

# Multidimensional Scaling (MDS) with XLSTAT

[demoMDS.xls](#)

## Multidimensional Scaling

Multidimensional Scaling (MDS) is a data analysis method which is widely used in marketing and psychometrics.

The aim of the methods is to build a mapping of a series of individuals from a proximities matrix (similarities or dissimilarities) between these individuals. In the ideal case where we have a matrix giving the distances between some points on a surface (for example the cities of a country), the Multidimensional Scaling allows to rebuild the exact map of the points (within about a symmetry and/or rotation).

To build an optimal representation, the Multidimensional Scaling algorithm minimizes a criterion called "Stress". The closer the stress is to zero, the better the representation.

## Dataset for Multidimensional Scaling

An Excel sheet containing both the data and the results for use in this tutorial can be downloaded by clicking [here](#).

The data correspond to a survey performed over 10 testers which have been asked to rate (the score ranges from 1 to 5) five chocolate bars, where only the product P1 is already available on the market.

	P1	P2	P3	P4	P5
Ind1	2	1	4	5	3
Ind2	3	1	2	5	4
Ind3	1	2	4	3	5
Ind4	3	2	4	5	1
Ind5	3	2	4	5	1
Ind6	2	1	3	5	4
Ind7	1	3	4	2	5
Ind8	3	1	2	4	5
Ind9	3	1	2	5	4
Ind10	1	4	5	2	1
<b>Average</b>	<b>2.2</b>	<b>1.8</b>	<b>3.4</b>	<b>4.1</b>	<b>3.3</b>

## Goal of this Multidimensional Scaling analysis

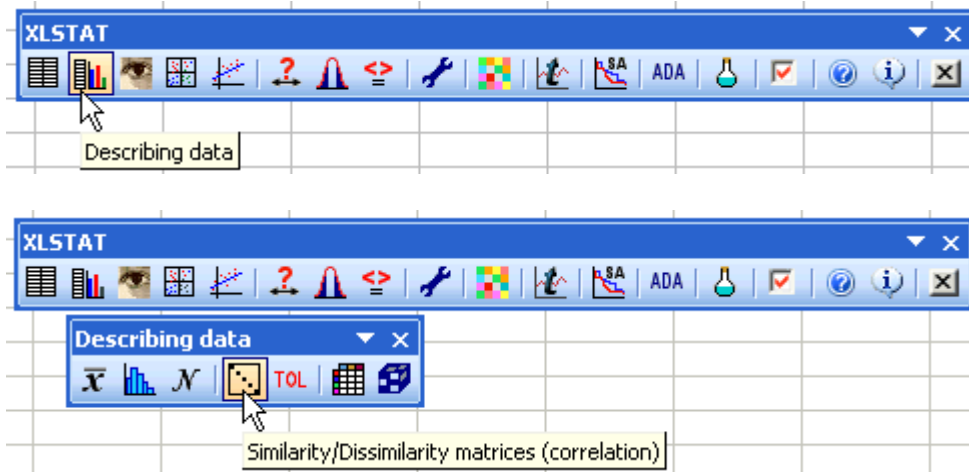
Our aim is to show how the products position themselves on a map, given the opinion of the testers.

