

Tests non-paramétriques

Frédéric Bertrand¹ & Myriam Maumy¹

¹IRMA, Université de Strasbourg
Strasbourg, France

Master 1^{re} Année
2015-2016

Sommaire

- 1 Contexte du test
 - Conditions d'application
- 2 Absence d'ex æquo dans les observations
 - Statistique de test
 - Règle de décision et conclusion du test
- 3 Présence d'ex æquo dans les observations
- 4 Comparaisons multiples
- 5 Application

2^{ème} partie

Tests non paramétriques : Le test de Kruskal-Wallis

Introduction

Nous observons, de manière indépendante, une variable aléatoire X de loi continue, sur $k \geq 3$ populations, ou sur une population divisée en $k \geq 3$ sous-populations.

Nous supposons ainsi que nous disposons de k échantillons aléatoires **indépendants** $(X_{1,1}, \dots, X_{1,n_1}), \dots, (X_{k,1}, \dots, X_{k,n_k})$ et de $k \geq 3$ séries d'observations $(x_{1,1}, \dots, x_{1,n_1})$ pour la première, $\dots, (x_{k,1}, \dots, x_{k,n_k})$ pour la dernière. Nous notons $\mathcal{L}_i(X)$ la loi de la variable aléatoire X sur la (sous-)population d'ordre i avec $1 \leq i \leq k$.

Sans faire d'hypothèses spécifiques, le test de Kruskal-Wallis ne permet pas de tester l'égalité des moyennes ni celle des médianes.

Sommaire

Hypothèses du test

Le test de **Kruskal-Wallis** est utilisé pour tester les hypothèses suivantes :

$$\mathcal{H}_0 : \mathcal{L}_1(X) = \dots = \mathcal{L}_j(X) = \dots = \mathcal{L}_k(X)$$

contre

\mathcal{H}_1 : Les lois $\mathcal{L}_1(X), \dots, \mathcal{L}_k(X)$ ne sont pas toutes identiques.

- 1 Contexte du test
 - Conditions d'application
- 2 Absence d'ex æquo dans les observations
 - Statistique de test
 - Règle de décision et conclusion du test
- 3 Présence d'ex æquo dans les observations
- 4 Comparaisons multiples
- 5 Application

Statistique de test

Calculons :

- 1 le rang $R_{i,j}$ de $X_{i,j}$ parmi les n_{\bullet} valeurs ;
- 2 puis la somme des rangs associée à chaque échantillon :

$$R_{i,\bullet} = \sum_{j=1}^{n_i} R_{i,j};$$

- 3 la moyenne des rangs de chaque échantillon :

$$\overline{R}_{i,\bullet} = R_{i,\bullet} / n_i.$$

Statistique de test (suite)

La statistique de Kruskal-Wallis $KW_{n_{\bullet}}$ prend en compte l'écart entre la moyenne des rangs de chaque échantillon et la moyenne de tous les rangs, qui vaut $(n_{\bullet} + 1)/2$:

$$\begin{aligned} KW_{n_{\bullet}} &= \frac{12}{n_{\bullet}(n_{\bullet} + 1)} \sum_{i=1}^k n_i \left(\overline{R}_{i,\bullet} - \frac{n_{\bullet} + 1}{2} \right)^2 \\ &= \frac{12}{n_{\bullet}(n_{\bullet} + 1)} \sum_{i=1}^k \frac{R_{i,\bullet}^2}{n_i} - 3(n_{\bullet} + 1). \end{aligned}$$

Propriétés

Lorsque l'hypothèse nulle \mathcal{H}_0 est vraie, la variable aléatoire KW_{n_\bullet} a les trois propriétés suivantes.

- 1 Pour $i = 1, \dots, k$, $W_i = n_i \overline{R_{i\bullet}}$ est la statistique de Wilcoxon qui compare le i -ème traitement aux $k - 1$ autres traitements.

