

Creating a contingency table in XLSTAT

[contingency_file](#)

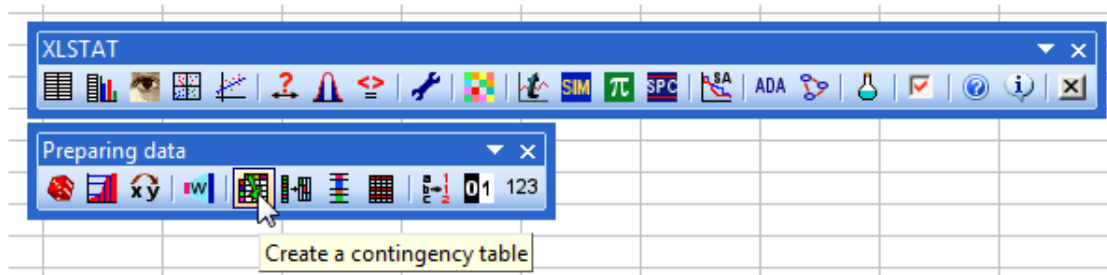
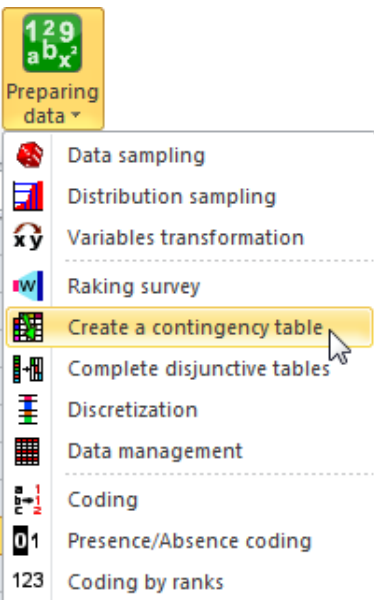
Dataset to create a contingency table

A MS Excel spreadsheet with both the data and the results can be downloaded by clicking [here](#). In this tutorial we show how to generate a contingency table.

The dataset contains information on 20 clients: their age, city of residence and gender. We will create a contingency table on the Age and City variables.

Creating a contingency table

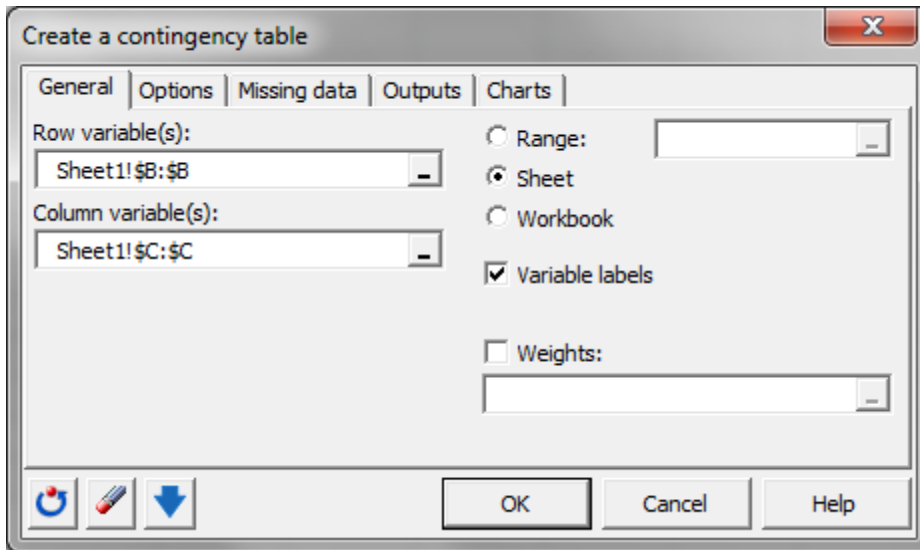
After opening XLSTAT, select the **XLSTAT / Preparing data / Create a contingency table** command, or click on the corresponding button of the **Preparing data** toolbar (see below).



