of the MDS analysis, we would most likely obtain a two-dimensional representation of the locations of the cities, that is, we would basically obtain a two-dimensional map.

In general then, MDS attempts to arrange "objects" (major cities in this example) in a space with a particular number of dimensions (two-dimensional in this example) so as to reproduce the observed distances. As a result, we can "explain" the distances in terms of underlying dimensions; in our example, we could explain the distances in terms of the two geographical dimensions: north/south and east/west.

**Orientation of axes.** As in factor analysis, the actual orientation of axes in the final solution is arbitrary. To return to our example, we could rotate the map in any way we want, the distances between cities remain the same. Thus, the final orientation of axes in the plane or space is

-  [Variance Components](#)
-  [Statistical Advisor](#)
-  [Distribution Tables](#)
-  [References Cited](#)
-  [Send Comments](#)
-  [Business Solutions](#)
-  [Free Resources](#)
-  [About Textbook](#)

mostly the result of a subjective decision by the researcher, who will choose an orientation that can be most easily explained. To return to our example, we could have chosen an orientation of axes other than north/south and east/west; however, that orientation is most convenient because it "makes the most sense" (i.e., it is easily interpretable).

[To index](#)

## Computational Approach

MDS is not so much an exact procedure as rather a way to "rearrange" objects in an efficient manner, so as to arrive at a configuration that best approximates the observed distances. It actually moves objects around in the space defined by the requested number of dimensions, and checks how well the distances between objects can be reproduced by the new configuration. In more technical terms, it uses a function minimization [algorithm](#) that evaluates different configurations with the goal of maximizing the goodness-of-fit (or minimizing "lack of fit").

**Measures of goodness-of-fit: Stress.** The most common measure that is used to evaluate how well (or poorly) a particular configuration reproduces the observed distance matrix is the stress measure. The raw stress value *Phi* of a configuration is defined by:

$$\text{Phi} = \sum [d_{ij} - f(\delta_{ij})]^2$$

In this formula,  $d_{ij}$  stands for the reproduced distances, given the respective number of dimensions, and  $\delta_{ij}$  (*delta<sub>ij</sub>*) stands for the input data (i.e., observed distances). The expression  $f(\delta_{ij})$  indicates a *nonmetric*, monotone transformation of the observed input data (distances). Thus, it will attempt to reproduce the general rank-ordering of distances between the objects in the analysis.

There are several similar related measures that are commonly used; however, most of them amount to the computation of the sum of squared deviations of observed distances (or some monotone transformation of those distances) from the reproduced distances. Thus, the smaller the stress value, the better is the fit of the reproduced distance matrix to the observed distance matrix.

**Shepard diagram.** You can plot the reproduced distances for a particular number of dimensions against the observed input data (distances). This scatterplot is referred to as a *Shepard* diagram. This plot shows the reproduced distances plotted on the vertical (Y) axis versus the original similarities plotted on the horizontal (X) axis (hence, the generally negative slope). This plot also shows a step-function. This line represents the so-called *D-hat* values, that is, the result of the monotone transformation  $f(\delta_{ij})$  of the input data. If all reproduced distances fall onto the step-line, then the rank-ordering of distances (or similarities) would be perfectly reproduced by the respective solution (dimensional model). Deviations from the step-line indicate lack of fit.

[To index](#)

## How Many Dimensions to Specify?

If you are familiar with factor analysis, you will be quite aware of this issue. If you are not familiar with factor analysis, you may want to read the [Factor Analysis](#) section in the manual; however, this is not necessary in order to understand the following discussion. In general, the more dimensions we use in order to reproduce the distance matrix, the better is the fit of the reproduced matrix to the observed matrix (i.e., the smaller is the stress). In fact, if we use as many dimensions as there are variables, then we can perfectly reproduce the observed distance matrix. Of course, our goal is to *reduce* the observed complexity of nature, that is, to explain the distance matrix in terms of fewer underlying dimensions. To return to the example of distances

between cities, once we have a two-dimensional map it is much easier to visualize the location of and navigate between cities, as compared to relying on the distance matrix only.

**Sources of misfit.** Let's consider for a moment why fewer factors may produce a worse representation of a distance matrix than would more factors. Imagine the three cities *A*, *B*, and *C*, and the three cities *D*, *E*, and *F*; shown below are their distances from each other.

	A	B	C		D	E	F
A	0			D	0		
B	90	0		E	90	0	
C	90	90	0	F	180	90	0

In the first matrix, all cities are exactly 90 miles apart from each other; in the second matrix, cities *D* and *F* are 180 miles apart. Now, can we arrange the three cities (objects) on one dimension (line)? Indeed, we can arrange cities *D*, *E*, and *F* on one dimension:

*D*---90 miles---*E*---90 miles---*F*

*D* is 90 miles away from *E*, and *E* is 90 miles away from *F*; thus, *D* is 90+90=180 miles away from *F*. If you try to do the same thing with cities *A*, *B*, and *C* you will see that there is no way to arrange the three cities on one line so that the distances can be reproduced. However, we can arrange those cities in two dimensions, in the shape of a triangle:



Arranging the three cities in this manner, we can perfectly reproduce the distances between them. Without going into much detail, this small example illustrates how a particular distance matrix implies a particular number of dimensions. Of course, "real" data are never this "clean," and contain a lot of noise, that is, random variability that contributes to the differences between the reproduced and observed matrix.

