

1 Objectif

Tests paramétriques et non paramétriques de comparaison des dispersions pour K échantillons indépendants.

Les tests de comparaison de variances sont souvent présentés comme des préalables aux tests de comparaisons de moyennes, pour s'assurer de l'hypothèse d'homoscédasticité. Mais ce n'est pas leur seule finalité. Comparer les dispersions peut être une fin en soi.

Les tests paramétriques reposent principalement sur la normalité des données. Nous mettons en avant le test de Levene dans ce didacticiel. D'autres tests existent, nous les signalerons dans le texte.

Lorsque l'hypothèse de normalité est battue en brèche, lorsque les effectifs sont faibles, lorsque la variable est plus ordinale que continue, on a intérêt à passer aux tests non paramétriques. On parle alors de comparaison d'échelles ou de dispersions. En effet les procédures ne reposent plus sur les variances estimées. Nous utiliserons dans ce didacticiel les techniques les plus connues tels que le test de Ansari-Bradley, le test de Mood ou le test de Klotz. Ils ont un champ d'application plus large puisque non paramétriques. Ils présentent en revanche un inconvénient fort, ils sont inapplicables dès que les caractéristiques de tendance centrale conditionnelles (on dira la médiane pour simplifier) sont différentes.

Nous montrons la mise en œuvre de ces différents tests dans TANAGRA. Nous inspecterons et confronterons les résultats. Nous essayerons d'apporter des solutions lorsque les conditions d'utilisation des tests ne sont pas respectées.

Les aspects théoriques relatifs à ce didacticiel sont décrits dans deux supports de cours accessibles en ligne http://eric.univ-lyon2.fr/~ricco/cours/cours/Comp_Pop_Tests_Nonparametriques.pdf et http://eric.univ-lyon2.fr/~ricco/cours/cours/Comp_Pop_Tests_Parametriques.pdf.

2 Données

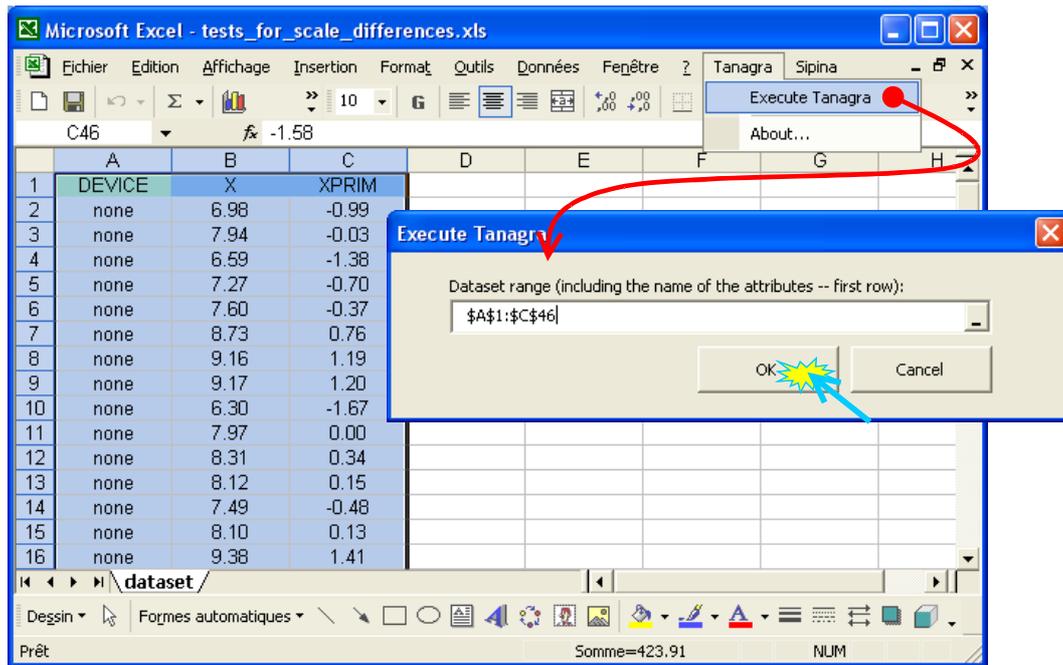
Les données décrivent les performances de dispositifs de chauffage écologique. La variable d'intérêt est la température à l'intérieur d'une cabane au petit matin de l'automne, au fin fond de la forêt, lorsque le loup n'est pas rentré dans sa tanière encore. Nous disposons de $n = 45$ observations. Un premier groupe de $n_1 = 15$ cabanes sert de témoin. Aucun système n'a été mis en place. Deux autres groupes ($n_2 = 15$ et $n_3 = 15$) sont constitués : l'un bénéficie d'un système basé sur un réchauffement naturel de l'air, l'autre s'appuie sur le réchauffement de l'eau. On cherche à comparer la disparité des températures d'un groupe à l'autre. Les données sont listées dans le fichier « **tests_for_scale_differences.xls** »¹.

