

1 Subject

Parametric and non parametric tests for differences in scale.

The tests of equal variability (or dispersion, or scale, or simply variance) are often presented as a preliminary test before the comparison of means, in order to verify the homoscedasticity assumption. But this is not their only purpose. Compare dispersions can be an end in itself. For example, we wish to compare the performance of two systems of heating. The average temperature at the center of the room is the same; however one can wish to compare the mode of diffusion of heat in different parts of the room.

The parametric tests are based primarily on the Gaussian distribution. The test becomes a test for homogeneity of variance. We highlight the Levene test in this tutorial. Other tests exist (Bartlett test for instance), we mention them in this tutorial.

When the normality assumption is questionable, when sample size is low, when the variable is ordinal and not continuous, it is more appropriate to use non parametric tests. These are called tests for equality of scales or dispersions. In fact the procedures are not based on estimated variances. We will use well known techniques such as the Ansari-Bradley test, the Mood or the Klotz test. They have a scope broader since nonparametric. Some of these tests have a drawback, they are not applicable when the distributions conditionals do not share the same parameter of central tendency (the median in general, but we can adjust the values by centering in relation to the median).

In this tutorial, we show how to implement these various tests with Tanagra.

2 Dataset

The data describe the performance of ecological heating devices¹. The variable of interest is the temperature inside a hut in the early morning fall. We have $n=15$ instances. The first is the reference group with $n_1=15$. The two others groups ($n_2=15$ and $n_3=15$) use new heating systems that we want to evaluate. We want to compare the differences in dispersion among the groups.

3 The parametric tests

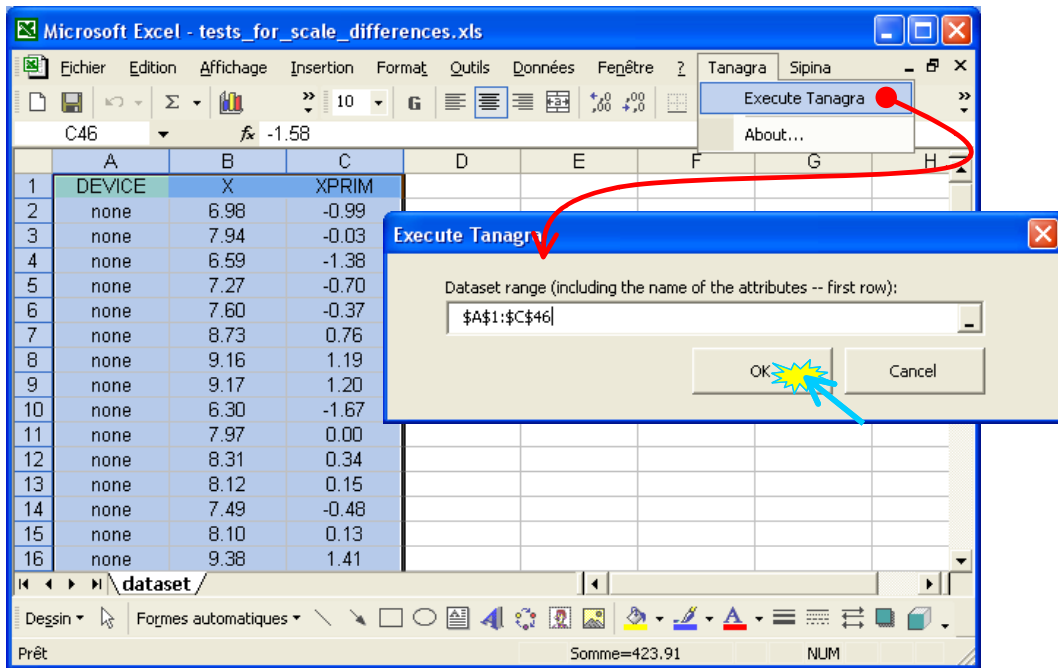
3.1 Importing the dataset

The simplest way to handle the data file is to open it into the Excel spreadsheet. We can send the values to Tanagra by using the TANAGRA.XLA add-in².