Sous l'hypothèse nulle \mathcal{H}_0 , nous en déduisons que

$$\mathbb{E}(W_i) = n_i(n_\bullet + 1)/2 \text{ et } \text{Var}(W_i) = n_i(n_\bullet - n_i)(n_\bullet + 1)/12.$$

Ainsi par conséquent, nous avons :

$$KW_{n_\bullet} = \frac{1}{n_\bullet} \sum_{i=1}^k (n_\bullet - n_i) \frac{(W_i - \mathbb{E}(W_i))^2}{\text{Var}(W_i)}.$$

Propriétés (suite)

- 2 **(Suite)** Nous calculons alors l'espérance et la variance de KW_{n_\bullet} sous l'hypothèse nulle \mathcal{H}_0 :

$$\mathbb{E}(KW_{n_\bullet}) = k - 1,$$

$$\text{Var}(KW_{n_\bullet}) = \frac{2 [3k^2 - 6k + n_\bullet(2k^2 - 6k + 1)]}{5n_\bullet(n_\bullet + 1)}$$

$$= \frac{6}{-5} \sum_{i=1}^k \frac{1}{n_i}$$

- 3 Il est possible de déterminer la distribution de KW_{n_\bullet} bien que le calcul soit complexe. Elle est donc tabulée pour les faibles valeurs des n_i .

Règle de décision et conclusion du test

- **Premier cas** : L'un des effectifs n_i , $1 \leq i \leq k$, est inférieur ou égal à 4, . Pour un seuil donné α , des tables de la loi de Kruskal-Wallis nous fournissent une valeur critique c_α . Alors nous décidons :

$$\begin{cases} \text{si } KW_{n_\bullet}(\text{obs}) \geq c_\alpha & \mathcal{H}_1 \text{ est vraie,} \\ \text{si } KW_{n_\bullet}(\text{obs}) < c_\alpha & \mathcal{H}_0 \text{ est vraie.} \end{cases}$$

Le niveau de signification réel du test est généralement strictement inférieur à α .

Règle de décision et conclusion du test (suite)

- **Second cas** : Si $n_i \geq 5$, pour tout $1 \leq i \leq k$, nous utilisons l'approximation $KW_{n_\bullet} \approx \chi^2(k - 1)$. Pour un seuil donné α , des tables de la loi du χ^2 nous fournissent une valeur critique c_α telle que $\mathbb{P}_{\mathcal{H}_0}(-c_\alpha < Z_{n_1, n_2} < c_\alpha) = 1 - \alpha$. Alors nous décidons :

$$\begin{cases} \text{si } KW_{n_\bullet}(\text{obs}) \geq c_\alpha & \mathcal{H}_1 \text{ est vraie,} \\ \text{si } KW_{n_\bullet}(\text{obs}) < c_\alpha & \mathcal{H}_0 \text{ est vraie.} \end{cases}$$

Lorsque nous rejetons \mathcal{H}_0 , nous décidons que \mathcal{H}_1 est vraie avec un risque d'erreur de première espèce α .

Lorsque nous conservons \mathcal{H}_0 , c'est avec un risque d'erreur de deuxième espèce β .

Sommaire

- 1 Contexte du test
 - Conditions d'application
- 2 Absence d'ex æquo dans les observations
 - Statistique de test
 - Règle de décision et conclusion du test
- 3 **Présence d'ex æquo dans les observations**
- 4 Comparaisons multiples
- 5 Application

Sommaire

- 1 Contexte du test
 - Conditions d'application
- 2 Absence d'ex æquo dans les observations
 - Statistique de test
 - Règle de décision et conclusion du test
- 3 Présence d'ex æquo dans les observations
- 4 **Comparaisons multiples**
- 5 Application

Présence d'ex æquo dans les observations : la méthode des rangs moyens

À chaque nombre appartenant à un groupe d'ex æquo nous attribuons le rang moyen du groupe auquel il appartient puis nous déterminons la somme $T = \sum_{j=1}^h (t_j^3 - t_j)$ où t_j désigne le nombre d'éléments du j -ème groupe d'ex æquo. Il est d'usage de substituer à KW_n , la variable KW_n^* , définie par :

$$KW_n^* = \frac{KW_n}{1 - \frac{n_0^3 - n}{n_0^3 - n}}$$

Test de Steel-Dwass-Critchlow-Fligner

$W_{n_i, n_{i'}}$ est la statistique de Wilcoxon qui compare le i -ème traitement au i' -ème traitement.

Les observations des deux groupes i et i' sont ordonnées puis regroupées en h classes d'ex æquo C_j , $1 \leq j \leq h$.

Notons d_j le nombre d'ex æquo de la classe C_j et $m_{i, i'} = n_i + n_{i'}$.

