

## Difference between 12 nations (Nations.sta)

Overview and Data File. This example is based on the data file Nations.sta. These data are discussed in Kruskal and Wish (1978, page 30). Open this data file via the File - Open Examples menu; it is in the Datasets folder. The data file contains the mean similarity ratings of 18 students for 12 countries. The countries were Brazil, Congo, Cuba, Egypt, France, India, Israel, Japan, Mainland China, Russia, USA, and Yugoslavia. A partial listing of this similarity matrix is shown below.

	1 BRAZIL	2 CONGO	3 CUBA	9 CHINA	10 RUSSIA	11 USA	12 YUGOSLAVIA
<b>BRAZIL</b>	0	4.83	5.28	2.39	3.06	5.39	3.17
CONGO	4.83	0	4.56	4	3.39	2.39	3.5
CUBA	5.28	4.56	0	5.5	5.44	3.17	5.11
EGYPT	3.44	5	5.17	4.39	4.39	3.33	4.28
FRANCE	4.72	4	4.11	3.67	5.06	5.94	4.72
INDIA	4.5	4.83	4	4.11	4.5	4.28	4
ISRAEL	3.83	3.33	3.61	3	4.17	5.94	4.44
JAPAN	3.5	3.39	2.94	4.17	4.61	6.06	4.28
CHINA	2.39	4	5.5	0	5.72	2.56	5.06
RUSSIA	3.06	3.39	5.44	5.72	0	5	6.67
USA	5.39	2.39	3.17	2.56	5	0	3.56
YUGOSLAV	3.17	3.5	5.11	5.06	6.67	3.56	0
Means	0	0	0	0	0	0	0
Std.Dev.	0	0	0	0	0	0	0
No.Cases	18						
Matrix	2						

Note that you can produce a similar matrix file by entering the distances into a new spreadsheet following the matrix data format conventions (as described in the Matrix File Format topic).

Specifying the Analysis. Select Multidimensional Scaling from the Statistics - Multivariate Exploratory Techniques menu to display the Multidimensional Scaling Startup Panel. Click the Variables button to display a standard variable selection dialog. Here, select all of the variables (i.e., objects or nations) for the analysis, and then click the OK button.



STATISTICA assumes that you want to calculate a two-dimensional solution for this similarity matrix, and that the initial solution is to be estimated via principal components analysis. Alternatively, you can also specify the initial configuration on the Options tab by selecting a STATISTICA raw data file with the initial coordinates.

Click the OK button to simply accept the default settings. First, the initial (starting) configuration will be computed, and the Parameter Estimation dialog will be displayed. (Note that you can later view these initial configurations by clicking the Start (initial) configuration button on the Results - Review & Save tab.)

Performing the Analysis. The iterative algorithm for finding an optimum configuration proceeds in two stages: First, the program will use a method known as steepest descent. The respective number of steepest descent iterations is listed under the first column (labeled iter. s) in the Parameter Estimation dialog.

After each iteration under steepest descent, STATISTICA will perform up to 5 additional iterations to "fine-tune" the configuration (see Technical Notes for details). The respective numbers of these iterations are listed under the second column in the Parameter Estimation dialog (labeled iter. t). In addition, the stress value (Kruskal, 1964) and coefficient of alienation (Guttman, 1968) are calculated and displayed at each step (see also the Introductory Overviews and Technical Notes). A detailed discussion of this iterative procedure can be found in Shiffman, Reynolds, and Young (1981, pages 366-370).

