

Computing a Spearman correlation coefficient and test if it is significant or not

[demoCorrSpearman.xls](#)

Dataset to run a Spearman correlation coefficient test

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

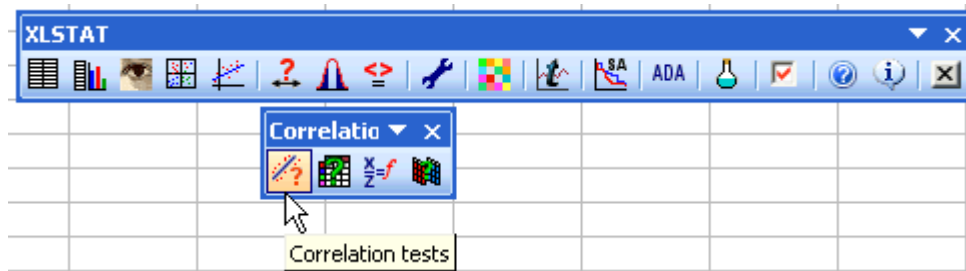
The data used in this example correspond to a survey where a given brand/type of potato chips has been evaluated by 100 consumers. Each consumer gave a rating on 1 to 5 scale for four attributes (Saltiness, Sweetness, Acidity, Crunchiness) - 1 means "little", and 5 "a lot" -, and then gave an overall liking score on a 1-10 likert scale. Our goal is to check how the attributes are correlated with the liking score.

Correlations are computed in many of the XLSTAT features. However two functions are dedicated to that: the **Describing data / Similarity/Dissimilarity matrices** feature, and the **Correlation/Association tests / Correlation tests** feature.

In this tutorial, we use the **Correlation/Association tests / Correlation tests** tool.

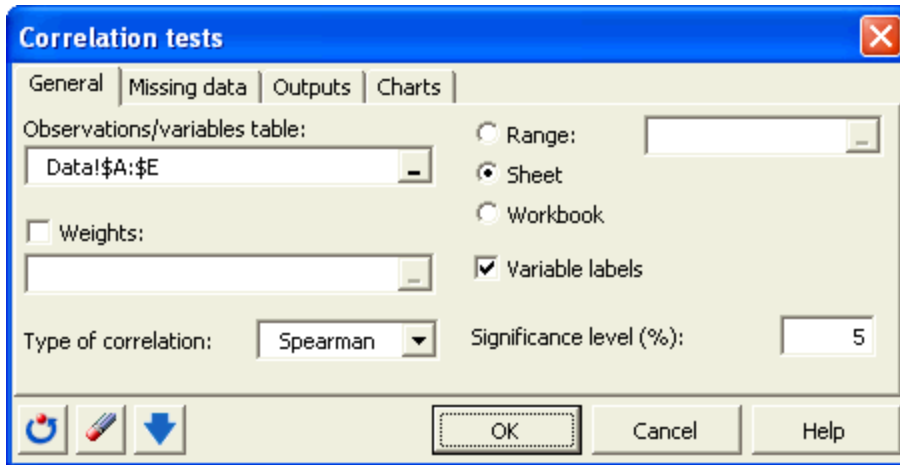
Setting up a Spearman correlation coefficient test

After opening XLSTAT, select the **Correlation/Association tests / Correlation tests** command, or click on the corresponding button of the **Correlation/Association tests** toolbar (see below).

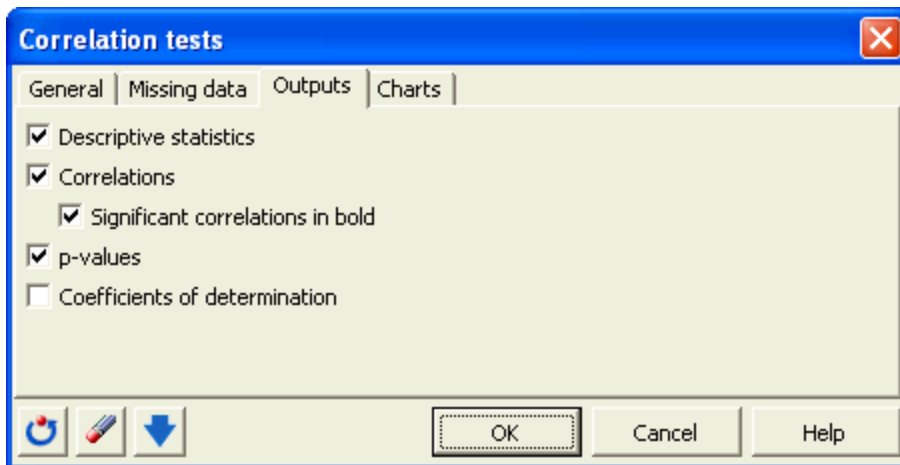


Once you've clicked on the button, the dialog box appears. We select the liking scores and the four attributes in the **Observations/Variables** box. As the first row of the table corresponds to headers, we leave the **Variable labels** option checked.