**Scree test.** A common way to decide how many dimensions to use is to plot the stress value against different numbers of dimensions. This test was first proposed by Cattell (1966) in the context of the number-of-factors problem in factor analysis (see [Factor Analysis](#)); Kruskal and Wish (1978; pp. 53-60) discuss the application of this plot to MDS.

Cattell suggests to find the place where the smooth decrease of stress values (eigenvalues in factor analysis) appears to level off to the right of the plot. To the right of this point, you find, presumably, only "factorial scree" - "scree" is the geological term referring to the debris which collects on the lower part of a rocky slope.

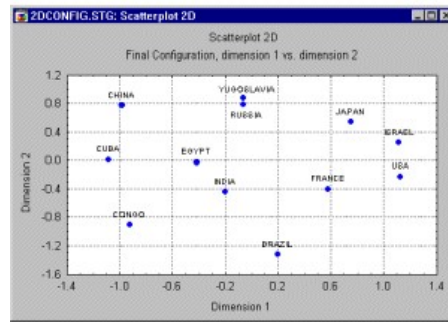
**Interpretability of configuration.** A second criterion for deciding how many dimensions to interpret is the clarity of the final configuration. Sometimes, as in our example of distances between cities, the resultant dimensions are easily interpreted. At other times, the points in the plot form a sort of "random cloud," and there is no straightforward and easy way to interpret the dimensions. In the latter case, you should try to include more or fewer dimensions and examine the resultant final configurations. Often, more interpretable solutions emerge. However, if the data points in the plot do not follow any pattern, and if the stress plot does not show any clear "elbow," then the data are most likely random "noise."

[To index](#)

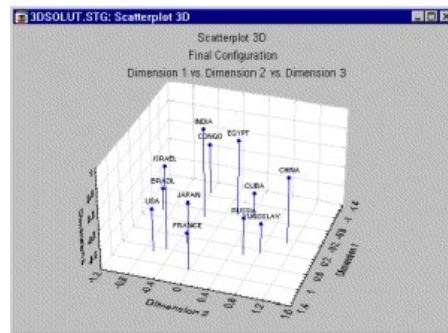
## Interpreting the Dimensions

The interpretation of dimensions usually represents the final step of the analysis. As mentioned earlier, the actual orientations of the axes from the MDS analysis are arbitrary, and can be rotated in any direction. A first step is to produce scatterplots of the objects in the different

two-dimensional planes.



Three-dimensional solutions can also be illustrated graphically, however, their interpretation is somewhat more complex.



In addition to "meaningful dimensions," you should also look for clusters of points or particular patterns and configurations (such as circles, manifolds, etc.). For a detailed discussion of how to interpret final configurations, see Borg and Lingoes (1987), Borg and Shye (in press), or Guttman (1968).

**Use of multiple regression techniques.** An analytical way of interpreting dimensions (described in Kruskal & Wish, 1978) is to use multiple regression techniques to regress some meaningful variables on the coordinates for the different dimensions. Note that this can easily be done via [Multiple Regression](#).

[To index](#)

## Applications

The "beauty" of MDS is that we can analyze any kind of distance or similarity matrix. These similarities can represent people's ratings of similarities between objects, the percent agreement between judges, the number of times a subjects fails to discriminate between stimuli, etc. For example, MDS methods used to be very popular in psychological research on person perception where similarities between trait descriptors were analyzed to uncover the underlying dimensionality of people's perceptions of traits (see, for example Rosenberg, 1977). They are also very popular in marketing research, in order to detect the number and nature of dimensions underlying the perceptions of different brands or products & Carmone, 1970).

In general, MDS methods allow the researcher to ask relatively unobtrusive questions ("how similar is brand A to brand B") and to derive from those questions underlying dimensions without the respondents ever knowing what is the researcher's real interest.

[To index](#)

## MDS and Factor Analysis

Even though there are similarities in the type of research questions to which these two

procedures can be applied, MDS and factor analysis are fundamentally different methods. Factor analysis requires that the underlying data are distributed as multivariate normal, and that the relationships are linear. MDS imposes no such restrictions. As long as the rank-ordering of distances (or similarities) in the matrix is meaningful, MDS can be used. In terms of resultant differences, factor analysis tends to extract more factors (dimensions) than MDS; as a result, MDS often yields more readily, interpretable solutions. Most importantly, however, MDS can be applied to any kind of distances or similarities, while factor analysis requires us to first compute a correlation matrix. MDS can be based on subjects' direct assessment of similarities between stimuli, while factor analysis requires subjects to rate those stimuli on some list of attributes (for which the factor analysis is performed).

In summary, MDS methods are applicable to a wide variety of research designs because distance measures can be obtained in any number of ways (for different examples, refer to the references provided at the beginning of this section).

[To index](#)