3 Les tests paramétriques

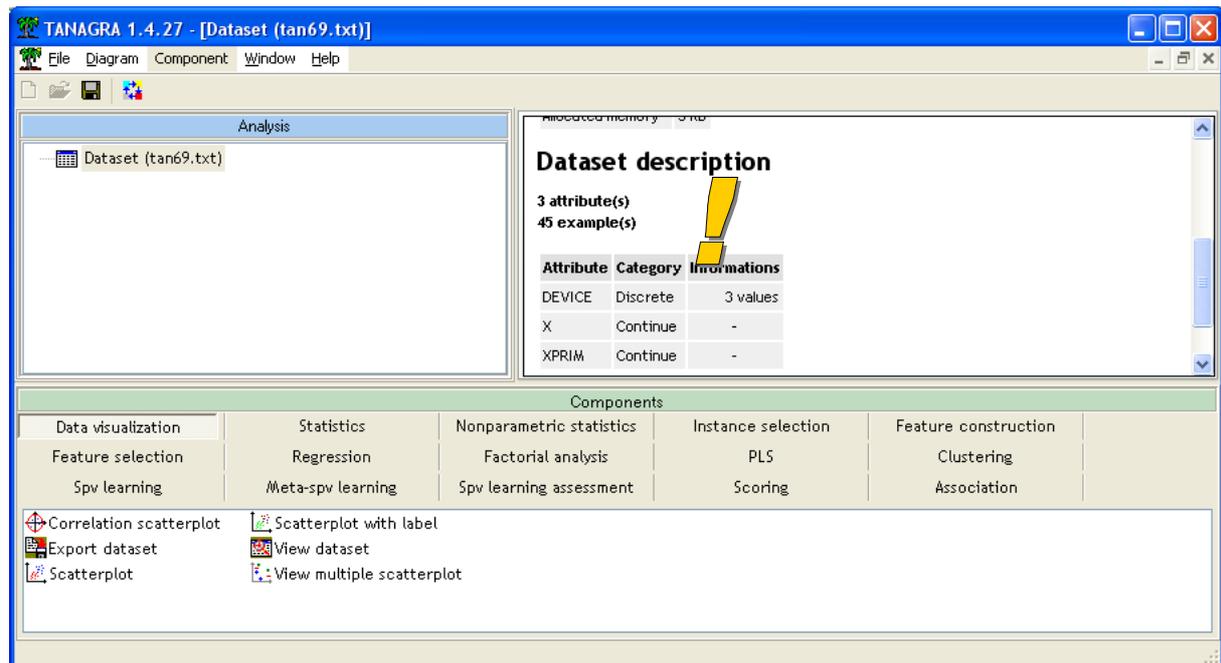
3.1 Importation des données

Le plus simple pour lancer Tanagra et charger les données est d'ouvrir le fichier XLS dans le tableur EXCEL. Nous sélectionnons la plage de données. La première ligne doit correspondre au nom des variables. Puis nous activons le menu TANAGRA / EXECUTE TANAGRA qui a été installé avec la macro complémentaire TANAGRA.XLA². Une boîte de dialogue apparaît. Nous vérifions la sélection. Si tout est en règle, nous validons en cliquant sur le bouton OK.

