

Running a Multiple Correspondence Analysis (MCA) with XLSTAT

[demoMCA.xls](#)

Multiple Correspondence Analysis

Multiple Correspondence Analysis (MCA) is a method that allows studying the association between two or more qualitative variables.

Multiple Correspondence Analysis is to qualitative variables what Principal Component Analysis is to quantitative variables. One can obtain maps where it is possible to visually observe the distances between the categories of the qualitative variables and between the observations. For detailed information on the method, see the [recent book](#) by Michael Greenacre and Jörg Blasius.

Dataset to run a Multiple Correspondence Analysis

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

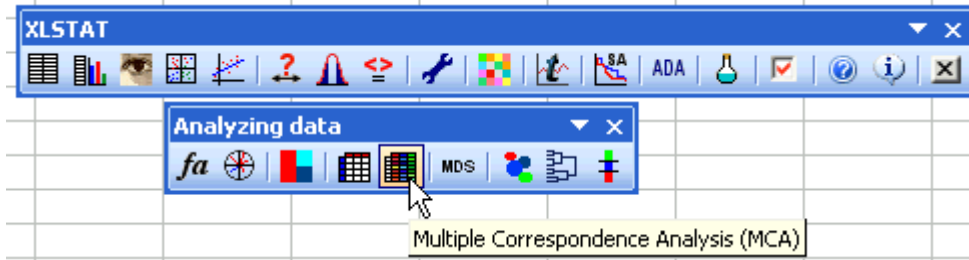
The data correspond to a survey conducted by a car dealer where 28 customers were asked five questions, one week after they had picked up their car after a mechanical repair. The questions were:

- Are you globally satisfied by the service? (Yes/No)
- Do you consider the problem is solved? (Yes/No/Don't know)
- How good was the welcome? (1 to 5)
- Is the quality/price ratio satisfactory? (Yes/No)
- Will you use our services again? (Yes/No/Don't know)

By running a Multiple Correspondence Analysis (MCA), we want to identify the relationships between the various possible answer to the questions.

Setting up the Multiple Correspondence Analysis dialog box

After opening XLSTAT, select the **XLSTAT / Analyzing data / Multiple Correspondence Analysis** command, or click on the corresponding button of the **Analyzing data** toolbar (see below).

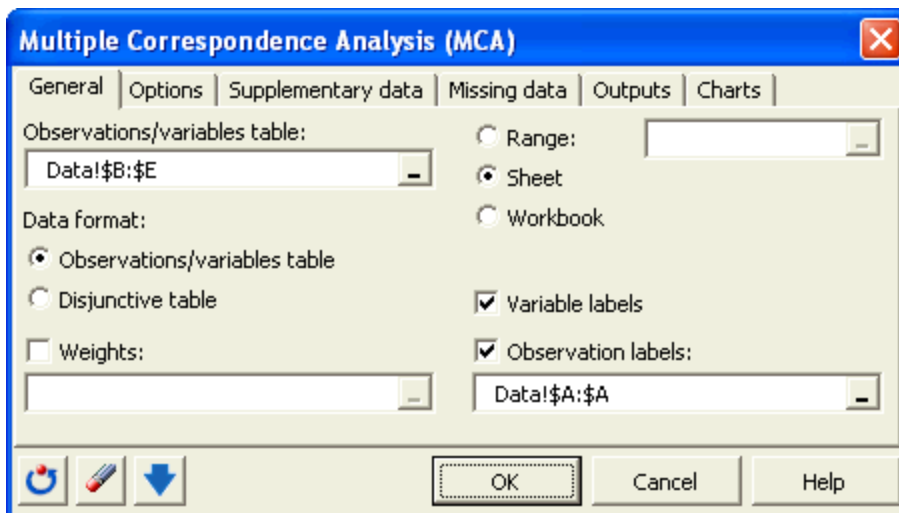


Once you've clicked on the button, the Multiple Correspondence Analysis dialog box appears.

The **format** of the data is here **Observations/Variables**.