## Setting up a Multidimensional Scaling analysis

### Creating a proximity matrix

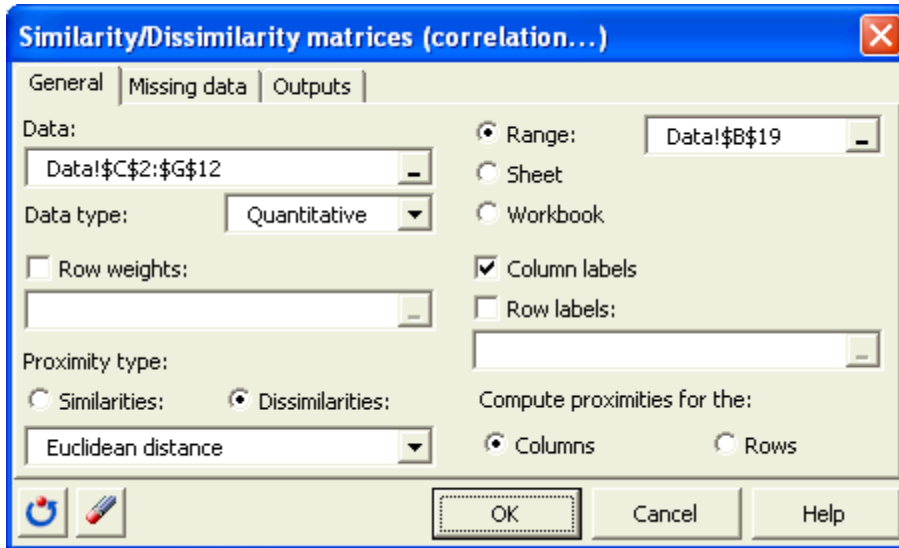
A proximity matrix is needed to perform a Multidimensional Scaling analysis, but here we have a individuals x products table. Therefore we need first to compute the dissimilarities between products, which can be done by using the **Similarity /Dissimilarity matrix** tool of XLSTAT.

Once XLSTAT is activated, select the **XLSTAT / Describing data / Similarity/Dissimilarity matrix** command, or click on the corresponding button of the **Describing data** toolbar (see below).

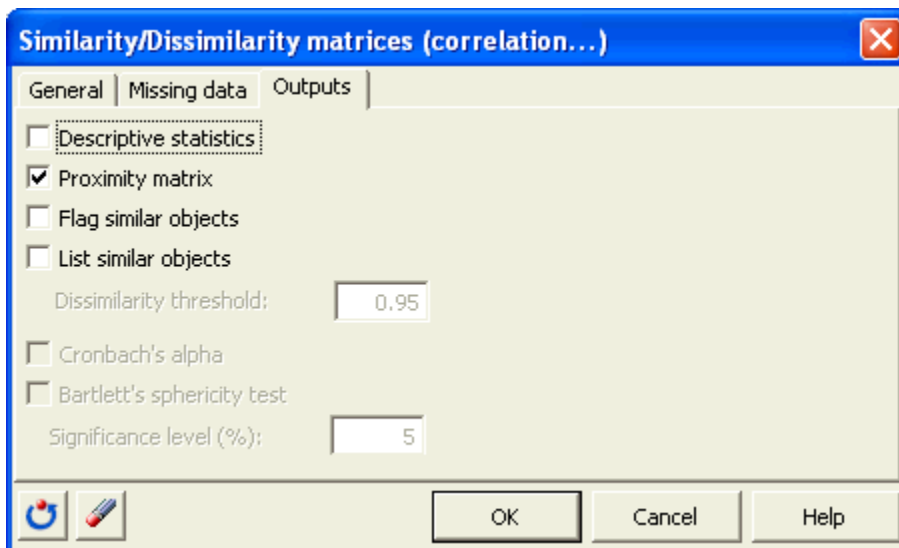


The dialog box appears. You can then select the data on the Excel sheet and choose the appropriate options as shown below.

We decide to display the proximities matrix just below the original data, on the same sheet.



In the **outputs** tab, only the **proximities** are activated.



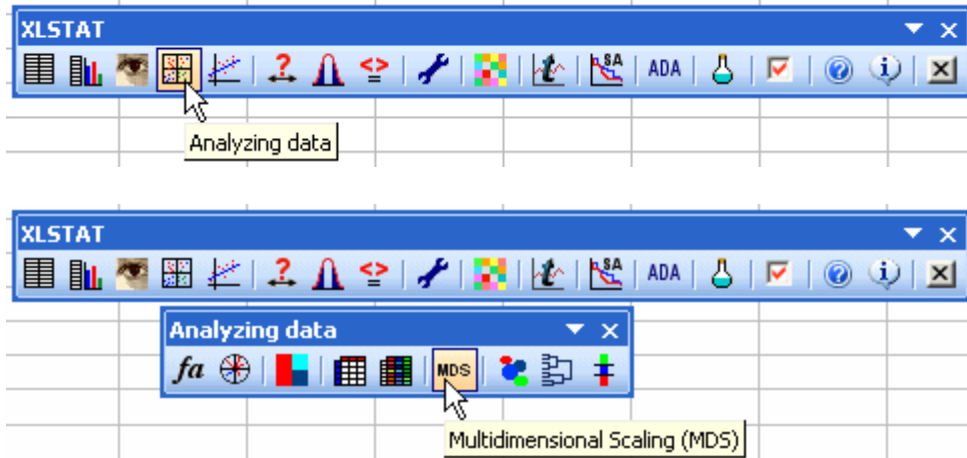
The computations begin once you have clicked on **OK**. You then obtain the matrix of the euclidean distances between the products which will be the base for the Multidimensional Scaling analysis.