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



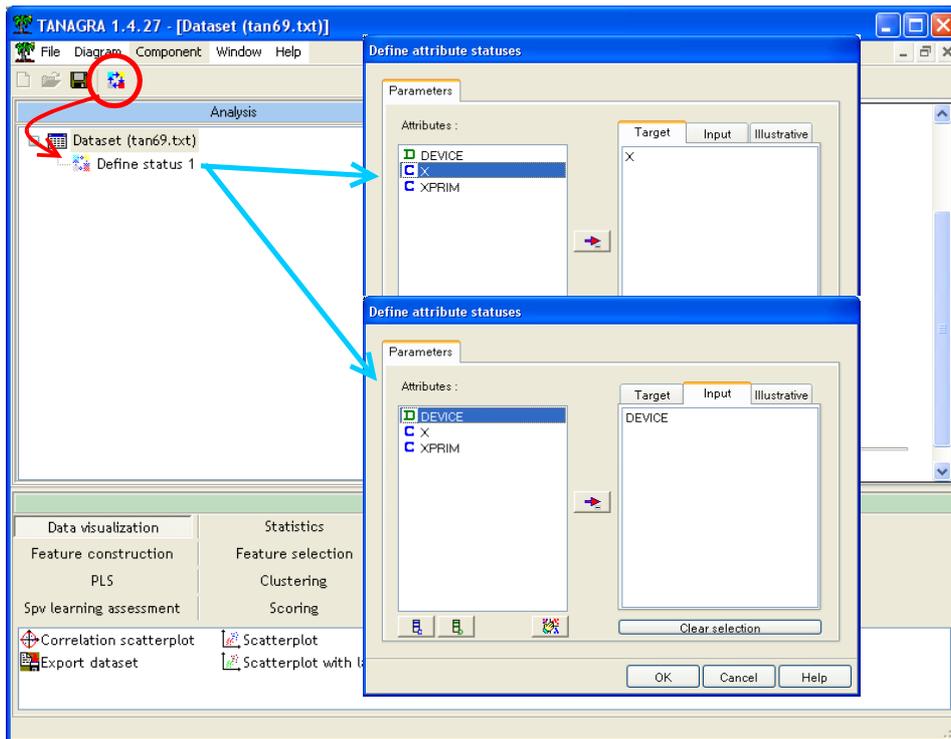
TANAGRA est automatiquement lancé. Un nouveau diagramme est créé. Nous devons disposer de 45 observations et 3 variables (GROUP, X et XPRIM). La variable d'intérêt est X. Nous expliciterons le rôle de XPRIM un peu plus loin.