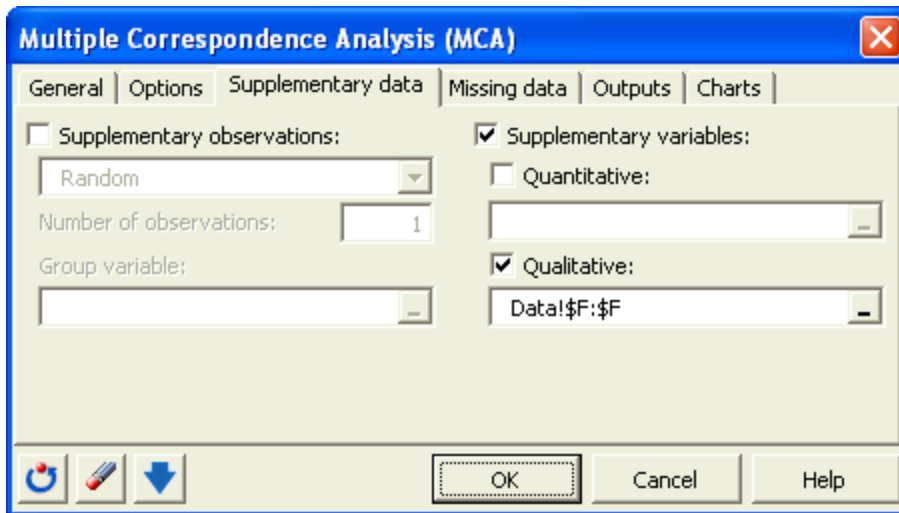
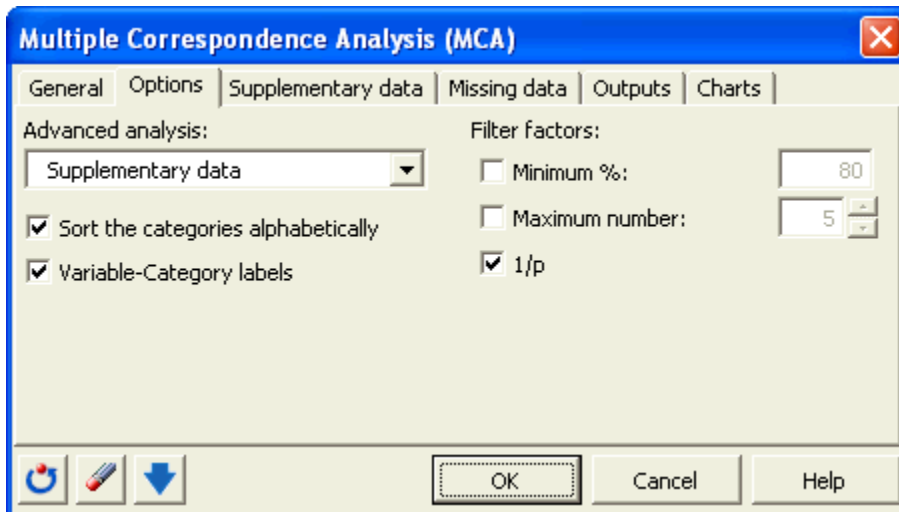
We select the data on the Excel sheet, using the column selection method: just click on the name of the columns you want to select (see the tutorial on how to select data for more information on this topic).

The **Observations labels** are selected in the corresponding field, and the **Variable labels** option is left activated as the first row of the table contains the name of the variables.