Proximity matrix (Euclidean distance):

	P1	P2	P3	P4	P5
P1	0	5,477	6,633	6,403	7,141
P2	5,477	0	5,477	9,849	8,426
P3	6,633	5,477	0	6,557	7,416
P4	6,403	9,849	6,557	0	7,348
P5	7,141	8,426	7,416	7,348	0

## Setting up the Multidimensional Scaling dialog box

Then click on the **XLSTAT / Analyzing data / Multidimensional Scaling command**, or click on the corresponding button of the **Analyzing data** toolbar (see below).



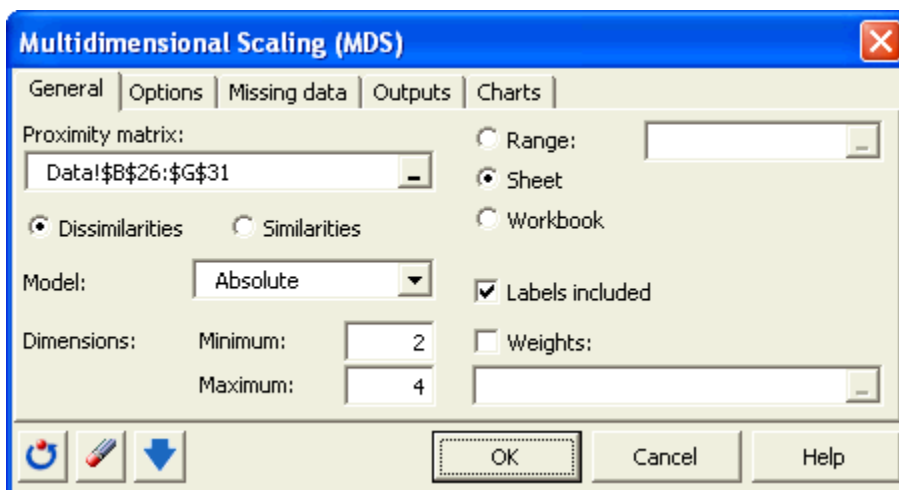
The Multidimensional Scaling dialog box appears.

You can then select the distances matrix on the Excel sheet and choose the appropriate options as shown below.

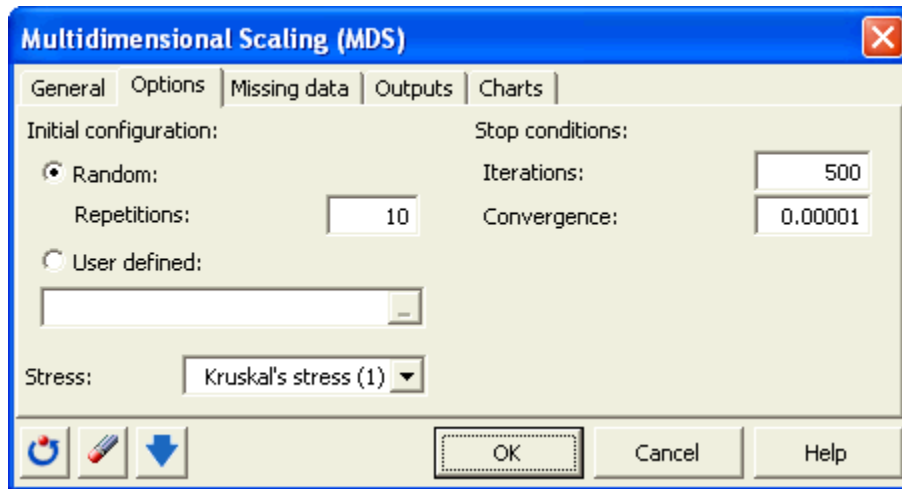
The **absolute** model has been chosen; this model makes that the distances in the final representation are as close as possible to the initial Euclidean distances.

