



FICHE MÉTHODOLOGIQUE

Série de Fiches Méthodologiques en Recherche et en Rédaction Scientifique

Fiche Méthodologique n°10

Tests non paramétriques sur SPSS pour comparer deux ou plusieurs moyennes sur des échantillons appariés (test de Wilcoxon et test de Friedman)

Non-parametric tests on SPSS to compare two or more means on matched samples

مقارنة متوسطين أو أكثر من عينات متطابقة بالإختبارات غير المعيارية على

A. Bezzaouha¹, A. Bouamra¹, A. Ammimer¹, A. Ben Abdelaziz², Réseau Maghrébin PRP2S³

1. Service d'Epidémiologie - CHU de Blida - Faculté de médecine de Blida - Université Saad Dahleb de Blida (Algérie)

2. Laboratoire de Recherche LR19SP01 « Mesure et Appui à la Performance des Etablissements de Santé ». CHU Sahloul Sousse. Université de Sousse (Tunisie)

3. Réseau Maghrébin Pédagogie Recherche Publication en Sciences de la Santé

Cette série...

Le Réseau Maghrébin PRP2S et la Rédaction de la revue « La Tunisie Médicale » ont l'honneur de vous présenter, régulièrement à partir du numéro de janvier 2020, une série des fiches techniques en épidémiologie et en bio statistique. Ces fiches méthodologiques décrivent, d'une manière standardisée, les modes d'usage des concepts, des outils et des méthodes utilisés lors des différentes phases de la rédaction médicale scientifique depuis la phase de la recherche documentaire jusqu'à la phase de la communication médicale scientifique.

Cette série est rédigée par des experts de méthodologie de recherche dans les universités du Grand Maghreb et les facultés sœurs au Nord de la Méditerranée. Chaque fiche répond à trois questions essentielles (Quoi ? Pourquoi ? Comment) du concept étudié, en se basant sur un article publié dans la revue Tunis Med.

Le coordinateur de la série « Fiches Méthodologiques ».

Professeur Ahmed Ben Abdelaziz (Président du Réseau Maghrébin PRP2S)
ahmedbenabdelaziz.prp2s@gmail.com

Série des Fiches méthodologiques

Sommaire

Fiche n°1 (janvier 2020):

Comment calculer la taille d'un échantillon pour une étude observationnelle

Serhier Z, et al. (Faculté de Médecine et de Pharmacie de Casablanca. Maroc)

Fiche n°2 (février 2020):

La recherche qualitative: méthodes, outils, analyse

Soulimane A. (Faculté de Médecine, Université Djillali Liabes, Sidi Bel Abbès, Algérie)

Fiche n°3 (mars 2020)

Et Allah ...créa la variabilité

Barhoumi T, et al (Réseau Maghrébin PRP2S)

Fiche n°4 (mai 2020)

Réussir votre recherche bibliographique sur PubMed

Ben Abdelaziz A, et al (Réseau Maghrébin PRP2S)

Fiche n°5 (juin 2020)

Réussir la rédaction de votre « Protocole de Recherche » en sciences de la santé

Ben Abdelaziz A, et al (Réseau Maghrébin PRP2S)

Fiche n°6 (juillet 2020)

Analyse multi variée par régression logistique

Ben Salem K, et al (Réseau Maghrébin PRP2S)

Fiche n°7 (octobre 2020)

Tests non paramétriques pour comparer deux ou plusieurs moyennes sur des échantillons indépendants

Bezzaoucha A, et al (Réseau Maghrébin PRP2S)

Fiche n°8 (novembre 2020)

Comment évaluer la concordance entre deux mesures qualitatives par le test Kappa ?

Khiari H, et al (Réseau Maghrébin PRP2S)

Fiche n°9 (décembre 2020)

Comment comparer plusieurs moyennes par le test d'Analyse de Variance (ANOVA) ?