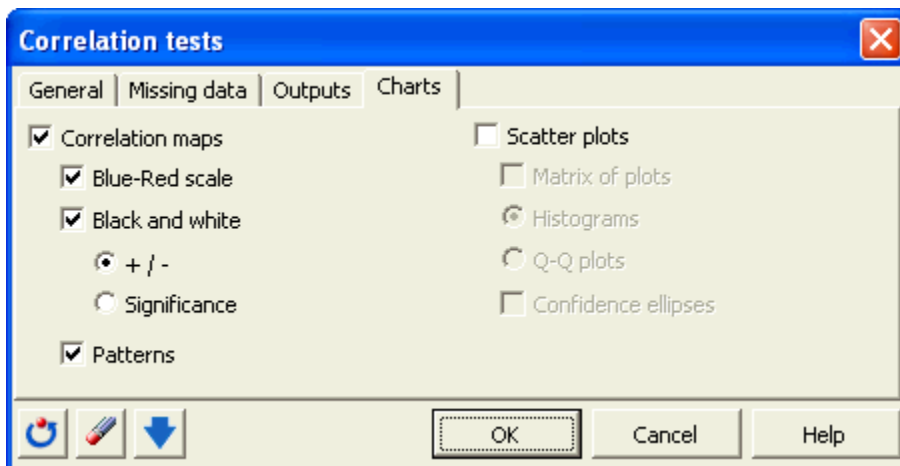
Because the data are not continuous but ordinal, we choose to use the **Spearman correlation coefficient** instead of the Pearson correlation coefficient which is the usual one for continuous data.



In the **Outputs** tab, we choose the results we want to display.



In the **Charts** tab, we select the correlations maps we want to display.



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

Interpreting the results of a Spearman correlation coefficient test

The first results are the descriptive statistics for the liking data and the attributes. The correlations matrix is then displayed.

Correlation matrix (Spearman):					
Variables	liking score	Saltiness	Sweetness	Acidity	Crunchiness
Liking score	1	0,272	0,030	-0,035	0,466
Saltiness	0,272	1	-0,122	-0,077	0,113
Sweetness	0,030	-0,122	1	-0,223	-0,081
Acidity	-0,035	-0,077	-0,223	1	0,158
Crunchiness	0,466	0,113	-0,081	0,158	1

Values in bold are significantly different from 0 with a significance level $\alpha=0,05$

The correlations between the liking scores and the attributes are mostly low. The correlation between the liking scores and the Crunchiness is higher than average with 0.466. This suggests that the only criterion for which we have "the more, the better" is the crunchiness. For the other criteria it is likely that there is an intermediate optimal level, above which the consumers will show their dissatisfaction.

Values displayed in bold are significant at 0.05 significance level. That means that the risk of being wrong when rejecting the null hypothesis that the correlations are not significantly different from 0 is less than 5%.

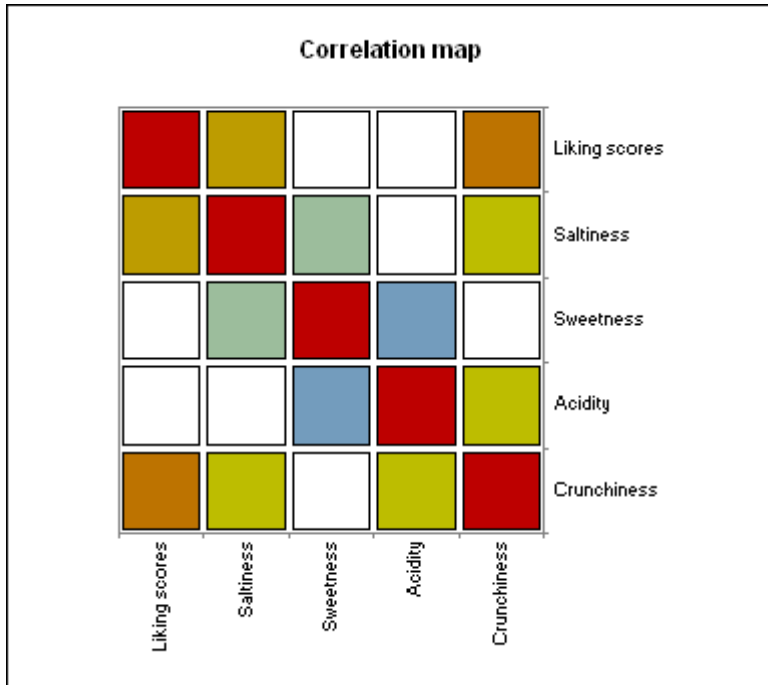
The corresponding p-values (the estimated risk) are displayed in the next table. The lower the p-value, the more different from 0 the correlation. Note: the tests are two-sided.