Parameter Estimation: Nations.sta

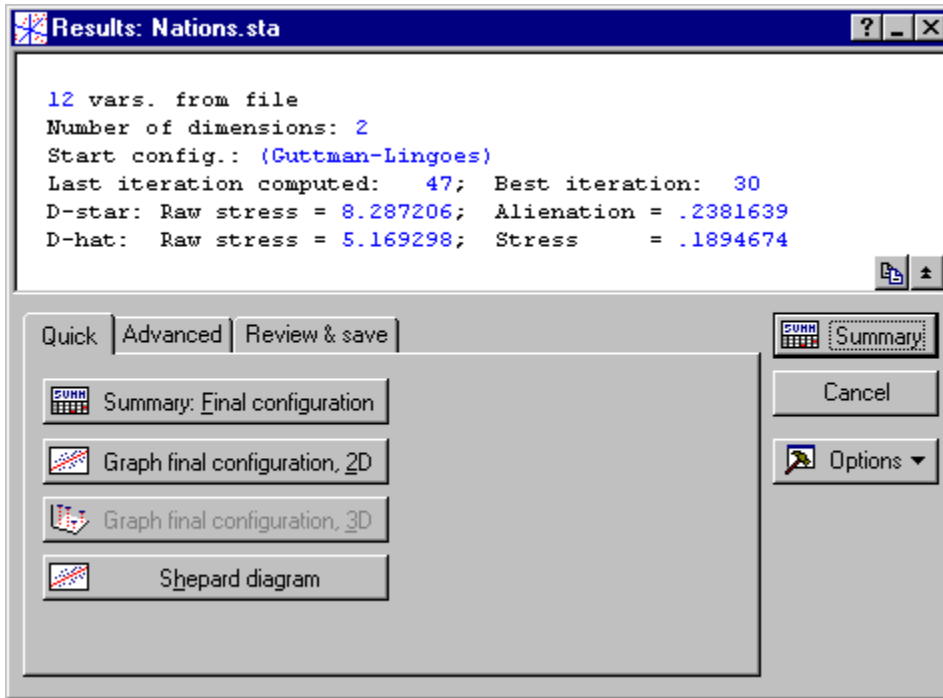
iter.	[dim=2]	D-star	D-star	D-hat	d-hat
s: t:	cosin step	raw stress	alienation	raw stress	stress
15	1	.968	.485	5.310299	.1920341
16	1	.887	.404	5.304970	.1919377
17	1	.694	.230	5.300331	.1918537
18	1	.711	.197	5.295306	.1917628
19	1	.908	.322	5.284995	.1915760
20	1	.960	.456	5.264813	.1912098
21	1	.898	.411	5.242467	.1908036
22	1	.700	.234	5.228747	.1905538
23	1	.662	.180	5.217968	.1903573
24	1	.877	.281	5.201593	.1900584
25	1	.956	.429	5.182857	.1897157
26	1	.561	.188	5.178150	.1896296
27	1	.665	.168	5.175625	.1895833
28	1	.950	.354	5.172096	.1895187
29	1	.968	.487	5.169822	.1894770
30	1	.698	.247	5.169298	.1894674
30	*			5.169298	.1894674
13	1	.697	.184	5.320363	.1922159
14	1	.938	.349	5.315846	.1921343

Estimation procedure converged

Buttons: Cancel, OK

After STATISTICA has determined the best two-dimensional configuration, it will display the final stress value; click the OK button to display the Results dialog.

Results. In general, you can examine the results in spreadsheets or graphs via the options available on the Results dialog.



First, examine the table of actual distances and estimated distances.

Reproduced and observed distances. To evaluate the fit of the two-dimensional solution, click the Summary statistics button on the Advanced tab.

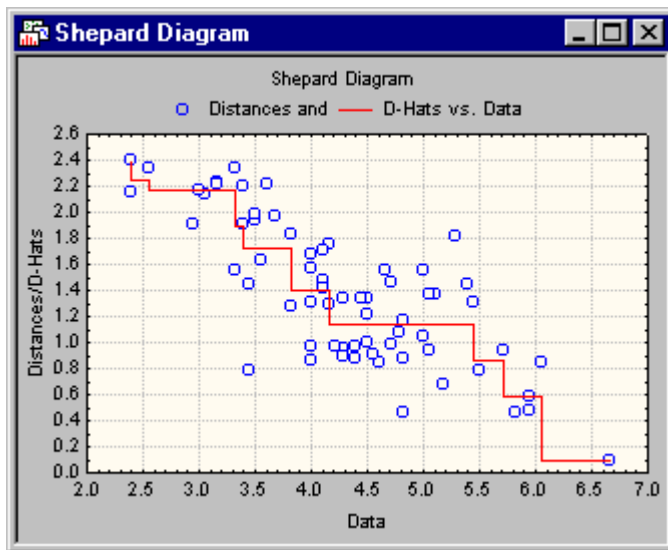
Final Configuration (Nations.sta)			
D-star: Raw stress = 8.287206; Alienation = .2381639			
D-hat: Raw stress = 5.169298; Stress = .1894674			
	Distance	D-star	D-hat
<b>D(12, 10)</b>	0.093247	0.093247	0.093247
D(11, 8)	0.849364	0.456247	0.591286
D(11, 5)	0.578455	0.462164	0.591286
D(11, 7)	0.481080	0.481080	0.591286
D(6, 4)	0.456247	0.578455	0.591286
D(10, 9)	0.938324	0.671463	0.865153
D(9, 3)	0.791982	0.786270	0.865153
D(10, 3)	1.312809	0.791982	1.133368
D(11, 1)	1.450855	0.842240	1.133368
D(3, 1)	1.817224	0.849364	1.133368
D(4, 3)	0.671463	0.861000	1.133368
D(12, 3)	1.362808	0.872578	1.133368
D(10, 5)	1.373663	0.874488	1.133368
D(12, 9)	0.933320	0.886758	1.133368
D(4, 2)	1.038965	0.907474	1.133368
D(11, 10)	1.561636	0.933320	1.133368
D(2, 1)	1.171765	0.938324	1.133368

There are four columns in this spreadsheet. The columns labeled D-hat and D-star contain the monotone transformations of the input data (see the Introductory Overviews): D-stars are rank images calculated according to Guttman (1968); D-hats are monotone regression estimates calculated according to Kruskal (1964).