Khiari H, et al (Réseau Maghrébin PRP2S)

Fiche n°10 (décembre 2020)

Tests non paramétriques sur SPSS pour comparer deux ou plusieurs moyennes sur des échantillons appariés. (test de Wilcoxon et test de Friedman)

Bezzaoucha A et al (Réseau Maghrébin PRP2S)

Correspondance

Ahmed Ben Abdelaziz

Laboratoire de Recherche LR19SP01 « Mesure et Appui de la Performance des Etablissements de Santé ». Université de Sousse (Tunisie)

ahmedbenabdelaziz.prp2s@gmail.com

EUDE DE CAS

Des auteurs tunisiens (associés à des auteurs du Hidjaz d'Arabie) ont étudié, dans un article intitulé « Modulation of drug resistance and biofilm formation of *Staphylococcus aureus* isolated from the oral cavity of Tunisian children » publié en 2017 dans « The Brazilian Journal of Infectious Diseases » [1], les activités antimicrobiennes modulatrices de l'extrait de *Lactobacillus plantarum* (LPE). Cet extrait est isolé d'un lait fermenté traditionnel et a été testé sur 10 souches de staphylocoque doré. Dans un tableau de l'article, les auteurs ont donné, pour chacune des souches de staphylocoque, la concentration inhibitrice minimale (MIC) et la concentration bactéricide minimale (MBC) exprimées en µg/mL. Ces concentrations, moyennes d'expériences répétées au moins deux fois, ont été données pour le benzalchonium chlorure (BC) pour la chlorhexidine (CHX) et pour la tétracycline (TET), séparément ou associés au LPE. A titre d'illustration, voici les résultats tronqués rapportés pour une des souches de staphylocoque (S1) :

Les auteurs avaient auparavant précisé dans la partie *Material and methods*: «Data were analyzed using SPSS v. 17.0 software. The Friedman test, followed by the Wilcoxon signed ranks test were used to assess inter-group differences. In addition, statistical significance was set at p -value <0.05 » [Les données ont été analysées à l'aide du logiciel SPSS v. 17.0. Le test de Friedman, suivi du test des rangs signés de Wilcoxon ont été utilisés pour évaluer les différences entre les groupes. De plus, la signification statistique a été fixée à une valeur de $p <0,05$ (tdr)]. Bien que les auteurs n'ont pas donné la probabilité « p » relative aux tests exécutés, ni dans les tableaux ni dans la partie *Results*, ils se rattrapaient tardivement dans la partie *Discussion* à propos du tableau concerné par l'assertion

suivante: «The results showed a statistically significant difference between the antibacterial effects of drugs with and without LPE supplementation ($p <0.05$) » [Les résultats ont montré une différence statistiquement significative entre les effets antibactériens des médicaments avec et sans supplémentation en LPE ($p <0,05$) (tdr)]. . On peut penser que les auteurs, pour chacune des souches de staphylocoque, ont comparé d'abord les moyennes des MIC des trois groupes BC, CHX et TET sans apport de LPE (respectivement 8, 256 et 8) et ensuite les moyennes des MIC avec apport de LPE (4, 128 et 4). Il ne semble pas que les auteurs aient comparé les six moyennes simultanément par un seul test. Au lieu des moyennes de MIC, les auteurs pouvaient aussi comparer les moyennes de MBC. Nonobstant les écarts considérables des valeurs des moyennes, il s'agit d'une situation où on compare simultanément plusieurs moyennes mesurées non pas sur des échantillons indépendants (fiche technique n°7) mais sur des échantillons appariés puisqu'il s'agit de la même souche de staphylocoque. Les auteurs ont donc eu raison d'utiliser le test de Friedman qui est indiqué pour comparer plusieurs moyennes mesurées sur des échantillons appariés. Il peut paraître encore plus clair que les auteurs ont comparé les moyennes des MIC, sans apport ou avec apport de LPE, pour chacune des substances testées. Par exemple, on peut comparer pour la CHX, la moyenne 256 à la moyenne 512. Les mêmes remarques pourraient être formulées à propos de la MBC. Pour comparer deux moyennes sur des échantillons appariés, les auteurs ont là aussi fait le bon choix en exécutant le test de Wilcoxon. Les tests de Friedman et de Wilcoxon ont l'avantage en outre d'être des tests non paramétriques qui ne sont pas soumis à des conditions particulières d'application. Les auteurs ont conclu en tout cas que le LPE avait une activité antimicrobienne modulatrice avérée.