Other options can give similar results, but a scale effect might be introduced, which we want to remove here.

We ask that the analysis is being performed from 4 to 2 dimensions to evaluate the distortion associated to the decrease in dimensions.



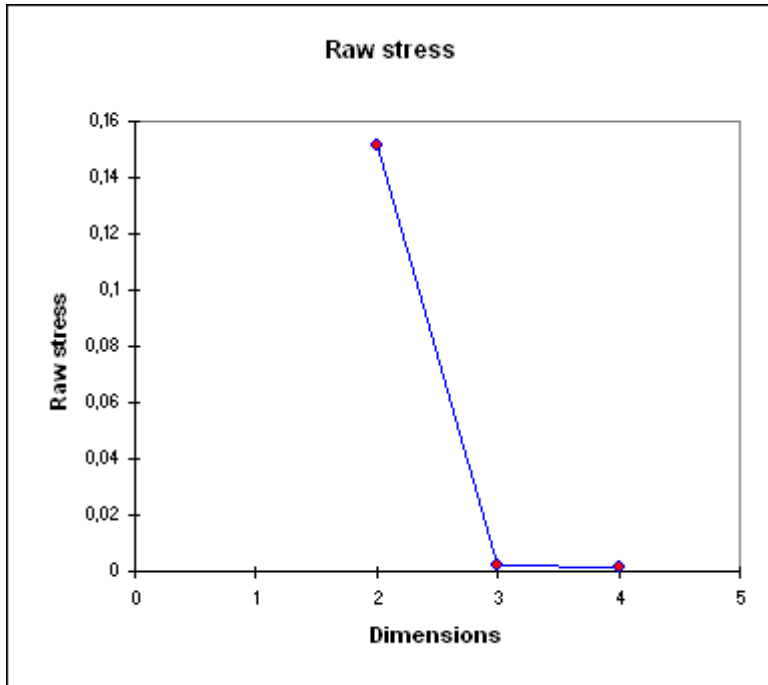
*Note: unless you give the algorithm a starting point in the **Options** tab, the starting points are randomly chosen. Therefore, it is likely that the results you will obtain are slightly different from the one you see on this page. However it shouldn't change the interpretation. To be sure that the algorithm finds a true optimum (in terms of stress) you can increase the number of repetitions, the maximum number of iterations, and the accuracy.*



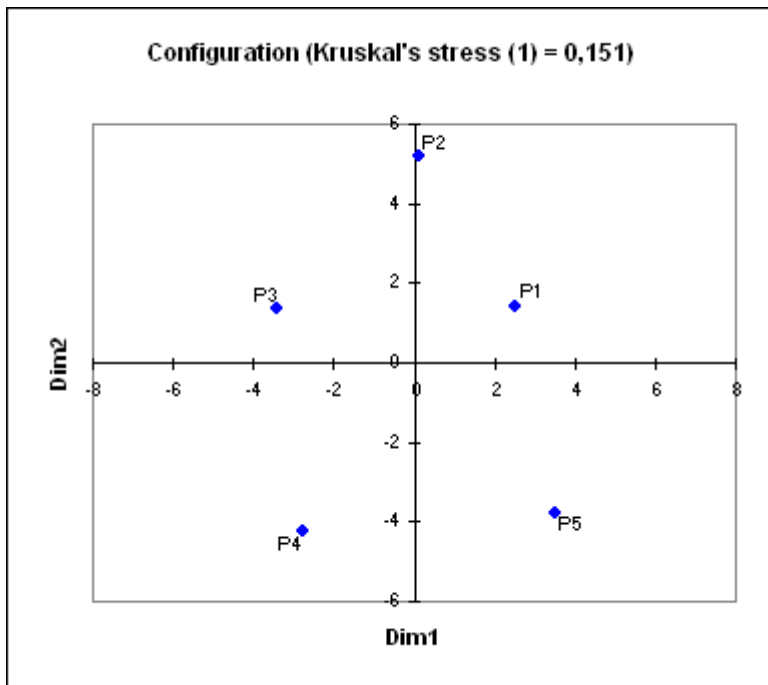
The computations begin once you have clicked on **OK**. After you have accepted to plot the map on the first two dimensions by clicking **Done**, the results are displayed on the Multidimensional Scaling sheet of the Excel workbook.