The rows in the spreadsheet, each representing one distance as specified in the similarity matrix, are sorted according to the size of D-star or D-hat. The second column of the spreadsheet contains the reproduced Distances from the current configuration. If the fit of the current model (i.e., the current number of dimensions) is very good, then the order of reproduced distances should be approximately the same as that for the transformed input data (i.e., D-star or D-hat values). Out-of-order elements indicate lack of fit. The first column of the spreadsheet references the elements of the original input matrix as  $D(X,Y)$ , where X is the respective row in the input matrix, and Y is the respective column.

For example,  $D(2,1)$  would be the element in the second row and the first column of the input matrix (i.e., in our example, the comparison between Congo and Brazil). It appears that, by and large, the order of distances was approximately reproduced by the two-dimensional solution.

Shepard diagram. Now examine the Shepard plot. As described in the Introductory Overviews, this plot is a scatterplot of the observed input data (similarities or dissimilarities) against the reproduced distances. The plot will also show the D-hat values, that is, the monotonically transformed input data, as a step function. To produce this plot, click the Shepard diagram button on the Quick or Advanced tab.

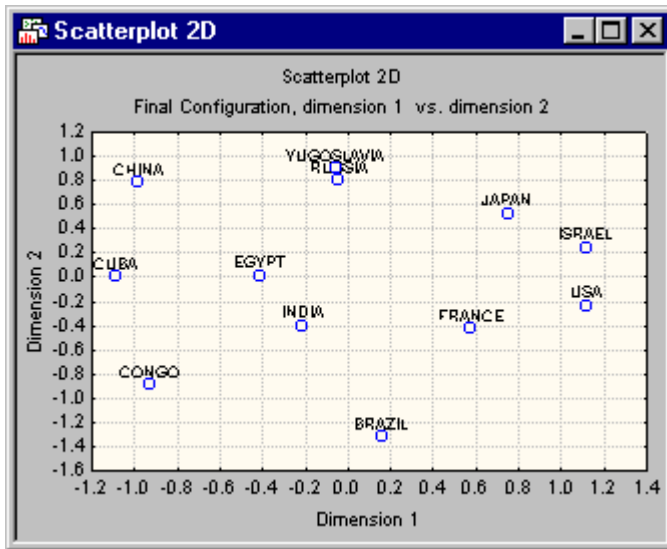


Most points in this plot are clustered around the step-line. Thus, you may conclude for now that this two-dimensional configuration is adequate for describing the similarities between countries.

Interpreting the configuration. In order to interpret this solution, you can display the configuration of nations in the two-dimensional space. Return to the Advanced tab and then click the Graph final configuration, 2D button. The Select two dimensions for scatterplot intermediate dialog will open in which you can select the dimensions for the 2D scatterplot. Select Dimension 1 as the First (X), Dimension 2 as the Second (Y), and then click the OK button to produce the plot.

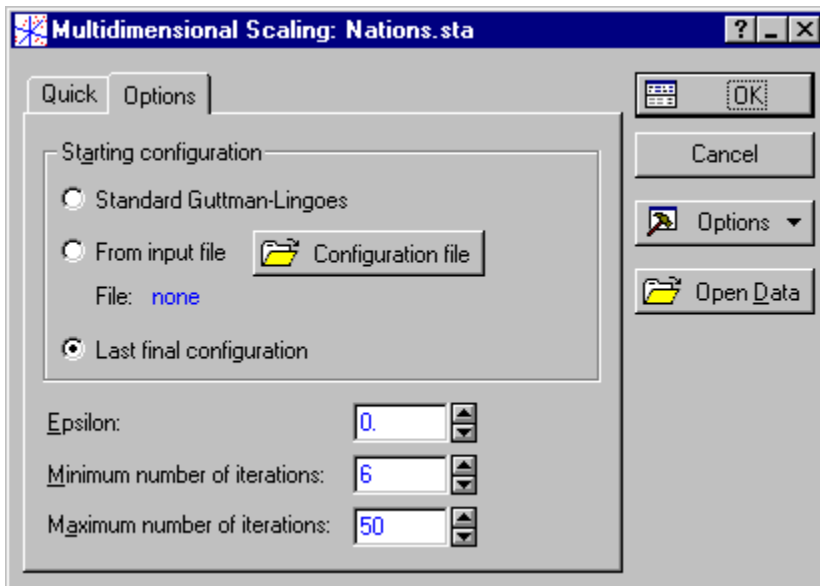
As described in the Introductory Overviews, the actual orientation of axes in multidimensional scaling is arbitrary (just as in Factor Analysis). Thus, you may rotate the configuration in order to achieve a more interpretable solution. Kruskal and Wish (1978) used a program called KYST (which uses a slightly different algorithm for multidimensional scaling) in order to analyze the present data, and they obtained a