	BC		BC + LPE ^a		CHX		CHX + LPE ^a		TET		TET + LPE ^a	
	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC
S1	8	16	4(2)	8(2)	256	512	128(2)	512	8	16	4(2)	8(2)

a Fold reductions are given in parentheses

QUIZZ

Question 1. D'après la synthèse de l'étude, la cible de l'étude est:

- Le LPE
- La MIC ou la MBC
- Le BC ou la CHX ou la TET

Question 2. Le test de Wilcoxon sert à comparer:

- Deux moyennes déterminées sur des échantillons appariés
- Au moins trois moyennes déterminées sur des échantillons appariés
- Deux ou plusieurs moyennes déterminées sur des échantillons indépendants

Question 3. Si on prend un groupe de sujets de 20 sujets convalescents du Covid-19 et qu'on mesure leur taux moyen d'anticorps neutralisants par quatre méthodes différentes, le test statistique le plus adapté pour comparer ces quatre taux moyens est:

a. Le test «F» paramétrique classique de l'analyse de la variance

b. Le test d'analyse de la variance de Kruskal-Wallis

c. Le test de Friedman

INTRODUCTION

Dans une précédente fiche technique (fiche n°7), les tests non paramétriques sur SPSS pour comparer deux ou plusieurs moyennes sur des échantillons indépendants ont été exposés. Mais on peut être confronté aussi à la comparaison de deux ou plusieurs moyennes déterminées sur des échantillons appariés (groupes liés) lorsqu'on tient compte du fait que les valeurs utilisées pour calculer ces moyennes correspondent deux à deux, trois à trois... Ces valeurs peuvent être mesurées sur un même individu ou des individus appariés selon une caractéristique donnée (selon le sexe, l'âge, la classe sociale comme cela peut se faire dans une étude cas-témoins...).

LES TESTS NON PARAMÉTRIQUES COMPARANT DES MOYENNES SUR DES GROUPES APPARIÉS : QUOI ?

La présente fiche expose deux tests non paramétriques pour comparer deux ou plusieurs moyennes sur des échantillons appariés: le test de Wilcoxon et le test de Friedman. Ces deux tests sont dits « non paramétriques » parce qu'ils ne sont pas soumis à des conditions particulières d'application. Lorsque la référence à la loi normale n'est manifestement pas possible pour comparer deux moyennes avec le test « t » de comparaison de deux moyennes déterminées sur deux groupes appariés, le test de Wilcoxon peut avantageusement le remplacer. Le test de Wilcoxon peut être appliqué à un très petit nombre de paires (de l'ordre de 6), sa puissance est appréciable puisqu'elle est de l'ordre de 95% de celle du test « t » paramétrique pour séries appariées. Le test de Friedman est le seul test appliqué pour comparer plus de

deux moyennes déterminées sur des groupes appariés. Le test de Friedman peut être assimilé à une analyse de variance lorsque les moyennes sont déterminées sur des échantillons liés.

LES TESTS NON PARAMÉTRIQUES COMPARANT DES MOYENNES SUR DES GROUPES APPARIÉS : POURQUOI ?

Les deux tests de Wilcoxon et de Friedman sont utilisés pour comparer respectivement deux et plusieurs moyennes déterminées sur des échantillons appariés lorsque la cible de l'étude (critère de jugement ou variable de réponse) est une variable quantitative ou une variable qualitative ordinale (données de la variable transformées en rangs). Si on voulait comparer au moins trois groupes appariés avec comme variable cible une variable qualitative nominale à deux ou plus de deux modalités, on comparerait plusieurs pourcentages par le test « Q » de Cochran qu'il ne faut pas confondre avec le test de Friedman. Si on voulait comparer deux groupes appariés, avec comme variable cible une variable qualitative nominale à deux modalités, on comparerait deux pourcentages avec le test de McNemar. Si ce critère de jugement était à plus de deux modalités, le test « Q » de Cochran serait à aussi utilisé pour comparer plusieurs pourcentages. Rappelons le fait que lorsqu'on ne tient pas compte de l'appariement, le test statistique est moins enclin à mettre en évidence une différence significative entre les moyennes ou entre les pourcentages qui pourtant existerait réellement.

LES TESTS NON PARAMÉTRIQUES COMPARANT DES MOYENNES SUR DES GROUPES APPARIÉS : COMMENT ?

A. Test de Wilcoxon

On veut comparer chez deux groupes d'enfants appariés selon le sexe (à un garçon d'un groupe, on fait correspondre un garçon de l'autre groupe...), les deux moyennes du quotient intellectuel (tableau 1). On ne retiendra pas l'hypothèse de normalité du quotient intellectuel bien que l'exécution du test de Kolmogorov Smirnov montre que la distribution des 38 valeurs du quotient suit une loi normale ($p = 0.10$). Aussi, utiliserons-nous le test de Wilcoxon pour comparer les deux groupes.