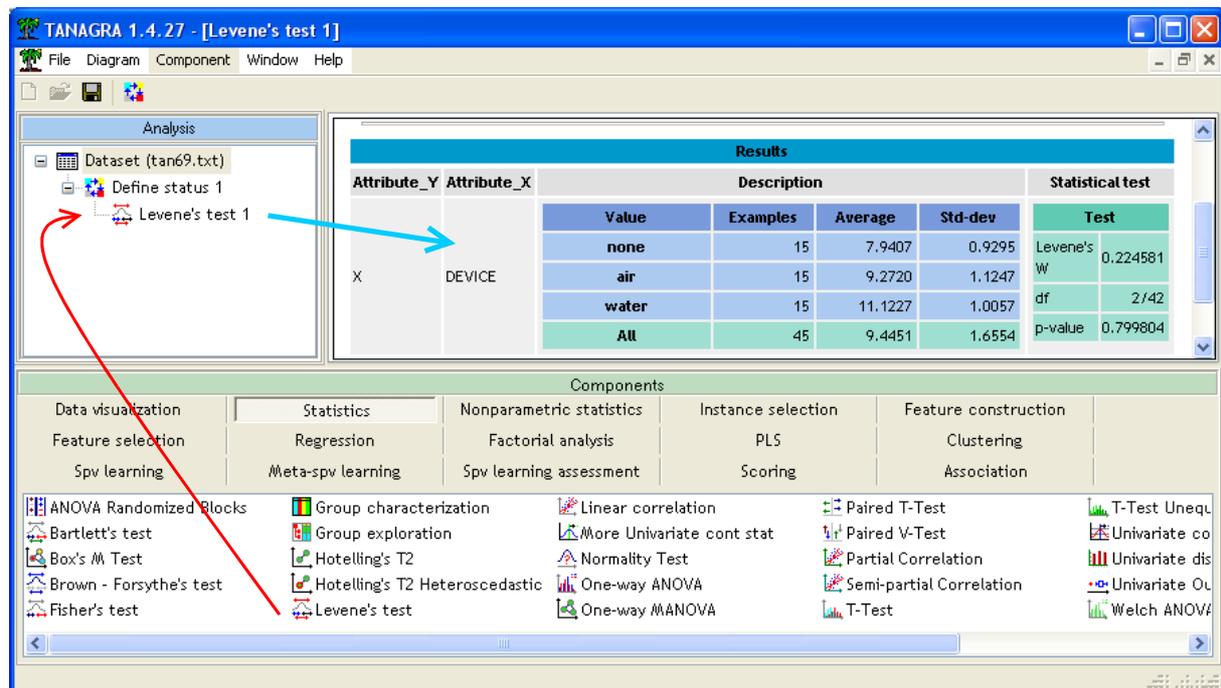
3.2 Test de Levene

Pour comparer les variances, nous utilisons le test de Levene. Nous insérons le composant DEFINE STATUS dans le diagramme, nous plaçons X en TARGET et GROUP en INPUT.

² Voir <http://tutoriels-data-mining.blogspot.com/2008/03/importation-fichier-xls-excel-macro.html> concernant l'installation et l'utilisation de la macro complémentaire TANAGRA.XLA.



Nous introduisons ensuite le composant LEVENE'S TEST (onglet STATISTICS). Nous cliquons sur le menu contextuel VIEW pour obtenir le résultat suivant.



Le test de Levene indique la compatibilité des données avec l'hypothèse d'égalité des variances dans les sous-groupes. Les écart-types conditionnels sont approximativement égaux à 1. Le résultat est sans ambiguïté, la probabilité critique du test est très élevée, $p = 0.799804$.

Remarque : Le test de Levene n'est pas le plus connu. On fait bien plus volontiers référence au test de Bartlett (composant BARTLETT'S TEST) dans la littérature. C'est une erreur. Ce dernier est très sensible à un écart par rapport à la normalité des données. Il est très peu robuste. On ne doit l'utiliser que si la conformité à la loi normale est assurée.

Le test de Brown-Forsythe (composant BROWN-FORSYTHE'S TEST) est une variante du test de Levene. Son champ d'application est plus vaste quant à la distribution sous-jacente des données³.

4 Tests non paramétriques

4.1 Inadéquation des tests non paramétriques usuels

Le test paramétrique conclut à l'hypothèse d'égalité des écart-types conditionnels. Voyons ce qu'il en est avec une procédure non paramétrique, le test de Ansari-Bradley.

Nous introduisons le composant ANSARI-BRADLEY SCALE TEST dans le diagramme. Nous cliquons sur le menu VIEW pour accéder aux résultats.

The screenshot shows the TANAGRA 1.4.27 interface. The main window displays the results of an 'Ansari-Bradley Scale Test 1'. The results are presented in a table with columns for 'Attribute_Y', 'Attribute_X', 'Description', and 'Statistical test'. The 'Description' table has columns for 'Value', 'Examples', 'Average', 'Scores sum', and 'Scores mean'. The 'Statistical test' table has columns for 'One-way Analysis', 'Chi-Square', 'd.f.', and 'p-value'.

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 palette at the bottom lists various statistical tests, including 'Ansari-Bradley Scale Test', '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'.

Patratras, la conclusion est tout autre. Il semble finalement que les dispersions soient différentes dans les sous-groupes au risque 5%⁴. A quelle procédure doit-on accorder crédit, quelle est la bonne solution ?

Les écart-types conditionnels quasi égaux semblent militer en faveur du test de Levene. Mais méfiance. Un écart-type n'est qu'un résumé statistique, il peut masquer des situations très disparates. Il se peut par exemple que certaines distributions conditionnelles soient bimodales, ou très asymétrique, etc.

³ Pour les références théoriques, outre les supports cités en introduction, voir <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm> ; concernant d'autres tutoriels relatif au test de comparaison de variances sous TANAGRA, voir <http://tutoriels-data-mining.blogspot.com/2008/04/analyse-de-variance-et-comparaison-de.html> et <http://tutoriels-data-mining.blogspot.com/2008/07/comparaison-de-populations-tests.html>

⁴ Les autres tests non paramétriques auraient donné des résultats similaires.