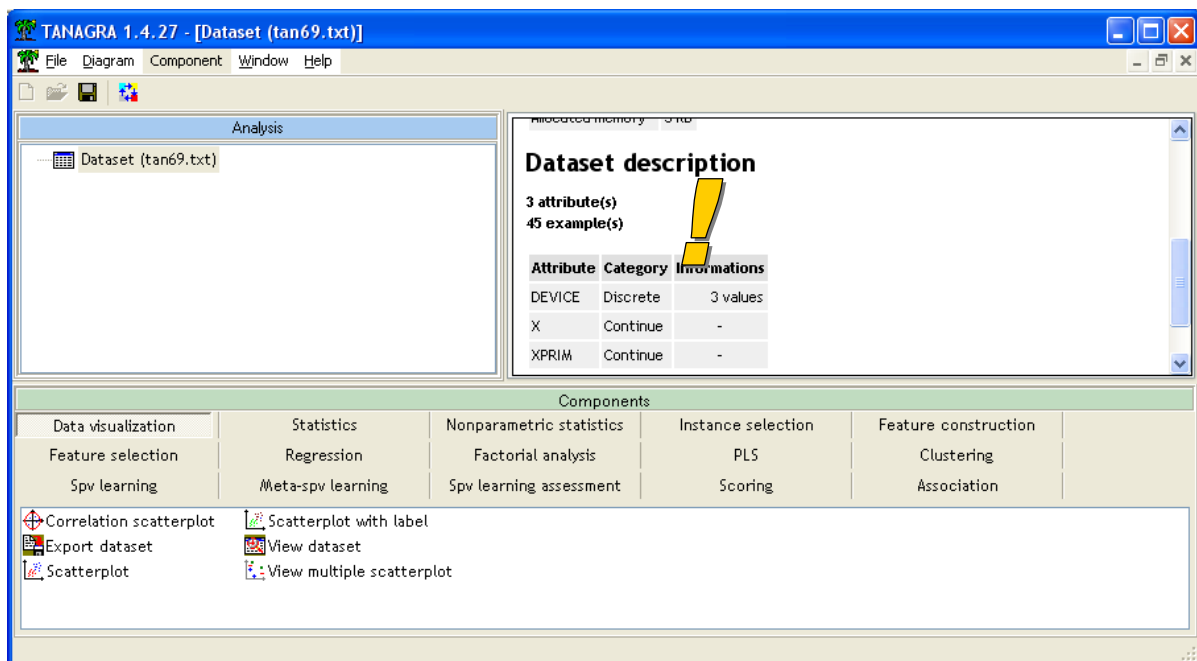
¹ http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/tests_for_scale_differences.xls

² See <http://data-mining-tutorials.blogspot.com/2008/10/excel-file-handling-using-add-in.html> for the installation and the utilization of the add-in. TANAGRA can handle also directly the XLS file format, see <http://data-mining-tutorials.blogspot.com/2008/10/excel-file-format-direct-importation.html>

We click on the TANAGRA / EXECUTE TANAGRA menu. We check carefully the cells selection in the dialog settings which appears.



TANAGRA is automatically launched. We note that we have 45 observations and 3 variables (DEVICE, X and XPRIM). The dependent variable is X. We use XPRIM later.