Test de Steel-Dwass-Critchlow-Fligner (suite)

Nous décidons qu'**au seuil global** α deux lois $\mathcal{L}_i(X)$ et $\mathcal{L}_{j'}(X)$, parmi les $k(k-1)$ comparaisons que nous allons faire, sont significativement différentes si :

$$\left| W_{n_i, n_{j'}} - \frac{n_i(m_{i,j'}) + 1}{2} \right| \geq d'(k; +\infty; 1 - \alpha) \sqrt{\frac{n_i n_{j'} (m_{i,j'} + 1)}{24}} \times \sqrt{\left(1 - \frac{\sum_{j=1}^h (d_j^3 - d_j)}{m_{i,j'}^3 - m_{i,j'}} \right)}$$

où $d'(k; +\infty; 1 - \alpha)$ est le quantile d'ordre $1 - \alpha$ pour la loi du maximum du module studentisé pour k moyennes et $+\infty$ degrés de liberté.

Test de Steel-Dwass-Critchlow-Fligner (suite)

Contrairement aux trois autres approches présentées ci-après, le **test Steel-Dwass-Critchlow-Fligner** n'est pas qu'une procédure de comparaisons multiples : c'est une **alternative complète** au test de **Kruskal-Wallis**.

Tests non-paramétriques

Tests non-paramétriques

Autres tests de comparaisons multiples

Si nous rejetons l'hypothèse nulle \mathcal{H}_{l_0} , nous nous demandons quelles sont les lois $\mathcal{L}_i(X)$ qui diffèrent.

Les formules ci-après sont valables en absence ou en présence d'ex æquo. En absence d'ex æquo, le terme $1 - T/(n_{\bullet}^3 - n_{\bullet})$ est égal à 1.

Application de la méthode de Sheffé

Nous décidons qu'**au seuil** α **deux lois** $\mathcal{L}_i(X)$ et $\mathcal{L}_{j'}(X)$ sont significativement différentes si :

$$\left| \overline{R_{i,\bullet}} - \overline{R_{j',\bullet}} \right| \geq \sqrt{\chi^2(k-1; 1-\alpha)} \sqrt{\frac{n_{\bullet}(n_{\bullet}+1)}{12} \left(1 - \frac{T}{n_{\bullet}^3 - n_{\bullet}} \right)} \times \sqrt{\frac{1}{n_i} + \frac{1}{n_{j'}}},$$

où $\chi^2(k-1; 1-\alpha)$ est le quantile d'ordre $1 - \alpha$ pour la loi du χ^2 à $k-1$ degrés de liberté.

Application de l'inégalité de Bonferroni

Nous décidons qu'**au seuil global** α deux lois $\mathcal{L}_i(X)$ et $\mathcal{L}_{j'}(X)$, parmi les $k(k-1)$ **comparaisons** que nous allons faire, sont significativement différentes si :

$$|\overline{R_{i,\bullet}} - \overline{R_{j',\bullet}}| \geq u \left(1 - \frac{\alpha}{k(k-1)} \right) \sqrt{\frac{n_{\bullet}(n_{\bullet}+1)}{12} \left(1 - \frac{T}{n_{\bullet}^2} \right)} \\ \times \sqrt{\frac{1}{n_i} + \frac{1}{n_{j'}}},$$

où $u(1 - \alpha/(k(k-1)))$ est le quantile d'ordre $1 - \alpha/(k(k-1))$ pour la loi normale centrée-réduite.

Remarque

Il s'agit d'une application des inégalités de Bonferroni. Cette procédure est plus puissante que la précédente.
Il est également possible d'utiliser une approche séquentielle comme la procédure de Holm-Bonferroni (test de Dunn) ou de Holm-Sidak.

Tests non-paramétriques

Analogue de la méthode de Tukey-Kramer

Nous décidons qu'**au seuil global** α deux lois $\mathcal{L}_i(X)$ et $\mathcal{L}_{j'}(X)$, parmi les $k(k-1)$ **comparaisons** que nous allons faire, sont significativement différentes si :

$$|\overline{R_{i,\bullet}} - \overline{R_{j',\bullet}}| \geq q(k; +\infty; 1 - \alpha) \sqrt{\frac{n_{\bullet}(n_{\bullet}+1)}{12} \left(1 - \frac{T}{n_{\bullet}^2} \right)} \\ \times \sqrt{\frac{1}{2} \left(\frac{1}{n_i} + \frac{1}{n_{j'}} \right)},$$

où $q(k; +\infty; 1 - \alpha)$ est le quantile d'ordre $1 - \alpha$ pour la loi de l'étendue studentisée pour k moyennes et $+\infty$ degrés de liberté.