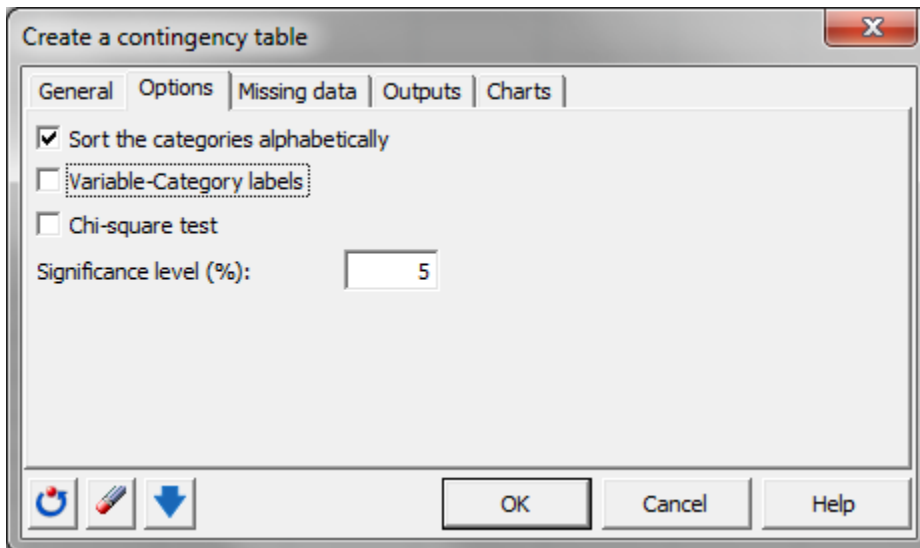
Once you've clicked on the relevant button, the dialog box appears. Select the category variable you wish to use in rows. Select the variable Age by selecting the entire column. Then select the variable to be used in Column, select the variable City.

The output should be displayed in a new sheet so tick the option Sheet.

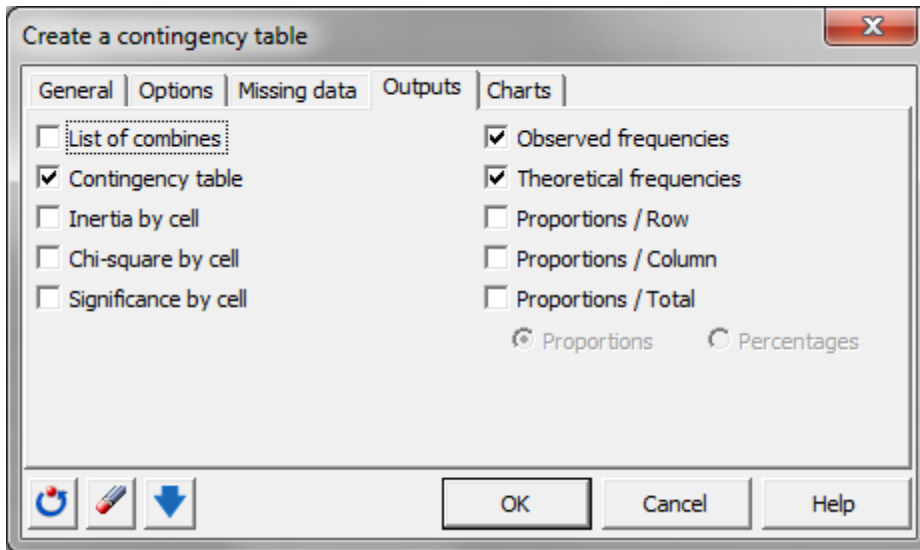
The columns B and C contain variable labels so the option Variable labels should be ticked.



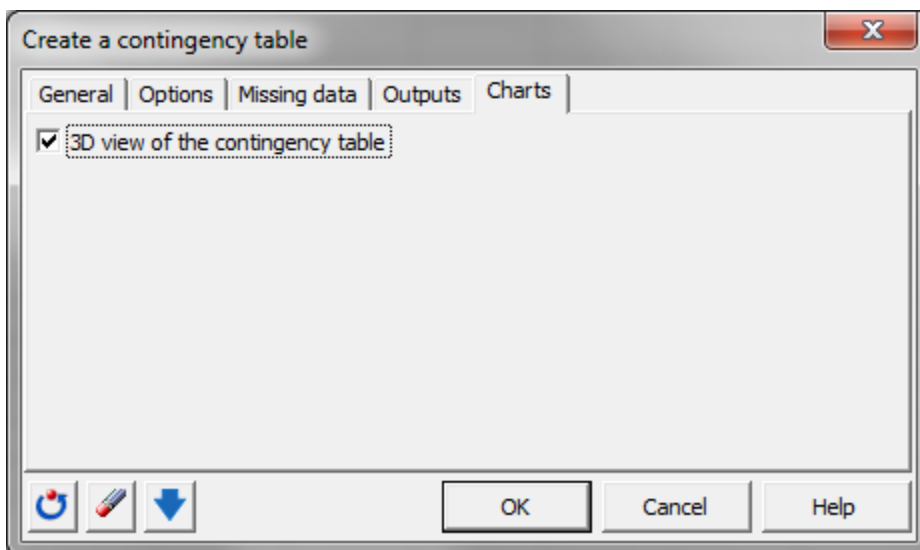
now you should proceed to the next tab **Options**. Here you can decide how the categories of the variable should be treated. Also you can do a Chi-square test. In this case, however, we will only select the option **Sort the categories alphabetically**.