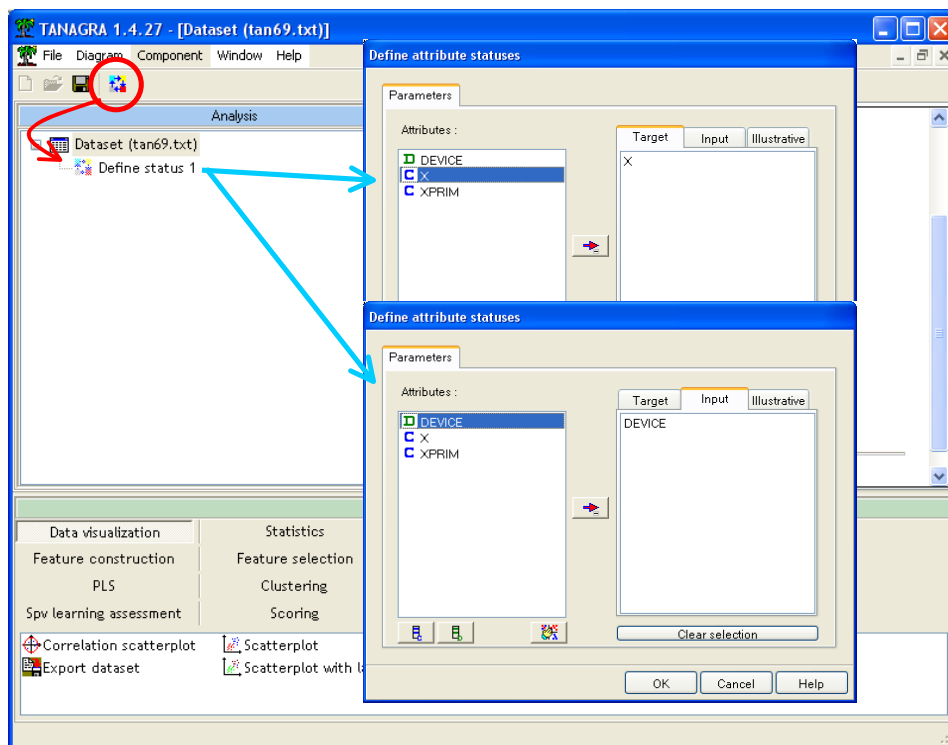


3.2 Levene test

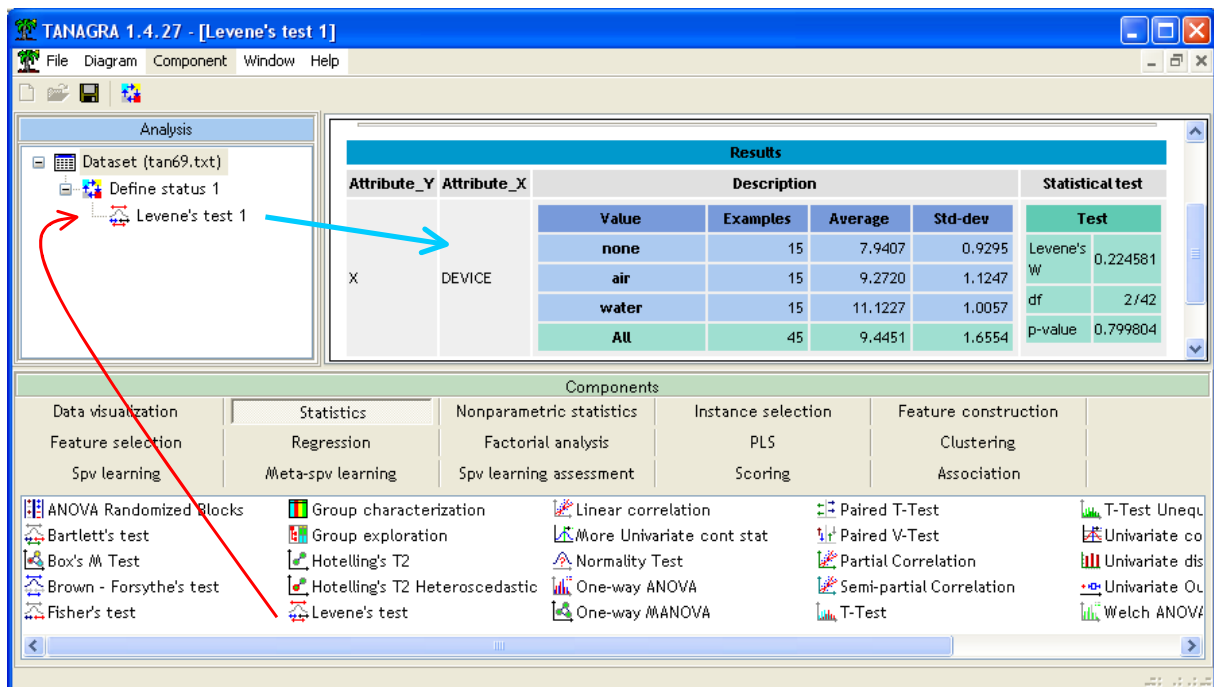
In order to compare the variances, we use the Levene test. It relies on the normality assumption³.