Tests non-paramétriques

Remarque

Il s'agit d'une procédure analogue à celle de Tukey-Kramer dans le cas paramétrique et valide asymptotiquement. Elle est généralement plus puissante que les deux approches précédentes.

Sommaire

- 1 Contexte du test
 - Conditions d'application
- 2 Absence d'ex æquo dans les observations
 - Statistique de test
 - Règle de décision et conclusion du test
- 3 Présence d'ex æquo dans les observations
- 4 Comparaisons multiples
- 5 Application

Application

Des forestiers ont réalisé des plantations d'arbres en trois endroits.

Plusieurs années plus tard, ils souhaitent savoir si la hauteur moyenne des arbres est identique dans les trois forêts. Chacune des forêts constitue une population et dans chacune d'entre elles, un échantillon d'arbres est tiré au sort. Puis la hauteur de chaque arbre est mesurée en mètres. Le tableau ci-après donne les valeurs des mesures de hauteur ainsi que les moyennes et les variances corrigées estimées dans chacun des groupes.

Données

	Forêt 1	Forêt 2	Forêt 3
Données expérimentales (en mètres)	23,4	18,9	22,5
	24,4	21,1	22,9
	24,6	21,1	23,7
	24,9	22,1	24,0
	25,0	22,5	24,0
	26,2	23,5	24,5
Moyennes	24,750	21,533	23,600
Variances corrigées	0,831	2,487	0,568

La Figure 1 correspond aux boîtes à moustaches de la hauteur en fonction des trois forêts.

Représentation graphique

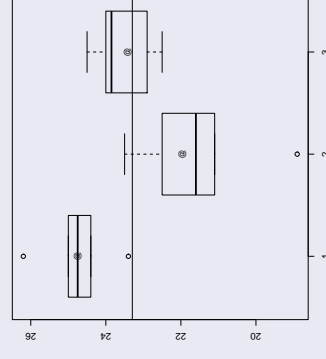


FIGURE: Boîtes à moustaches (à gauche) – Comparaisons multiples (à droite)

Statistique du test

Nous sommes en présence d'ex æquo, nous devons donc utiliser la statistique de test $KW_{n_s}^*$ à la place de la statistique de test KW_{n_s} .

Nous calculons la valeur de $KW_{n_s}^*$ sur l'échantillon :
 $KW_{n_s}^*(obs) = 11,51$.



Règle de décision à l'aide d'une valeur critique

Pour un seuil $\alpha = 5\%$ la valeur critique d'un Khi-deux à 2 degrés de liberté, est $c_{0,05} = 5,99$. Comme $KW_{n_s}^*(obs) \leq c_{0,05}$, nous décidons de rejeter l'hypothèse nulle \mathcal{H}_0 , et que l'hypothèse alternative \mathcal{H}_1 est vraie. Il y a une influence significative, au seuil $\alpha = 5\%$, de la forêt sur la hauteur des arbres. Le risque associé à cette décision est un risque de première espèce qui vaut $\alpha = 5\%$.

Remarque

La valeur de $KW_{n_s}(obs)$ est égale à 11,48. Nous remarquons la différence apportée par la correction pour prendre en compte les ex æquo.



Comparaisons multiples avec Minitab

Règle de décision à l'aide d'une p -valeur

En utilisant un logiciel de statistique nous calculons la p -valeur du test de Kruskal-Wallis. Il faut bien vérifier qu'elle tient compte des ex æquo. Elle vaut dans cas 0,003.

Comme la p -valeur est $\leq 0,05$, nous décidons, au seuil $\alpha = 5\%$, de rejeter l'hypothèse nulle \mathcal{H}_0 , et que l'hypothèse alternative \mathcal{H}_1 est vraie. Il y a une influence significative, au seuil $\alpha = 5\%$, de la forêt sur la hauteur des arbres. Le risque associé à cette décision est un risque de première espèce qui vaut $\alpha = 5\%$.



