

# Running a Student's t test on two independent samples

[demoStudentTest.xls](#)

## Dataset to run a Student's t test on two independent samples

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

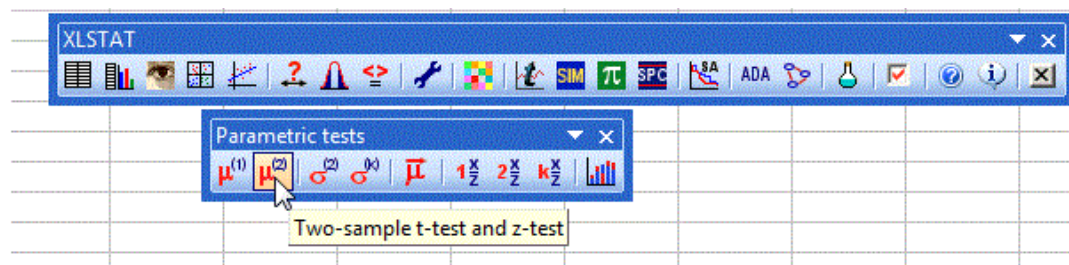
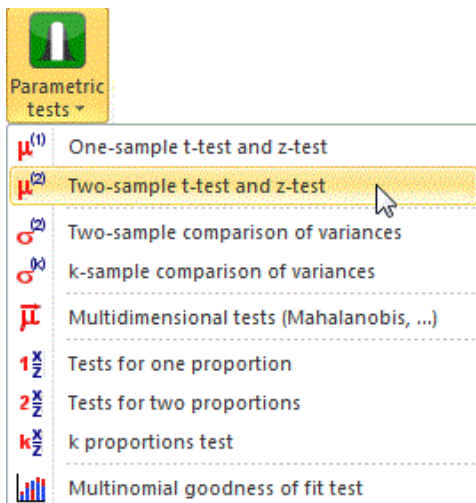
The data are from [Fisher M. (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7, 179 -188] and correspond to 100 Iris flowers, described by four variables (sepal length, sepal width, petal length, petal width) and their species. The original dataset contains 150 flowers and 3 species, but we have isolated for this tutorial the observations belonging to the versicolor and virginica species. Our goal is to test if for the four variables, there is a clear difference between the two species.



Iris versicolor and virginica.

## Setting up a Student's t test on two independent samples

After opening XLSTAT, select the **XLSTAT / Parametric tests / Two-sample t and z-test** command, or click on the corresponding button of the **Parametric tests** toolbar (see below).

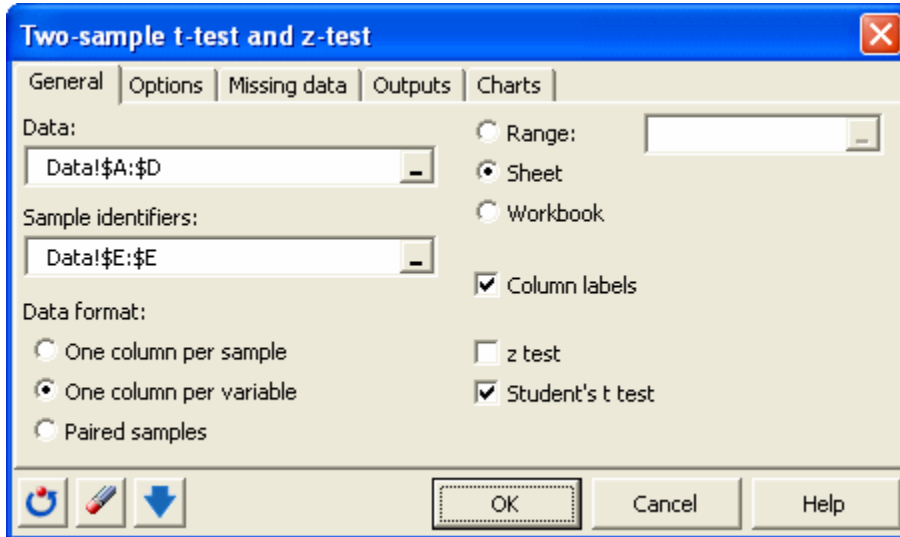


Once you've clicked on the button, the dialog box appears.