³ <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm>

We insert the DEFINE STATUS component: we set X as TARGET, DEVICE as INPUT.



Then, we add the LEVENE TEST component (STATISTICS tab). We click on the VIEW menu in order to obtain the following results.



The results show that the observed scores are compatible with the equality of variances assumption at the significance level of 5% (p-value = 0.799804).

Note: Users often prefer the Bartlett test (BARTLETT'S TEST into TANAGRA), perhaps because it is available in the various tools. But the Levene test is much more reliable. It is less sensible to departures of the normality assumption. In fact, the Bartlett test is more efficient only when the normality assumption is true⁴. Another alternative is the Brown and Forsythe test (BROWN-FORSYTHE'S TEST into TANAGRA)⁵. Its scope is also larger than the Bartlett test.

4 Non parametric tests

4.1 Inadequacy of the usual non parametric tests on our dataset

The previous parametric tests lead to the acceptance to the null hypothesis. The conditional variances are the same into the subgroups. Now, we want to use a non parametric Ansari Bradley test. We add the ANSARI-BRADLEY SCALE TEST component (NON PARAMETRIC STATISTICS tab) into the diagram. We click on the VIEW menu.

The screenshot shows the TANAGRA 1.4.27 interface. The 'Analysis' tree on the left includes 'Dataset (tan69.txt)', 'Define status 1', 'Levene's test 1', and 'Ansari-Bradley Scale Test 1'. The 'Results' table displays the following data:

Attribute_Y	Attribute_X	Description				Statistical test		
		Value	Examples	Average	Scores sum	Scores mean	One-way Analysis	
X	DEVICE	none	15	7.9407	158.0000	10.5333	Chi-Square	6.35786
		air	15	9.2720	228.0000	15.2000	d.f.	2
		water	15	11.1227	143.0000	9.5333	p-value	0.04163
		All	45	9.4451	529.0	11.7556		

The 'Components' panel at the bottom lists various non-parametric tests, with 'Ansari-Bradley Scale Test' highlighted. Other tests include Friedman's ANOVA by Ranks, Goodman-Kruskal Tau, Kendall's tau, Mann-Whitn, Categorical r, FYTH 1-way ANOVA, Kendall Tau-b, Klotz Scale Test, Median test, Cochran's Q-test, Goodman Kruskal Gamma, Kendall Tau-c, Kruskal-Wallis 1-way ANOVA, Mood Runs, Contingency Chi-Square, Goodman-Kruskal Lambda, Kendall's Concordance W, K-S 2-sample test, and Mood Scale.

The conclusion is different. At the significance level of 5%, it seems that the dispersion is different according to the groups (p-value of the test is $p = 0.04163$)⁶.

What is the correct conclusion? Which test is the most reliable in our context?

We know that the non parametric tests for equal dispersion are sensitive to the departure from the equality of central tendency parameters of the conditional distribution. The reject of the null hypothesis can be the consequence of the difference of the conditional means.

⁴ <http://www.itl.nist.gov/div898/handbook/eda/section3/eda357.htm>

⁵ <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm>

⁶ The other non parametric tests for the equality of dispersion give the same conclusion.

We see in the screenshot above that the mean seem different from one group to another ($\bar{x}_1 = 7.94$; $\bar{x}_2 = 9.27$; $\bar{x}_3 = 11.12$). One way to alleviate this drawback is to remove from the scores in each group the conditional central tendency indicator (e.g. the median).