## Interpreting the results of Multidimensional Scaling


The first table shows the evolution of the stress when the number of dimensions increases. We notice a strong rupture between the 2 dimensional and 3 dimensional representations and a stability between 3 and 4 (it is mathematically normal that a 4D representation for 5 points is perfect).



A map is created on the Dim1 x Dim2 plan, for the 2-dimensional configuration.

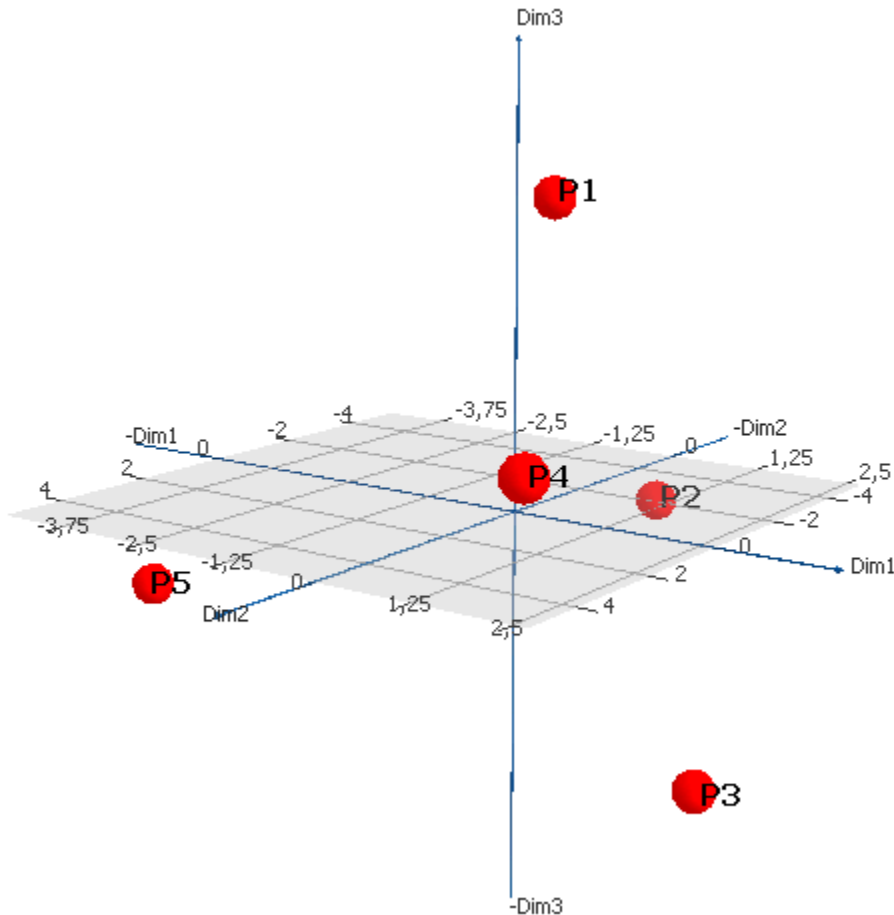


It is also possible to build maps for some other couples of axes for the 3 or 4D configurations. However it is not recommended to view in two dimensions the map for a configuration which has been built for more dimensions as there might be a projection effect that makes false any interpretation. The 2D map should be used only for the 2 dimensions configuration. To view on a 2D map the 4D configuration, one should first do a PCA on the coordinates.

In order to obtain an even better quality of representation, we have used XLSTAT-3DPlot to plot the data in three dimensions. To do that, we just needed to select the data (see below) and to click on the XLSTAT-3DPlot button  in the "Visualizing data" toolbar.