very similar solution. They then rotated their solution by approximately 45 degrees, and interpreted the rotated dimensions as developed vs. underdeveloped, and pro-western vs. pro-communist. Looking at the plot below (and mentally rotating it by 45 degrees), this interpretation seems to hold quite well (remember that this study was conducted in the 1970's).



(Note that in the plot above the scaling has been adjusted via the All Options - Scaling tab.) In general, 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 Gutman, (1968).

Continuing the Analysis. Now click the Cancel button on the Results dialog to return to the Multidimensional Scaling Startup Panel.



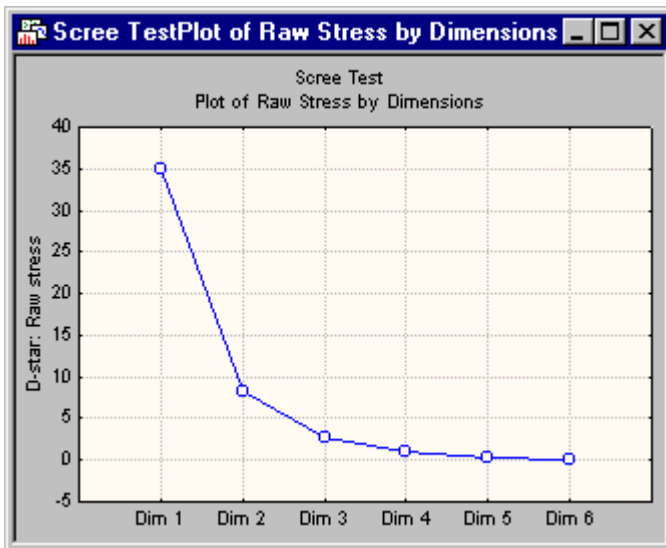
Note that now the default settings on the Options tab are different than when the program was first started. Multidimensional Scaling will remember the configuration from the previous analysis (unless you specify a new data file, or if you select new cases). Also, the default Number of dimensions on the Quick tab is now 1. You could now click the OK button to compute the one-dimensional solution, using the

configuration for the first dimension from the previous analysis as the starting configuration. In this manner, you may efficiently evaluate several consecutive solutions, starting with several dimensions (and working your way down to the one-dimensional solution).

Scree test: Plotting the stress values. This example began with the two-dimensional solution. Actually, if you are unsure about the dimensionality underlying the matrix, you should plot the stress values for consecutive numbers of dimensions. Then find the place where the smooth decrease of stress values appears to level off to the right of the plot. To the right of this point, presumably, one finds only "factorial scree" - "scree" is the geological term referring to the debris which collects on the lower part of a rocky slope (see, for example, Kruskal and Wish, 1978, pages 53-56, for a discussion of this plot). Shown below is this scree plot. (Note that this plot was created by first making a new spreadsheet containing the D-star: Raw stress values (that can be found in the Results dialog's summary box) for consecutive dimensions (1 through 6) for the present data

	1
	D-star: Raw stress
Dim 1	34.90807
Dim 2	8.289225
Dim 3	2.700427
Dim 4	0.984465
Dim 5	0.1885643
Dim 6	0.046122

and then selecting Line Plot (Variables) from the Graphs - 2D Graphs menu.)



Based on the plot above, the two-dimensional solution would, indeed, have been chosen. Perhaps you also would have looked at the three-dimensional solution. You can be the judge of whether the three-dimensional solution is more meaningful than the two-dimensional one. Shown below is a 3D scatterplot of the solution when 3 is specified as the Number of dimensions on the Multidimensional Scaling Startup Panel - Quick tab. To produce this graph, click the Graph of final configuration, 3D button on the Results - Quick tab (this button is dimmed if 1 or 2 dimensions were specified on the Startup Panel; note that

when you click this button you will be prompted to select the dimensions to plot in the 3D graph via the Select three dimensions for scatterplot dialog).