Concernant les tests non paramétriques, on sait qu'ils sont sensibles à une disparité des paramètres de localisation (de tendance centrale). Or on constate dans la copie d'écran ci-dessus que les moyennes conditionnelles sont différentes d'un groupe à l'autre, avec : $\bar{x}_1 = 7.94$; $\bar{x}_2 = 9.27$; $\bar{x}_3 = 11.12$. Le test de Ansari-Bradley indique bien des différences entre les sous-groupes, mais elles ne sont pas forcément dues à des disparités de dispersion. Il faut transformer les données pour dépasser cet écueil.

4.2 Transformation des données

La stratégie la plus simple consiste à centrer les données autour des médianes conditionnelles. Pour la valeur x_{ik} , correspondant à une observation n°i du groupe k, nous calculons

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

où \tilde{x}_k représente la médiane dans le groupe k.

Prenons le premier groupe (DEVICE = NONE), la médiane est $\tilde{x}_1 = 7.97$. Nous en déduisons ($x'_{11} = 6.98 - 7.97 = -0.99$; $x'_{21} = 7.94 - 7.97 = -0.03$; $x'_{31} = 6.59 - 7.97 = -1.38$; etc.)

Nous faisons de même pour les autres groupes. La variable XPRIM de notre ensemble de données a été élaborée ainsi.

4.3 Test de Ansari-Bradley

Nous insérons un nouveau composant DEFINE STATUS dans le diagramme. Nous plaçons XPRIM en TARGET, DEVICE en INPUT.

Statistical test	
One-way Analysis	
Scores mean	Chi-Square
10.5333	6.35786
15.2000	d.f.
9.5333	2
11.7556	p-value
	0.04163

Nous introduisons de nouveau le composant ANSARI-BRADLEY SCALE TEST (onglet NONPARAMETRIC STATISTICS).

The screenshot shows the TANAGRA 1.4.27 software interface. The main window displays the results of an 'Ansari-Bradley Scale Test 2'. The 'Results' pane contains a table with the following data:

Attribute_Y	Attribute_X	Description				Statistical test	
		Value	Examples	Average	Scores sum	Scores mean	
		none	15	-0.0293	189.8333	12.6556	
		air	15	0.0520	172.3333	11.4889	
		water	15	-0.0973	166.8333	11.1222	
		All	45	-0.0249	529.0	11.7556	

The 'Statistical test' table shows:

One-way Analysis	
Chi-Square	0.44584
d.f.	2
p-value	0.80018

The 'Components' pane at the bottom lists various statistical tests, including 'Ansari-Bradley Scale Test', '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'.

Le résultat est tout autre. Il est parfaitement en accord avec le test de Levene cette fois-ci. La statistique de test est $\chi^2 = 0.44584$, elle suit une loi du KHI-DEUX à 2 degrés de liberté, la probabilité critique est $p = 0.80018$.

La conclusion précédente, obtenue sans précautions sur la variable X, était manifestement entachée par les écarts entre les paramètres de localisation des groupes. La transformation opérée sur les données est essentielle dans cet exemple.