Results for a 3-dimensional representation space:			
Configuration:			
	Dim1	Dim2	Dim3
P1	-0,010	-1,131	3,356
P2	-0,129	-4,968	-0,532
P3	2,249	-0,643	-2,876
P4	2,243	4,455	1,226
P5	-4,352	2,288	-1,174

The result we obtain is as follow:



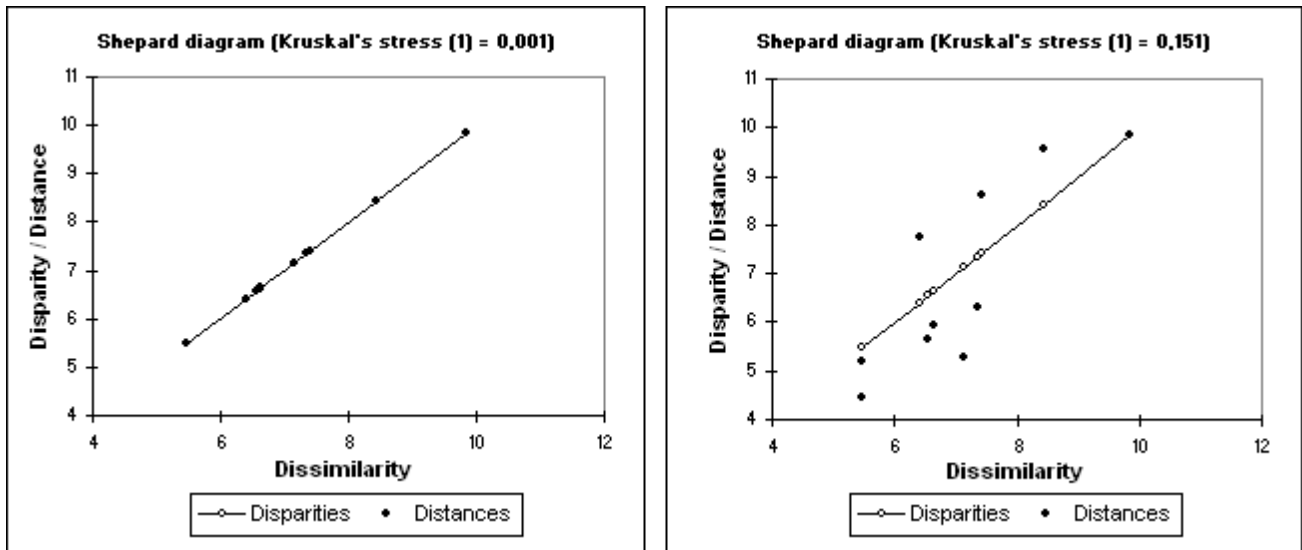
It is possible to see that the testers have collectively well distinguished the products among each other. We know that the product P2 contains more chocolate than P4, which is the one that

contains the less chocolate: on the 3D chart we can see that they are diametrically opposed. By looking at the initial data set, we can see that the testers have significantly preferred the product P2. We can also see that, although they have similar average scores, the products P3 and P5 are not close in the representation space: the testers' opinions are sometimes opposed. That can be explained by the fact that some peanuts have been added to the product P3, which is appreciated by some testers and not by some others.

## Conclusion about Multidimensional Scaling

As a conclusion, the Multidimensional Scaling method allows to map the products that have been rated by the testers. It allows a much richer interpretation than simple statistics would.

*Note: there is no rigorous statistical method to evaluate the quality and the reliability of a representation produced by an Multidimensional Scaling analysis. However by looking at the Shepard diagram, one can have a global idea of the quality of the representation. The Shepard diagram corresponds to a scatter plot, where the abscissas are the observed dissimilarities, and the ordinates, the distance on the configuration generated by the Multidimensional Scaling. The disparities are also displayed. The more the points are spread, the less the Multidimensional Scaling map is reliable. If the ranking of the abscissa is respected on the ordinates, then the chart is reliable. If the points are on the same line, then the quality is perfect.*



The left hand chart corresponds, for the data used in this example to the representation in a 4D space. The right hand chart corresponds to the representation in the 2D space. We notice a strong difference of the spread of the points between the two charts.

*Note: for an absolute model the disparities are equal to the dissimilarities, which is why they are confounded with the line for the 2D space, and hidden behind the distances on the 4D Sheppard diagram.*