4.2 Transformation of the dependent variable

The simplest way is to compute the deviation from the conditional median of the scores in each group. Let x_{ik} the score of the observation n^oi of the group n^ok , we get

$$x'_{ik} = x_{ik} - \tilde{x}_k$$

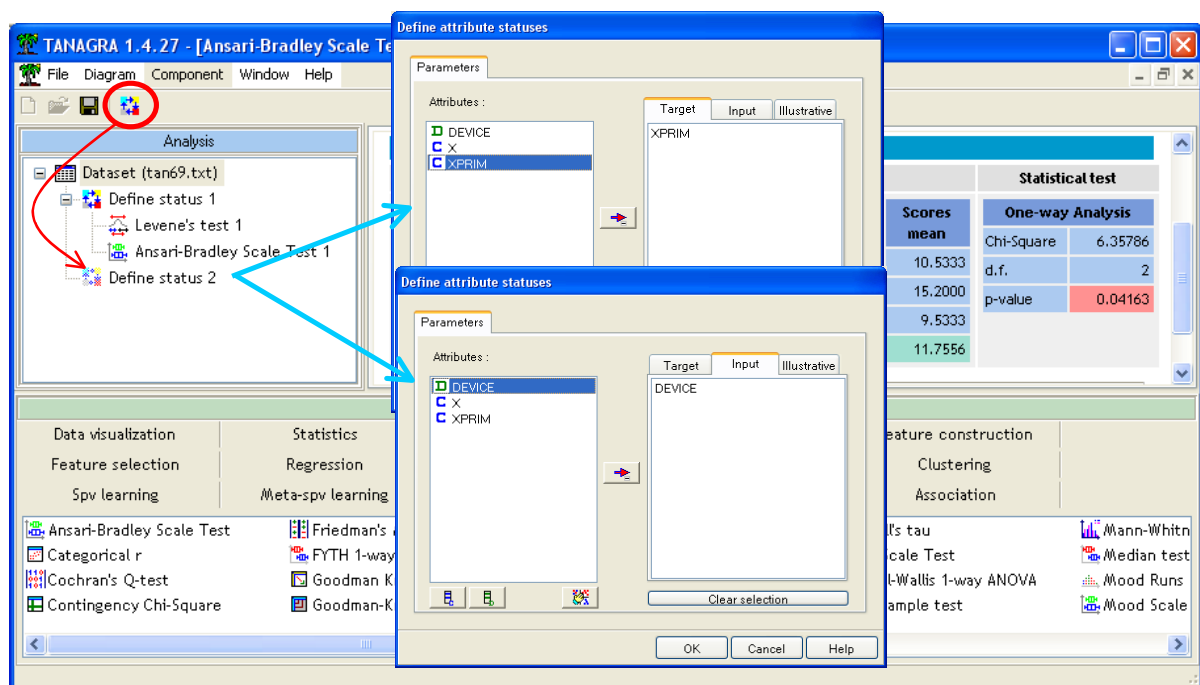
Where \tilde{x}_k is the median into the group n^ok .

For instance, for the first group (DEVICE = NONE), the median is $\tilde{x}_1 = 7.97$. We then obtain ($x'_{11} = 6.98 - 7.97 = -0.99$; $x'_{21} = 7.94 - 7.97 = -0.03$; $x'_{31} = 6.59 - 7.97 = -1.38$; etc.).