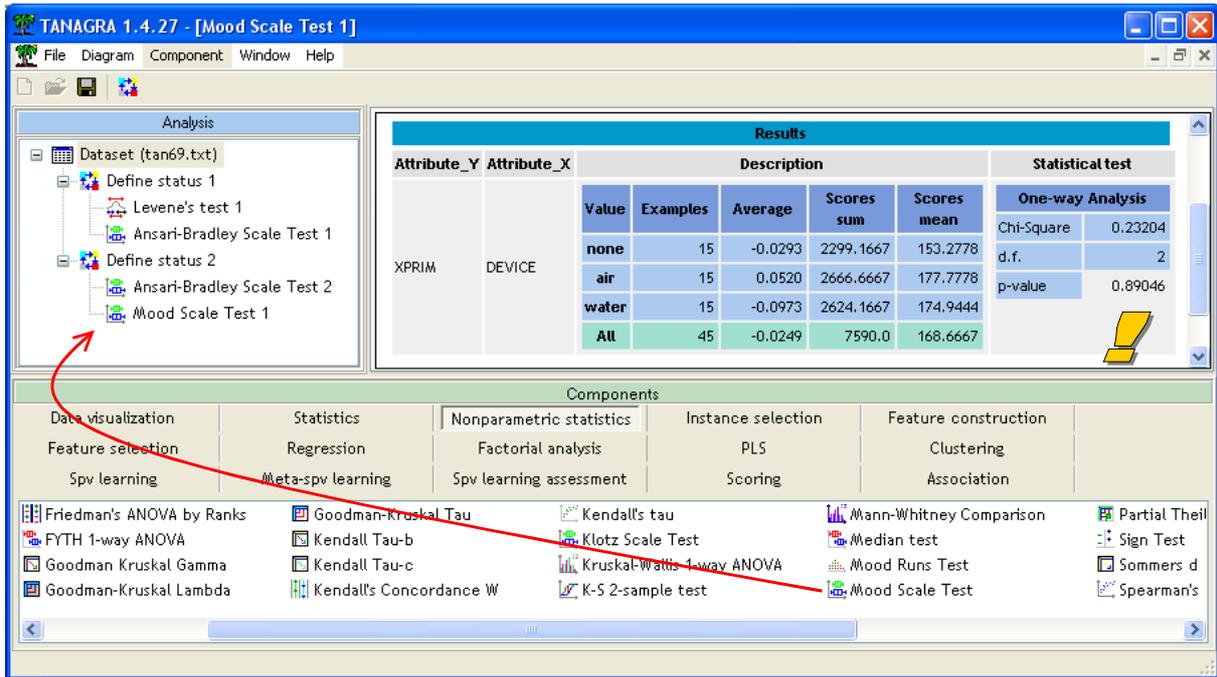
Remarque : Le centrage des données à l'aide de la médiane empirique ne pose pas de problèmes sur les effectifs assez importants, la distribution asymptotique de la statistique de test est conservée. C'est moins sûr lorsque l'on veut utiliser la distribution exacte sur les tous petits effectifs. Voir Capéraà et Van Cutsem, « Méthodes et modèles en statistique non paramétrique », PUL et DUNOD, 1988 ; page 179.

4.4 Test de Mood

Tout comme le test de Ansari-Bradley, le test de Mood repose sur une statistique de rang linéaire⁵. Ils produisent des résultats très similaires.

Nous insérons le composant MOOD SCALE TEST (onglet NONPARAMETRIC STATISTICS) dans le diagramme.