In the tab **Outputs**, select the **Contingency table** as well as the **observed and theoretical frequencies**.



Select the **3D view of the contingency table** in the tab **Charts**.



The computations begin once you have clicked on the **OK** button, and the results are displayed on a new sheet.

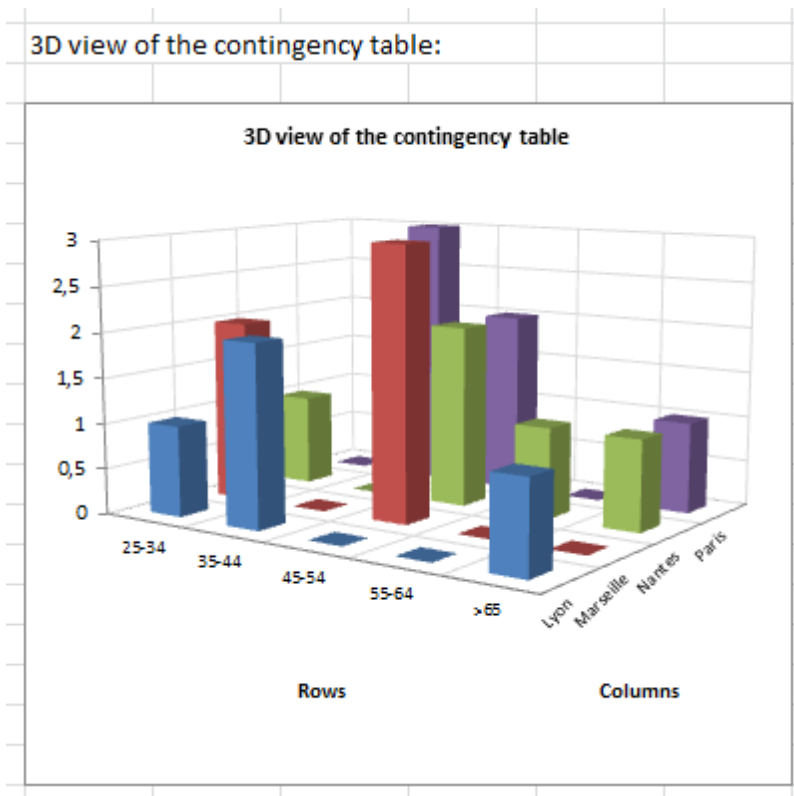
Results of the creation of a contingency table

The first result is the contingency table.

Notice that there isn't a client in all crossed categories. For example there is no client in Paris aged 25-34.

Contingency table (Age / City):				
	Lyon	Marseille	Nantes	Paris
25-34	1	2	1	0
35-44	2	0	0	3
45-54	0	3	2	2
55-64	0	0	1	0
>65	1	0	1	1

The next result is the 3-D plot, which is a good visualization of the data distribution.



Next are the two tables containing the frequencies Age/City.

You can compare the actual distribution of the clients and the theoretical distribution if the distribution was random.

Observed frequencies (Age / City):						Theoretical frequencies (Age / City):					
	Lyon	Marseille	Nantes	Paris	Total		Lyon	Marseille	Nantes	Paris	Total
25-34	1	2	1	0	4	25-34	0,800	1,000	1,000	1,200	4
35-44	2	0	0	3	5	35-44	1,000	1,250	1,250	1,500	5
45-54	0	3	2	2	7	45-54	1,400	1,750	1,750	2,100	7
55-64	0	0	1	0	1	55-64	0,200	0,250	0,250	0,300	1
>65	1	0	1	1	3	>65	0,600	0,750	0,750	0,900	3
Total	4	5	5	6	20	Total	4	5	5	6	20