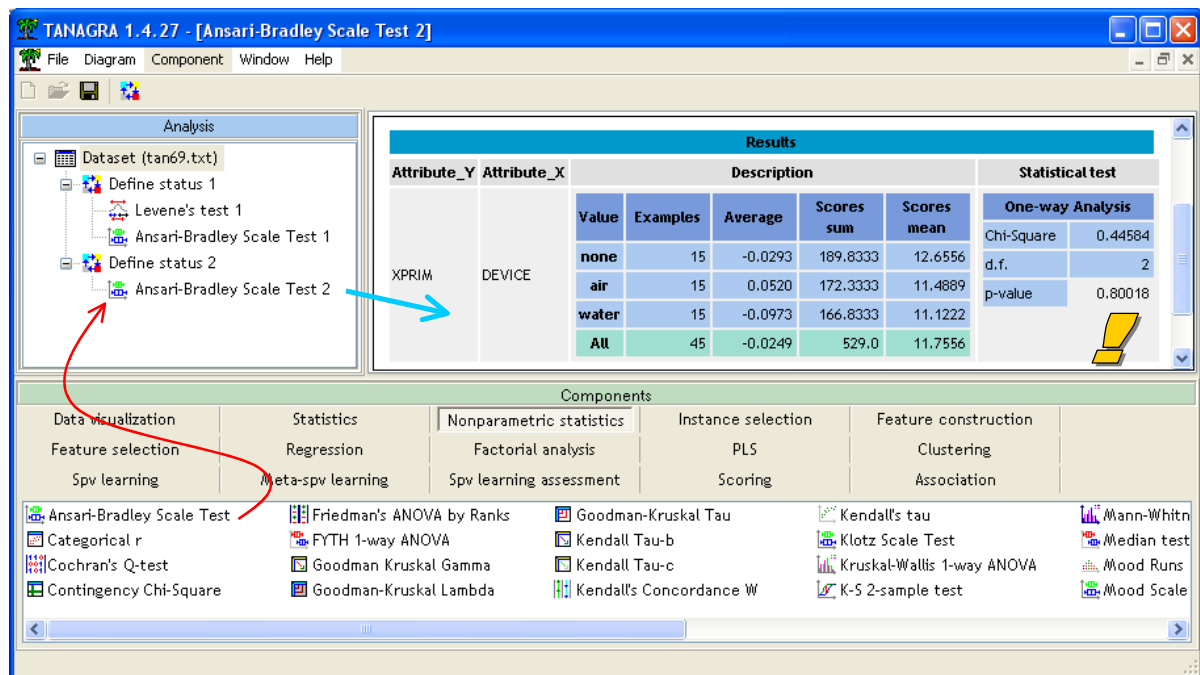
We perform the same transformation for the instance of the other groups. We thus obtain the variable XPRIM.

4.3 Ansari-Bradley test on XPRIM

We insert again the DEFINE STATUS component into the diagram. We set XPRIM as TARGET, DEVICE as INPUT.



We add the ANSARI-BRADLEY SCALE TEST component (NONPARAMETRIC STATISTICS tab).



The result is quite different. It is perfectly consistent with the Levene test now. The test statistic is $\chi^2 = 0.44584$. Under the null hypothesis, it follows a chi-square distribution with 2 degrees of freedom. The p-value of the test is $p = 0.80018$.

The previous Ansari-Bradley test conclusion, obtained without precautions on the variable X was clearly distorted by differences between the central tendencies of the conditional distributions. The transformation made on data is essential in this example.

Note: We must use the true value of the conditional medians to obtain the exact distribution of the test statistic under the null hypothesis. The process used here, the utilization of the estimated parameters during the data transformation, is only justified when we handle a large sample.