p-values:					
Variables	liking score	Saltiness	Sweetness	Acidity	Crunchiness
Liking score	0	0,006	0,767	0,733	< 0,0001
Saltiness	0,006	0	0,228	0,445	0,263
Sweetness	0,767	0,228	0	0,026	0,423
Acidity	0,733	0,445	0,026	0	0,117
Crunchiness	< 0,0001	0,263	0,423	0,117	0

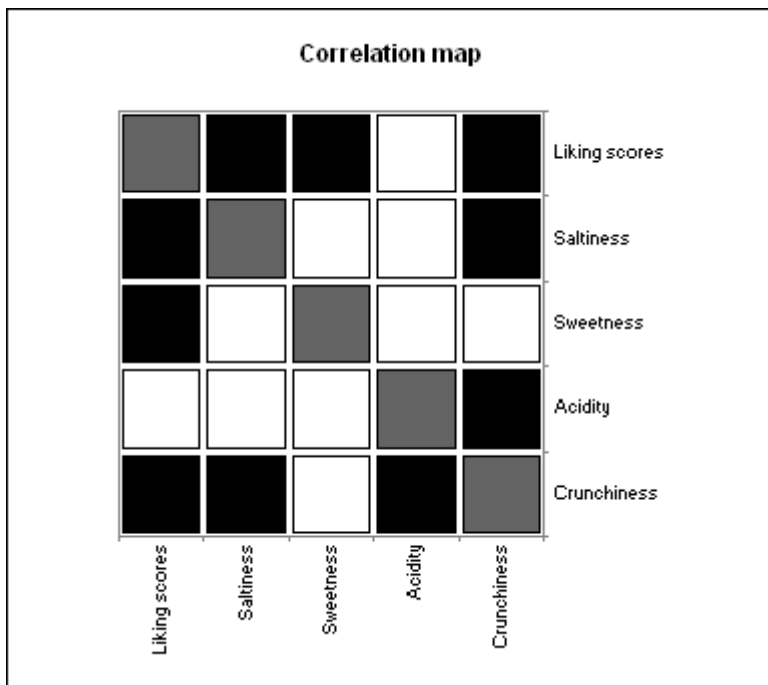
Values in bold are significantly different from 0 with a significance level $\alpha=0,05$

XLSTAT is among the few software programs that generates correlation maps. Correlation maps allow you to visually identify patterns in correlations. While this has more meaning when there are many dimensions, we take advantage of this example to show how the maps can be used.

The first correlation map uses a blue-red (cold-hot) scale to display the correlations. The blue color corresponds to a correlation close to -1 and the red color corresponds to a correlation close to 1. Green corresponds to a correlation close to 0.

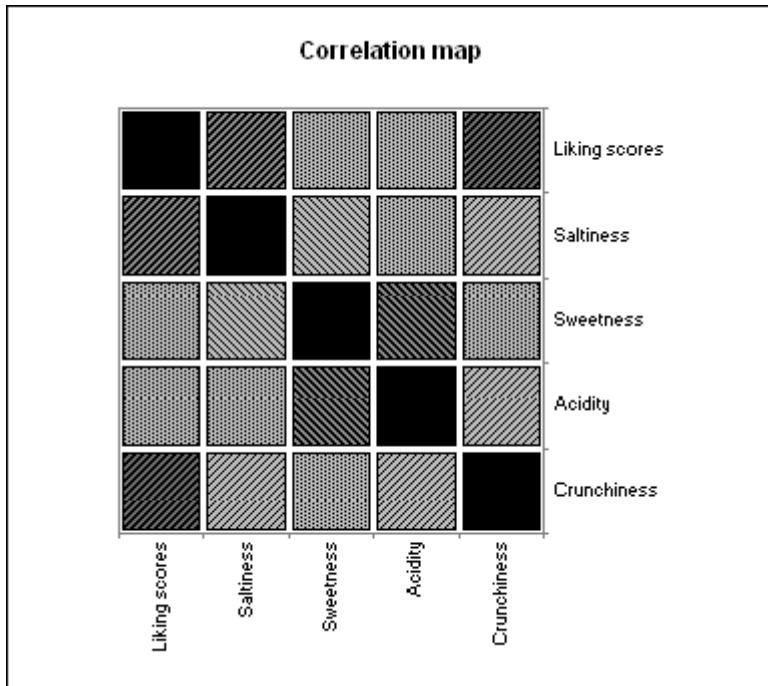


The second correlation map uses black and white to respectively identify positive and negative correlations. The diagonal is displayed in grey color.



The third correlation map uses patterns to identify both the sign and the intensity of the correlations:

- lines that go from the bottom left to the top right correspond to positive correlations, and vice versa;
- the tighter the lines, the closer the correlation to 0.



Note: if you only want to compute the correlation coefficient between two samples, you can use directly in the Excel sheet the XLSTAT XLSTAT_Spearman function. Example: on the "Data" sheet, type in the G2 cell, "`=XLSTAT_Spearman(A2:A101,B2:B101)`".