Données désempilées, comparaisons 2 à 2

```
MTB> %KrusMc c1-c3;  
SUBC> unstacked;  
SUBC> Fa1pha 0,05.
```



Résultats

Kruskal-Wallis: Multiple Comparisons

Test de Kruskal-Wallis sur the data

Group	N	Médiane	Rang moyen	Z
Forêt_1	6	24,75	14,5	2,81
Forêt_2	6	21,60	4,1	-3,04
Forêt_3	6	23,85	9,9	0,23
Global	18		9,5	

B H = 11,48 DL = 2 P = 0,003

H = 11,51 DL = 2 P = 0,003

(ajusté pour les nombres de même grandeur)

Frédéric Bertrand & Myriam Maumy

Tests non-paramétriques

Résultats

Kruskal-Wallis: All Pairwise Comparisons

Comparisons:	3
Ties:	3
Family Alpha:	0,05
Bonferroni Individual Alpha:	0,017
Bonferroni Z-value (2-sided):	2,394

Frédéric Bertrand & Myriam Maumy

Tests non-paramétriques

Résultats

Standardized Absolute Mean Rank Differences

|Rbar(i)-Rbar(j)| / Stdev

Rows: Group i = 1, ..., n
Columns: Group j = 1, ..., n

1. Table of Z-values

Forêt_1	0,00000	*	*
Forêt_2	3,37961	0,00000	*
Forêt_3	1,48703	1,89258	0

Frédéric Bertrand & Myriam Maumy

Tests non-paramétriques

Résultats

Adjusted for Ties in the Data

1. Table of Z-values			
Forêt_1	0,00000	*	*
Forêt_2	3,38486	0,00000	*
Forêt_3	1,48934	1,89552	0

2. Table of P-values

Forêt_1	1,00000	*	*
Forêt_2	0,00071	1,00000	*
Forêt_3	0,13640	0,05802	1

Frédéric Bertrand & Myriam Maumy

Tests non-paramétriques

Résultats

Sign Confidence Intervals controlled at a family error rate of 0,05

Desired Confidence: 90,951

Intervalle de confiance pour le test du signe de la médiane

Intervalle de confiance

N	Médiane	Confiance atteinte	Inférieur	Supérieur	Pos.
F_1 6	24,75	0,7813	24,40	25,00	2
F_2 6	24,10	0,9095	24,10	25,36	NLI
F_3 6	23,40	0,9688	23,40	26,20	1

Résultats

F_2 6	21,60	0,7813	21,10	22,50	2
F_3 6	23,85	0,9688	23,50	24,00	2
F_1 6	21,10	0,9095	20,44	22,80	NLI
F_2 6	23,85	0,7813	22,90	24,00	2
F_3 6	23,85	0,9688	22,78	24,15	NLI
F_1 6	21,10	0,7813	21,10	22,50	2

Résultats

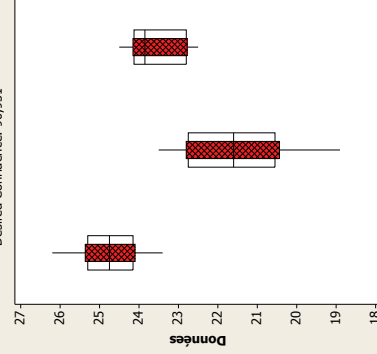
Kruskal-Wallis: Conclusions

The following groups showed significant differences (adjusted for ties):

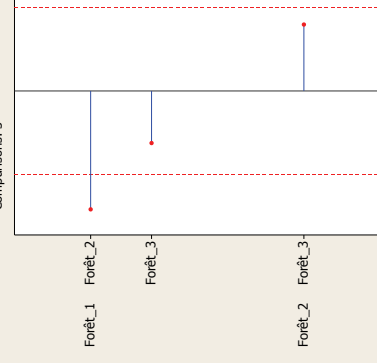
Groups Z vs. Critical value P-value
Forêt_1 vs. Forêt_2 3,38486 >= 2,394 0,0007

Multiple Comparisons Chart

Boxplots with Sign Confidence Intervals
Desired Confidence: 90,951



Pairwise Comparisons
Comparaisons: 3



Comparaisons multiples avec Minitab

Données désempilées, comparaisons avec un contrôle

```
MTB> %KrusMc c1 c2;  
SUBC> Control c3;  
SUBC> unstacked;  
SUBC> FaIpha 0,05.
```