4.4 Mood test

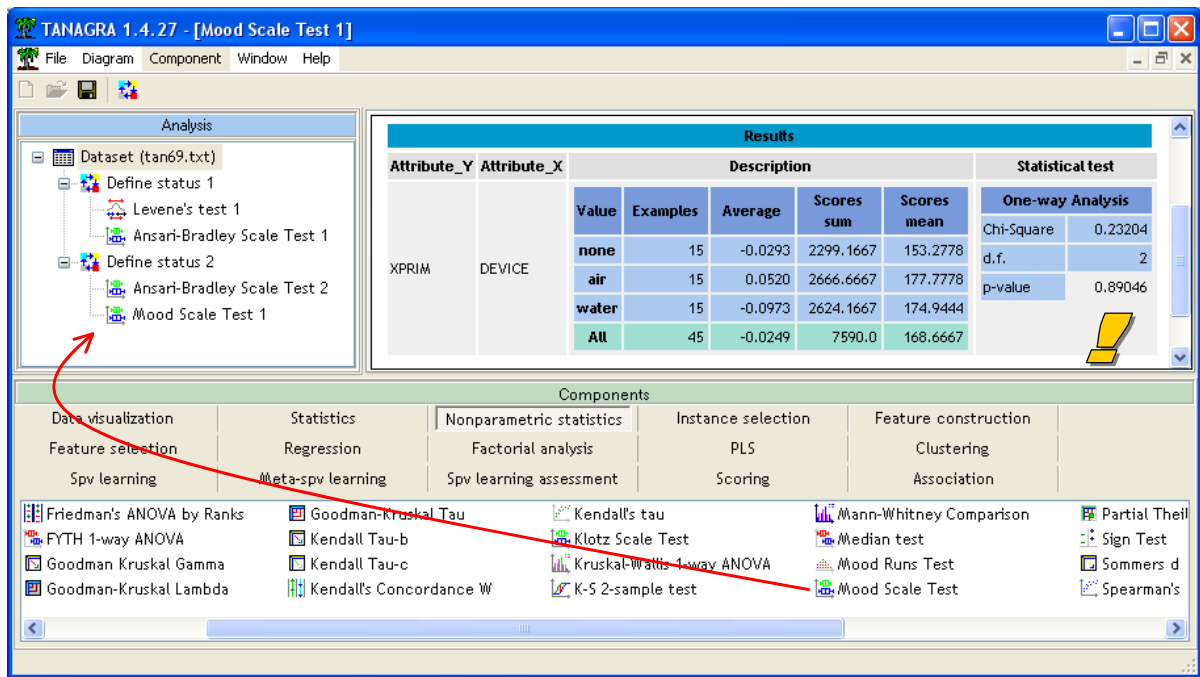
Like the Ansari-Bradley test, the Mood test relies on the ranks^{7,8}. They supply very similar results.

We insert the MOOD SCALE TEST component (NONPARAMETRIC STATISTICS tab) into the diagram. We obtain the following results.

We note that the conclusion is very similar to that of Ansari-Bradley test.

⁷ See <http://v8doc.sas.com/sashtml/stat/chap47/sect17.htm>

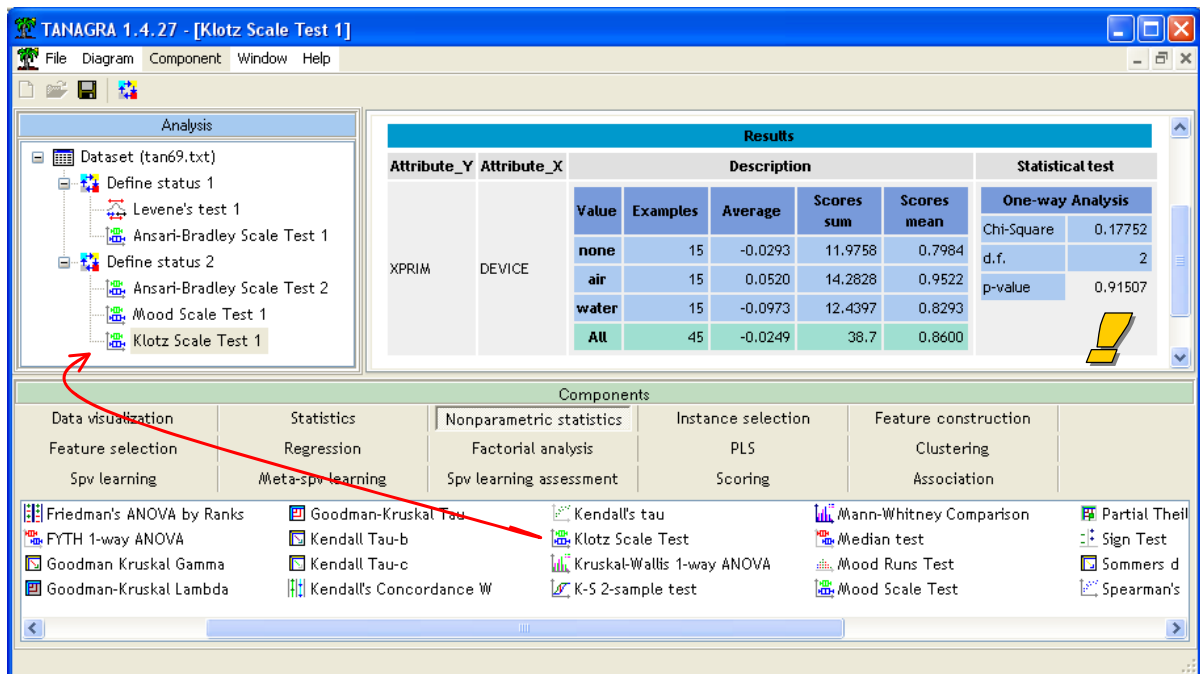
⁸ See also <http://www.technion.ac.il/docs/sas/analyst/chap10/sect3.htm>