⁵ Voir <http://v8doc.sas.com/sashtml/stat/chap47/sect17.htm>

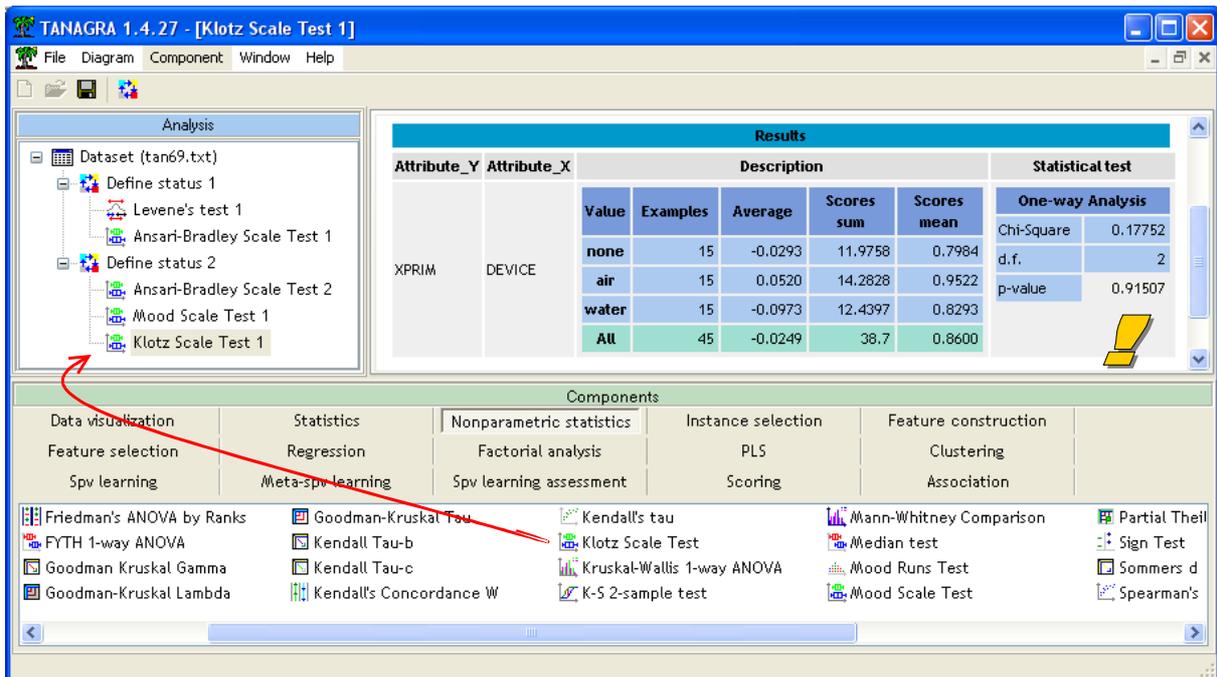


La conclusion est la même. L'hypothèse d'égalité des dispersions n'est pas remise en cause par les données.

4.5 Test de Klotz

Le test de Klotz est un autre avatar basé sur les statistiques linéaires des rangs. On le préférera si la distribution sous jacente des données est proche de la loi normale.

Nous insérons le composant KOTZ SCALE TEST pour la même conclusion, plus marquée encore avec un probabilité critique $p = 0.91507$.



5 Procédure graphique

Comme souvent en statistique, on aurait pu se dispenser de tous ces calculs compliqués en réalisant un graphique judicieusement choisi. Dans notre cas, les boîtes à moustaches (*box plot*) nous auraient très largement renseigné sur la configuration que nous sommes en train d'étudier. A l'évidence, les distributions sont décalées mais de dispersion égale (Figure 1).

Les procédures graphiques ne sont pas extrêmement développées encore dans TANAGRA (version 1.4.27). J'espère que cela évoluera dans un avenir proche.

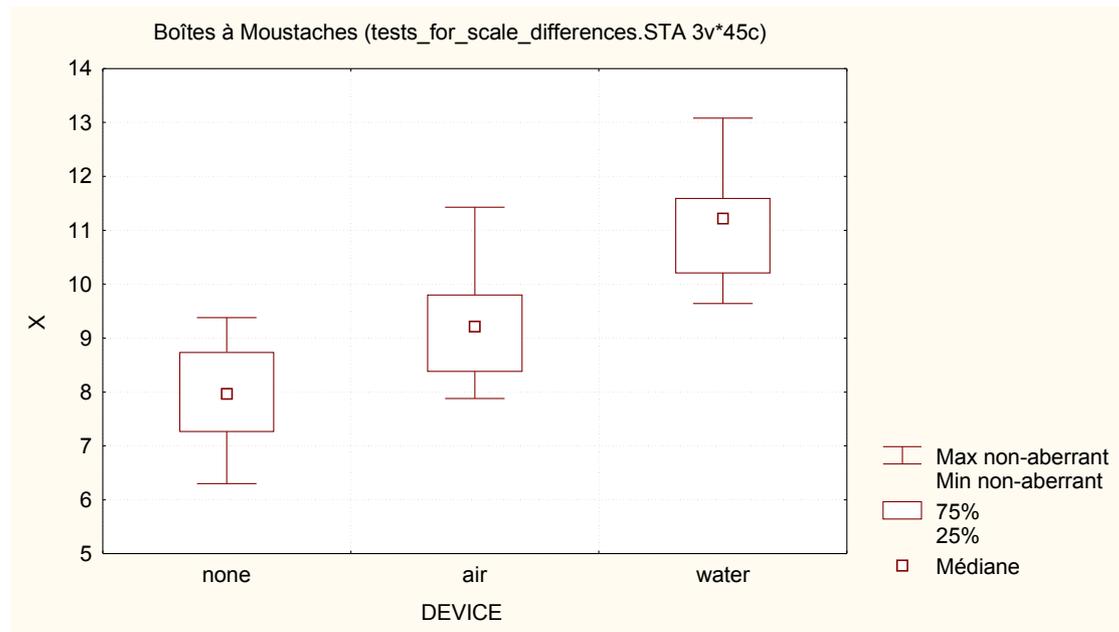


Figure 1 - Boîtes à moustaches de X conditionnellement à DEVICE