Sampling data in a distribution and running a Normality test with XLSTAT

demoNorm.xls

Goal of this tutorial

In this tutorial we first show how to generate a sample from a Normal distribution, then a sample from a uniform distribution. Then we show you how to run normality tests on both samples.

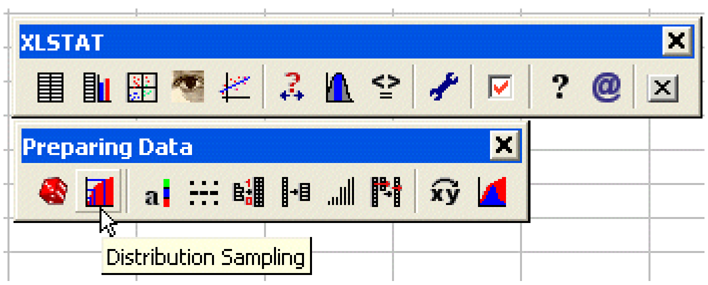
Dataset for sampling a distribution and running a Normality test

An Excel sheet with both the data and the results can be downloaded by clicking here.

We want to generate two samples, the first one sampled in a Normal N(2,4) distribution (mean = 2, variance = 4), the second one in a Uniform distribution, between -1.5 and 5 (mean = 2 and variance = 49/12 =4.08). To do that, we use the "Distribution sampling" tool available in the "Preparing data" section.

Setting up the sampling of data in a distribution

After opening XLSTAT, select the **XLSTAT / Preparing data / Distribution sampling** command, or click on the corresponding button of the **Preparing data** toolbar (see below).

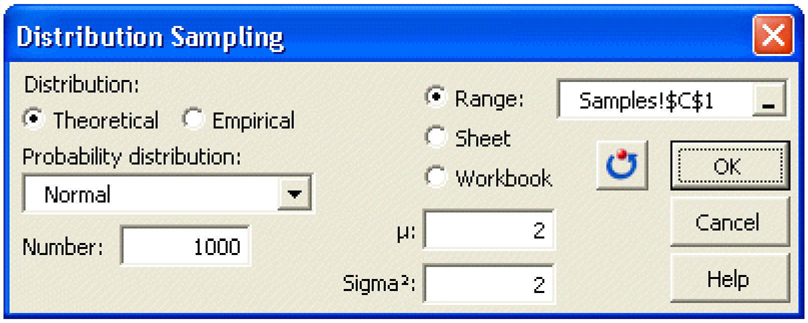


Once you've clicked on the button, the dialog box appears.

Select the distribution to use, and the corresponding parameters.

Then enter the size of the sample to generate.

The dialog box displayed below corresponds to the generation of a **1000 cases normal N(2,2) sample**.

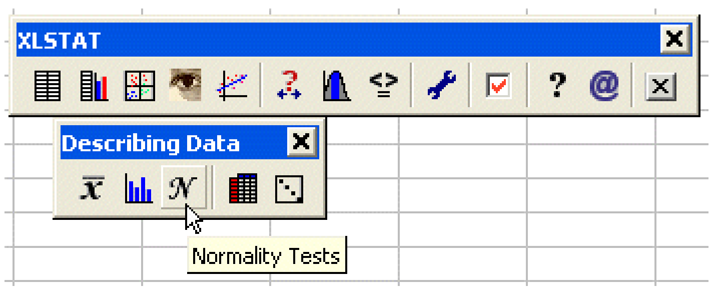


The computations begin once you have clicked on the **OK** button, and the sample is displayed.

A second sample is then generated using a **Uniform distribution between -1.5 and 5**.

Setting up the Normality test

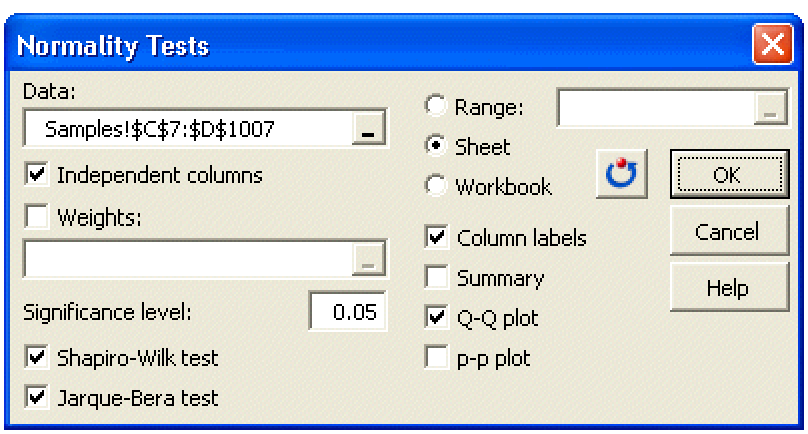
We then want to tests the normality of the two samples. Select the **XLSTAT / Describing data / Normality tests**, or click on the corresponding button of the **Describing data** toolbar.



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

Select the two samples, and activate the **Independent columns** option to confirm that the two samples are independent.

The **Q-Q plot** option is activated to allow us to visually check the normality of the samples.