Tableau 1. Quotient intellectuel (QI) de deux groupes de 19 enfants appariés

N° paire	QI (Groupe 1)	QI (Groupe 2)
1	100	95
2	110	71
3	80	80
4	92	100
5	76	80
6	98	76
7	84	70
8	88	70
9	80	100
10	110	95
11	86	75
12	90	88
13	104	96
14	98	91
15	92	92
16	77	80
17	100	94
18	110	101
19	104	106
m ± s	93.6 ± 11.3	87.4 ± 11.6

La base de données sur SPSS, après la création de deux variables numériques pour le quotient intellectuel (figure n°1), est celle de la figure 2. On clique sur Analyse > Tests non paramétriques > Boîte de dialogue ancienne version > 2 échantillons liés comme indiqué ci – dessous (figure n°3). Il faut introduire ensuite les deux variables par leur libellé « Groupe 01 » et « Groupe 02 » dans « Tests pour deux échantillons liés », on clique sur OK sans oublier de cocher le test de « Wilcoxon » (figure n°4). Les résultats suivants sont donnés (figure n°5). La statistique du test de Wilcoxon est la somme des rangs des différences de signe moins ou la somme des rangs des différences de signe plus, soit 32 ou 121. En fait ce qui importe, c'est la valeur de p ($p = 0.035$) correspondant au test de l'écart réduit ($z = -2.11$). Les deux moyennes sont donc significativement différentes. En moyenne, le groupe 1 a les valeurs les plus élevées pour le quotient intellectuel. Si on avait exécuté le test paramétrique t de comparaison de deux moyennes sur groupes appariés, on aurait trouvé $t = 2.16$, $p = 0.045$. Les deux moyennes auraient là aussi été significativement différentes. Mais si on n'avait pas tenu

compte de l'appariement et si on avait effectué le test t de comparaison de deux moyennes sur des échantillons indépendants, on aurait trouvé $t = 1.68$ ($df = 36$), $p = 0.10$. On aurait ainsi conclu à tort à l'absence de différence significative entre les deux moyennes.

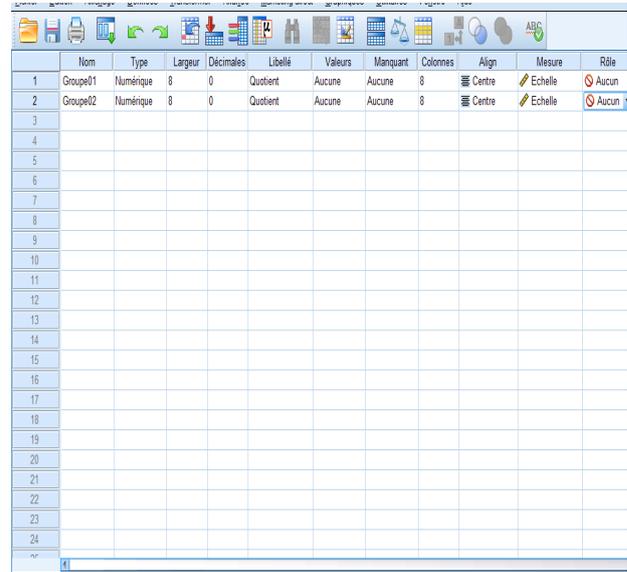


Figure 1. Capture d'écran de vue des deux variables en vue d'un appariement sur le logiciel SPSS

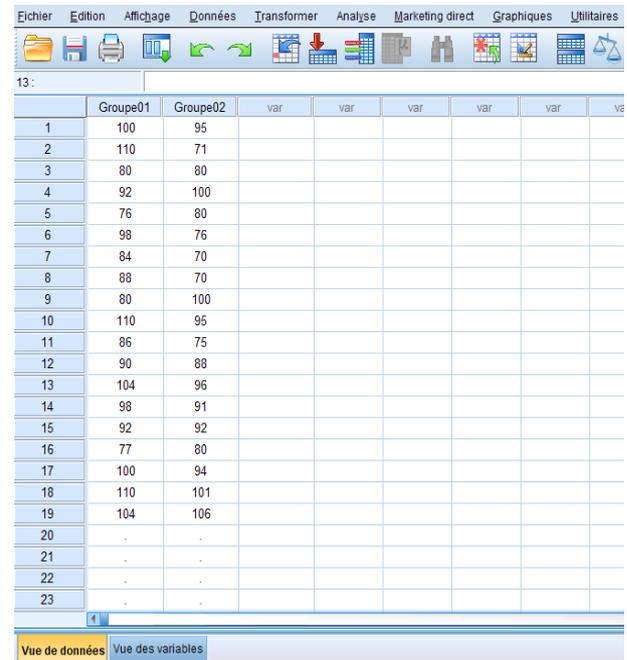


Figure 2. Capture d'écran de la base des données avec les deux groupes appariés sur le logiciel SPSS