4.5 Klotz test

The Klotz test is another version of the test for difference in scales based on the ranks. It is more convenient if the underlying distribution of the dataset is nearly Gaussian.

We insert the KOTZ SCALE TEST into the diagram. We obtain the same conclusion as the others.



5 Graphical plots

As often in statistics, we could dispense with all these complicated calculations by performing a graphical plot chosen wisely. In our case, the box plot⁹ gives a very relevant visualization of the configuration that we are studying. Clearly, the distributions are shifted but with the same dispersion (Figure 1).

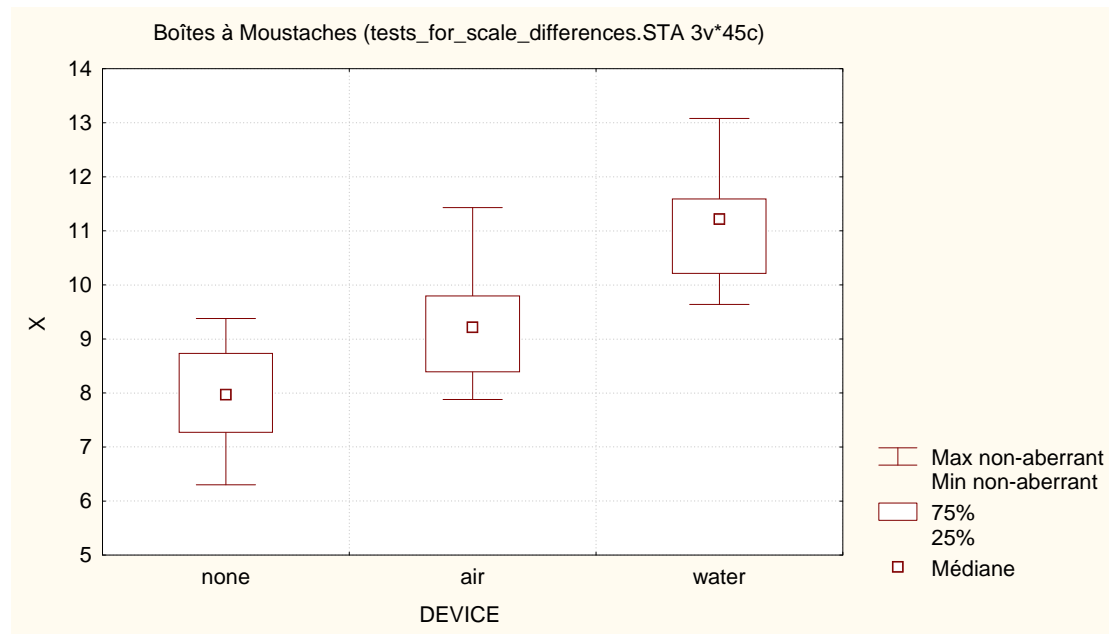


Figure 1 – Boxplot of X according to DEVICE

⁹ http://en.wikipedia.org/wiki/Box_plot