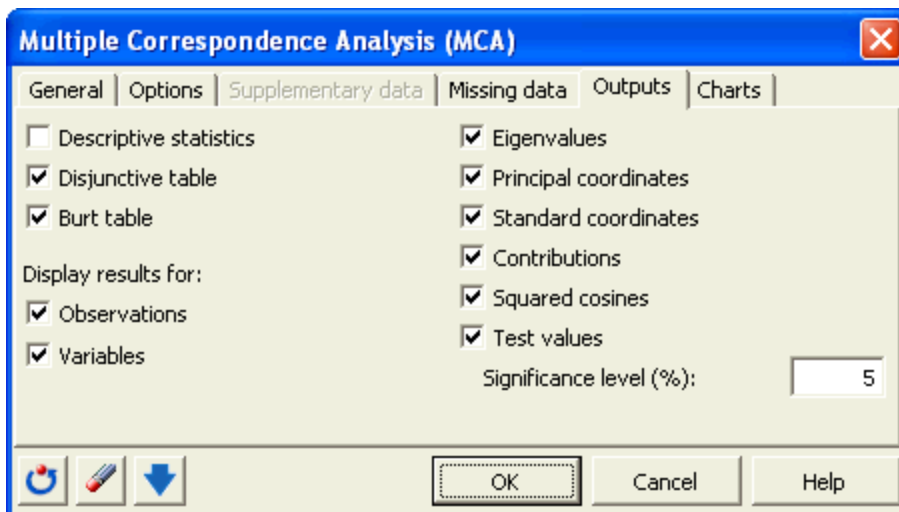


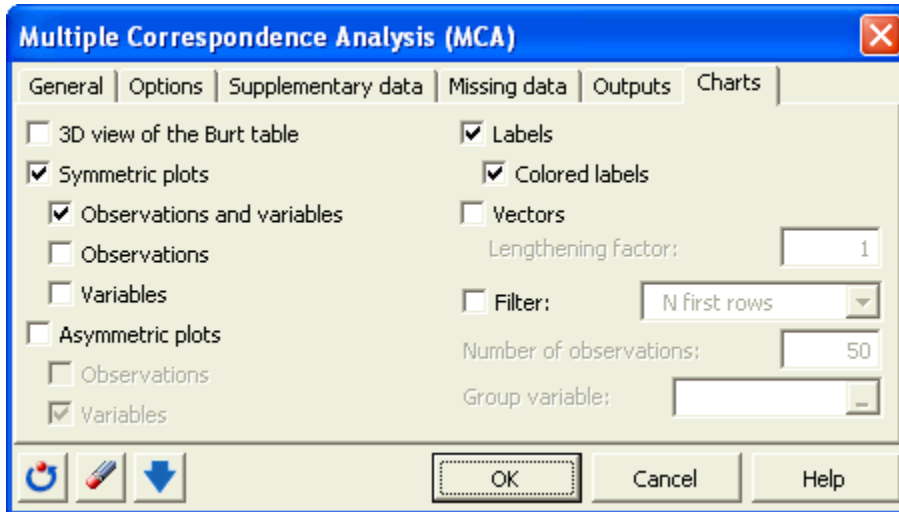
In the **Options** tab we activate the **Supplementary data** option and then go to the corresponding tab: the "Come back" variable is used as a supplementary variable because we don't want it to influence the computations; however, we want to know how the categories of this variable are positioned on the correspondence map.

The **1/p** option is our **filtering** choice: the detailed results corresponding to factors which eigenvalue is less than $1/p$ (where p is the number of active qualitative variables), will not be displayed.



The following **Outputs** and **Charts** options have been activated.





The computations begin once you have clicked on **OK**. The results will then be displayed.

Interpreting the results of a Multiple Correspondence Analysis

The first results displayed are the tables used for the computations (full disjunctive table, Burt's table).

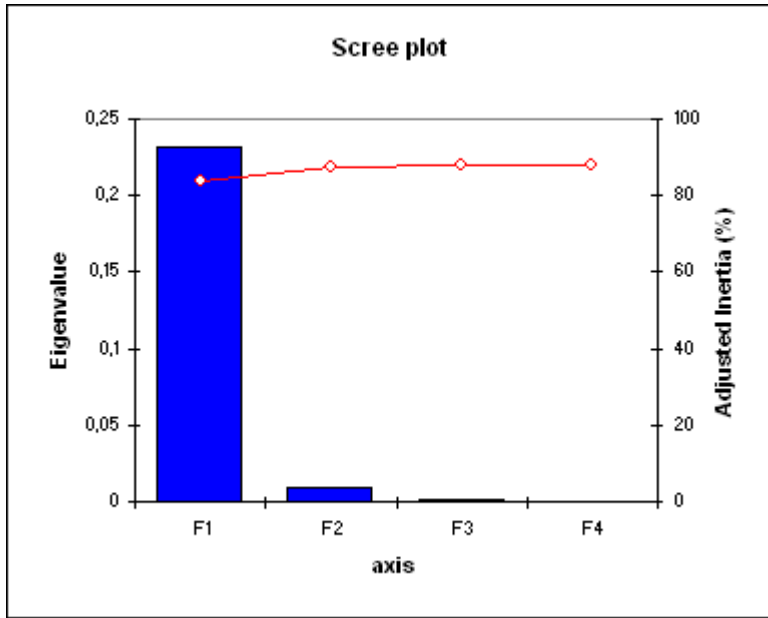
The total inertia is equal to 2. It depends only on the number of variables and categories and not on the linkage between the variables. Therefore, there is no possible statistical interpretation.

The next table shows the eight non null eigenvalues and the corresponding % of inertia. However, unlike with CA (correspondence analysis performed on only 2 variables), the % of inertia are here pessimistic estimates of the quality of the representation, the latter being for the user "how close is the representation to the reality".

Greenacre et al (2005) suggested an adjusted inertia which gives a better idea of the quality of the maps. We see here that while the usual computation gives us only 46.6% with the first two axes, the method based on the adjusted inertia gives us 87.3%.

Eigenvalues and percentages of inertia:								
	F1	F2	F3	F4	F5	F6	F7	F8
Eigenvalue	0,611	0,321	0,278	0,260	0,215	0,165	0,090	0,058
Inertia (%)	30,561	16,060	13,918	13,009	10,761	8,261	4,515	2,914
Cumulative	30,561	46,622	60,540	73,549	84,311	92,572	97,086	100,000
Adjusted In	0,232	0,009	0,001	0,000				
Adjusted In	83,997	3,264	0,518	0,067				
Cumulative	83,997	87,261	87,779	87,846				

The % displayed on the scree plot is based on the adjusted inertia.