Tests non-paramétriques

Frédéric Bertrand & Myriam Maumy

Résultats

Kruskal-Wallis: Multiple Comparisons

Test de Kruskal-Wallis sur the data

Group	N	Médiane	moyen	Z
Forêt_1	6	24,75	14,5	2,81
Forêt_2	6	21,60	4,1	-3,04
Forêt_3	6	23,85	9,9	0,23
Global	18		9,5	

B H = 11,48 DL = 2 P = 0,003

H = 11,51 DL = 2 P = 0,003

(ajusté pour les nombres de même grandeur)



Tests non-paramétriques

Frédéric Bertrand & Myriam Maumy

Résultats

Kruskal-Wallis: Comparisons to a Control

```
-----  
Comparisons:                2  
Ties:                        3  
Family Alpha:               0,05  
Bonferroni Individual Alpha: 0,025  
Bonferroni Z-value (2-sided): 2,241  
-----
```



Tests non-paramétriques

Frédéric Bertrand & Myriam Maumy

Résultats

Standardized Absolute Mean Rank Differences

|Rbar(i) - Rbar(Control)| / Stdev

Rows: Group i = 1, ..., n-1
Column: Control Group: Forêt_3

Z-value

Forêt_1 1,48703

Forêt_2 1,89258



Tests non-paramétriques

Frédéric Bertrand & Myriam Maumy

Résultats

Adjusted for Ties in the Data

Z-value	vs.	Critical value	P-value
Forêt_1	1,48934	< 2,24140	0,1364
Forêt_2	1,89552	< 2,24140	0,0580

Résultats

Sign Confidence Intervals controlled at a family error rate of 0,05

Desired Confidence: 88,701

Intervalle de confiance pour le test du signe de la médiane

F_1	N	Médiane	Confiance atteinte	Intervalle de confiance		Pos.
				Inférieur	Supérieur	
6	24,7	0,7813	0,7813	24,40	25,00	2
		0,8870	0,8870	24,19	25,25	NLI
		0,9688	0,9688	23,40	26,20	1

Frédéric Bertrand & Myriam Maumy

Tests non-paramétriques

Frédéric Bertrand & Myriam Maumy

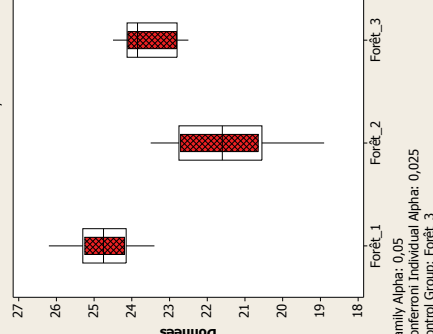
Tests non-paramétriques

Résultats

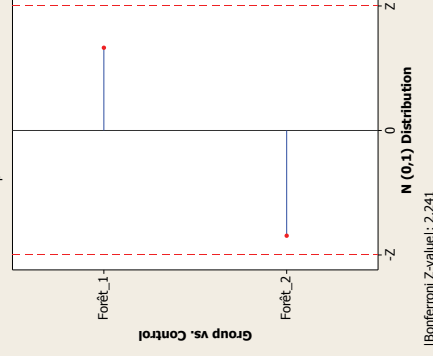
F_2	6	21,60	0,7813	21,10	22,50	2
		0,8870	0,8870	20,65	22,71	NLI
		0,9688	0,9688	18,90	23,50	1
F_3	6	23,85	0,7813	22,90	24,00	2
		0,8870	0,8870	22,82	24,10	NLI
		0,9688	0,9688	22,50	24,50	1

Multiple Comparisons Chart

Boxplots with Sign Confidence Intervals
Desired Confidence: 88,701



Comparisons to a Control
Comparisons: 2



Comparaisons multiples avec Minitab

Données empilées, comparaisons 2 à 2

Supposons que la colonne C1 contienne le facteur explicatif et la colonne C2 les valeurs observées de la réponse.

```
MTB> %KrusMc c1 c2;  
SUBC> Fa1pha 0,05.
```



Comparaisons multiples avec Minitab

Données empilées, comparaisons avec un contrôle

Supposons que la colonne C1 contienne le facteur explicatif et la colonne C2 les valeurs observées de la réponse. Attention pour faire des comparaisons avec un groupe contrôle, les observations pour ce groupe doivent être nécessairement dans une autre colonne, ici C3.

```
MTB> %KrusMc c1 c2;  
SUBC> Control c3;  
SUBC> Fa1pha 0,05.
```