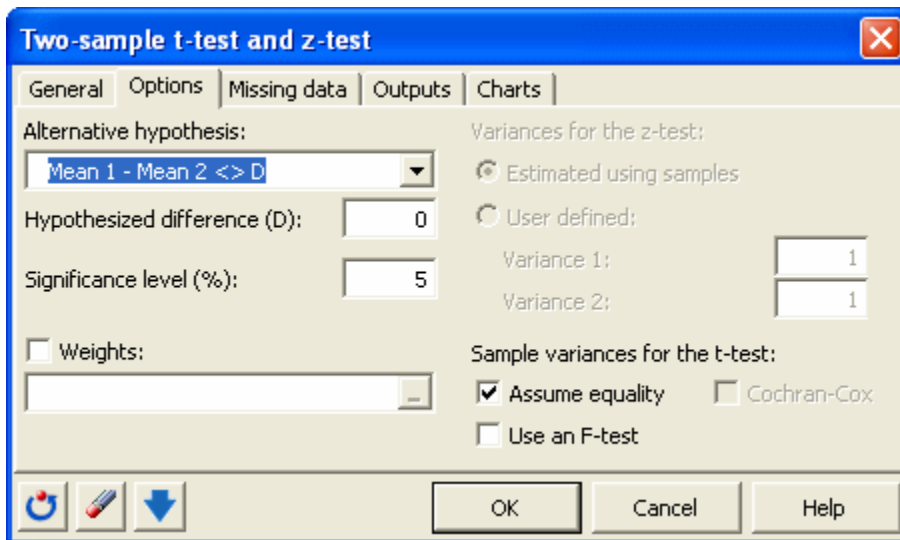
The **format** of the dataset is such that you have **one variable per column**.

If we were running the test for one variable only (out of the four), we could also have one column for each specie. The third format is for paired samples (in that case, there must be one column per sample).

We choose to run only the **t-test**.



In the **Options** tab, we leave the default options. In the **Charts** tab, we activate the option for the dominance diagram.



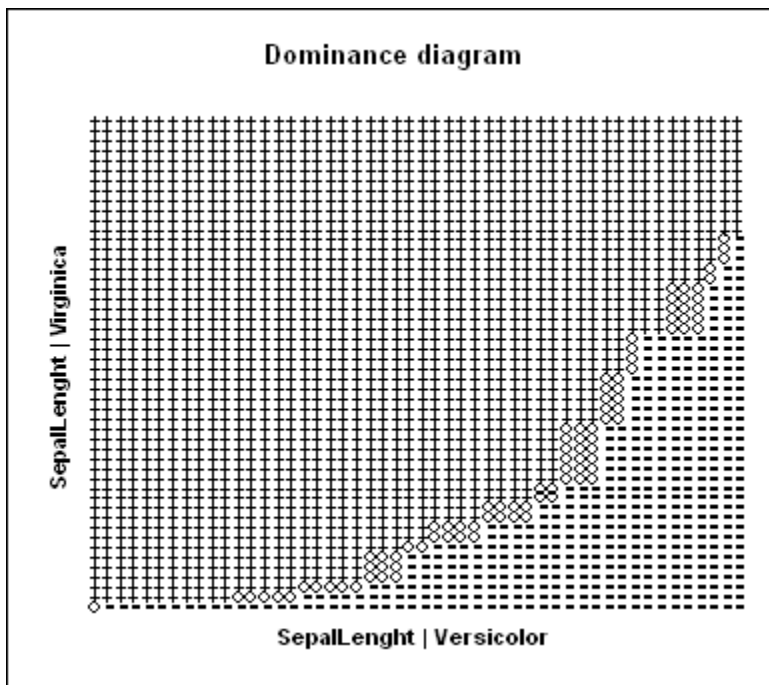
The computations begin once you have clicked on **OK**. The results will then be displayed in a new sheet.

# Interpreting the results of a Student's t test on two independent samples

The first results displayed are the statistics for the various samples. Next, the t tests and the dominance diagrams are displayed for each dimension one after the other.

The results for the first of the four variables are displayed below.

t-test for two independent samples / Two-tailed test (SepalLength):	
95% confidence interval on the difference between the means:	
] -8,819 ; -4,221 [	
Difference	-6,520
t (Observed)	-5,629
t (Critical value)	1,984
DF	98
p-value (Two-tailed)	< 0,0001
alpha	0,05
Test interpretation:	
H0: The difference between the means is equal to 0.	
Ha: The difference between the means is different from 0.	
As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.	
The risk to reject the null hypothesis H0 while it is true is lower than 0,01%.	



The difference appears to be very significant for the Sepal Length.