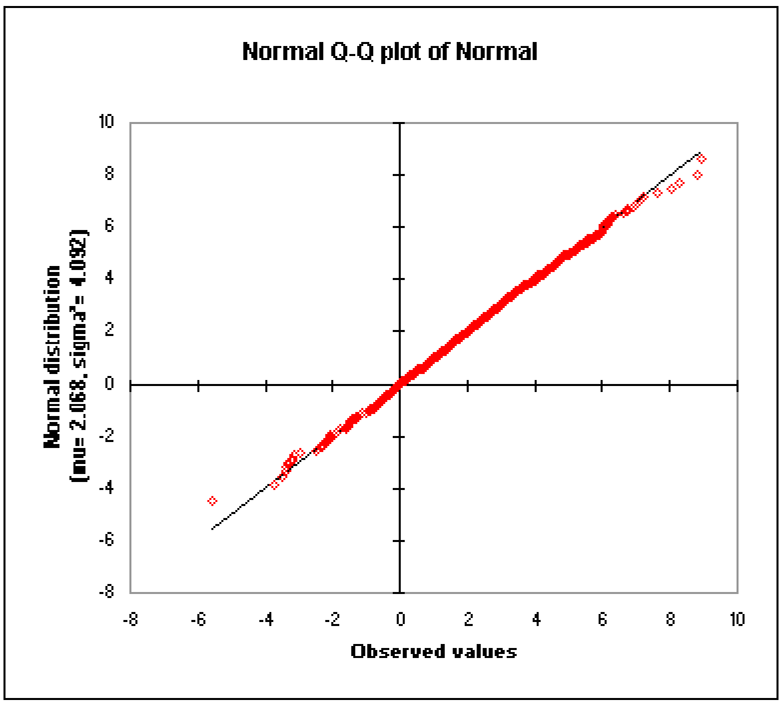


The computations begin once you have clicked on the **OK** button, and the results are displayed on a new sheet.

Interpreting the results of the Normality test

The results are first displayed for the first sample and then for the second sample.

The first result displayed is the Q-Q plot for the first sample. The Q-Q plot allows to compare the cumulative distribution function (cdf) of the sample (abscissa) to the cumulative distribution function of normal distribution with the same mean and standard deviation (ordinates). In the case of sample following a normal distribution, we should observe an alignment with the first bisecting line. In the other cases some deviations from the bisecting line should be observed.

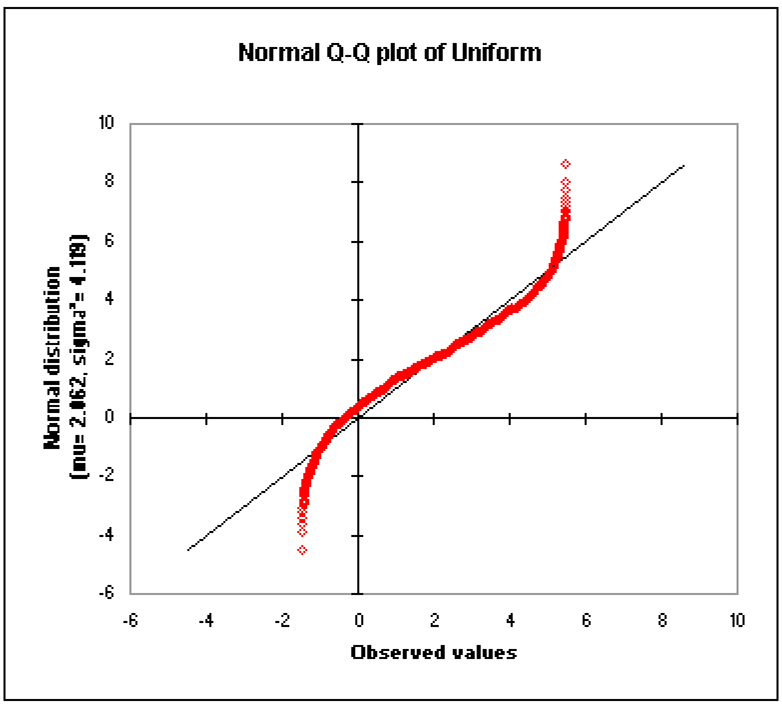


We can see here that the empiric cdf is very close to the bisecting line. The Shapiro-Wilk and Jarque-Bera confirm that we cannot reject the normality assumption for the sample. We notice that with the Shapiro-Wilk test, the risk of being wrong when rejecting the null assumption is greater than with the Jarque-Bera test.

Shapiro-Wilk test:	
W (observe	0.999
One-tailed	0.804
Alpha	0.05
Decision:	
At the level of significance alpha=0.050 the decision is to not reject the null hypothesis that the sample follows a normal distribution.	
In other words, the non-normality is not significant.	

Jarque-Bera test:	
JB (observ	2.239
JB (critical	5.991
DF	2
One-tailed	0.326
Alpha	0.05
Decision:	
At the level of significance alpha=0.050 the decision is to not reject the null hypothesis that the sample follows a normal distribution.	
In other words, the non-normality is not significant.	

The following results are for the second sample. Contrary to what we observed for the first sample, we notice here on the Q-Q plot that there are two strong deviations indicating that the the distribution is most probably not normal.



This gap is confirmed by the normality tests (see below) which allow to assert with no hesitation that we need to reject the hypothesis that the sample might be normally distributed.

Shapiro-Wilk test:

W (observe	0.955
One-tailed	< 0.0001
Alpha	0.05

Decision:
At the level of significance alpha=0.050 the decision is to reject the null hypothesis that the sample follows a normal distribution.
In other words, the non-normality is significant.

Jarque-Bera test:

JB (observ	59.154
JB (critical	5.991
DF	2
One-tailed	< 0.0001
Alpha	0.05

Decision:
At the level of significance alpha=0.050 the decision is to reject the null hypothesis that the sample follows a normal distribution.
In other words, the non-normality is significant.

Conclusion

As a conclusion, in this tutorial we have seen to generate two samples, one following a Normal distribution, the second following a Uniform distribution. We then confirmed on these samples the validity of the Shapiro-Wilk and Jarque-Bera tests: these tests have confirmed the normality assumption for the first sample, and allowed us to reject it for the second sample.