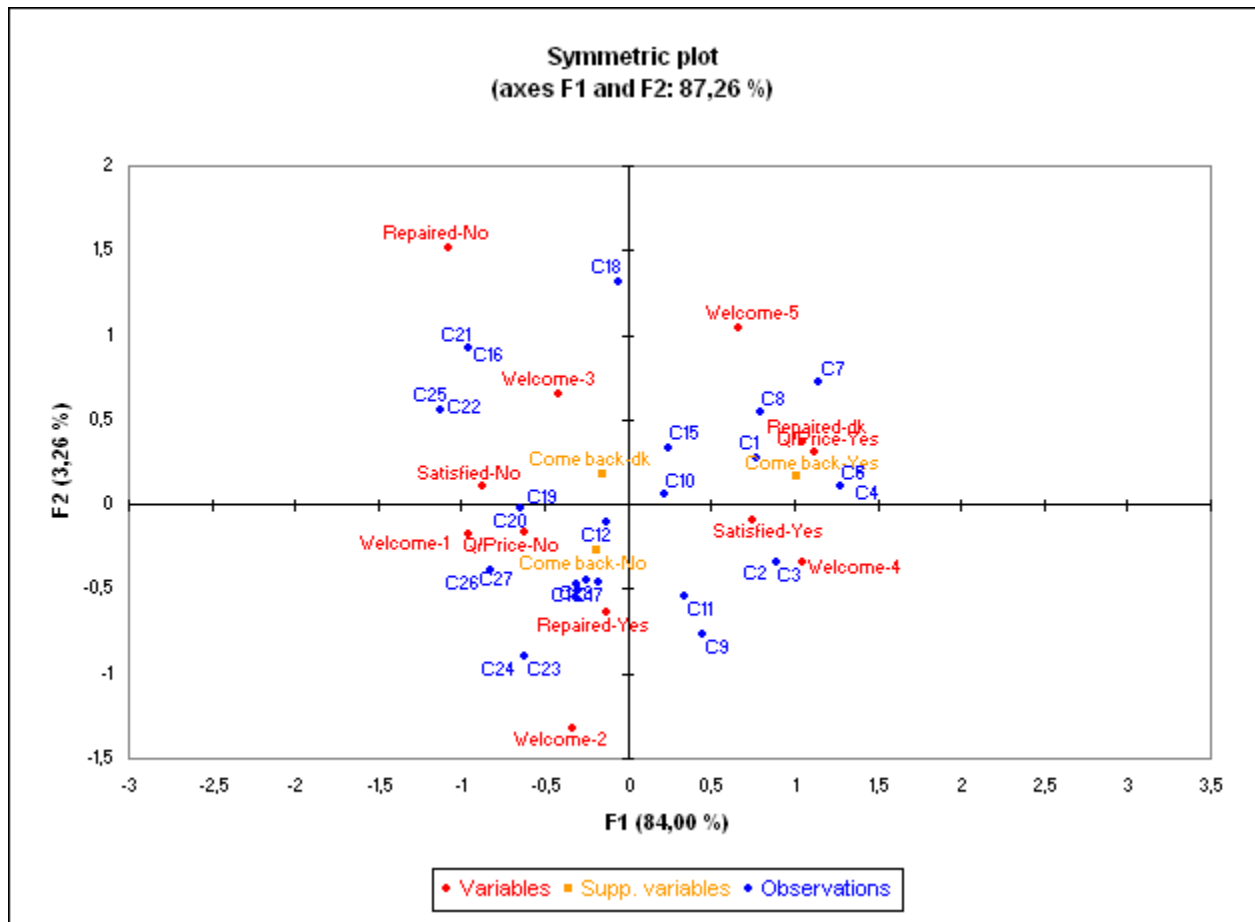


Then, a table displays the coordinates of the categories in the factors space. The results that correspond to the supplementary variable are displayed in blue color.

The coordinates of the observations are displayed further down.

The contributions, the test values and the squared cosines help in the interpretation of the results. Before interpreting that two categories are close on the map, one should check that their contribution to the axes of the map, or that their squared cosines are high.

The following chart corresponds to the correspondence map where both the categories and the observations are displayed on the first two axes.



In order to better visualize the relative positions of the categories, we have built with XLSTAT-3DPlot a visualization in the F1/F2/F3 space.