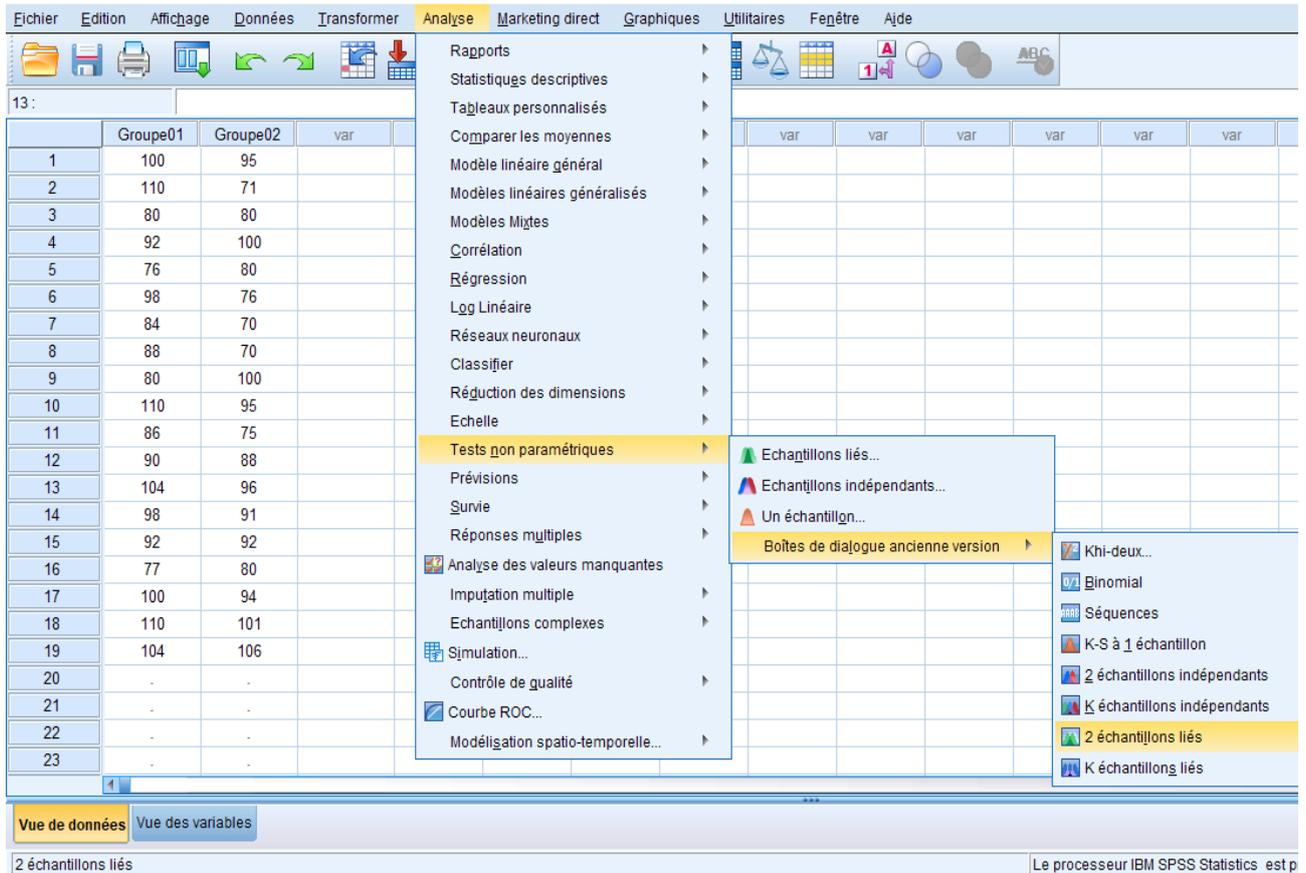


Figure 3. Chemin de la commande, sur le logiciel SPSS, de comparaison des deux moyennes par le test de Wilcoxon

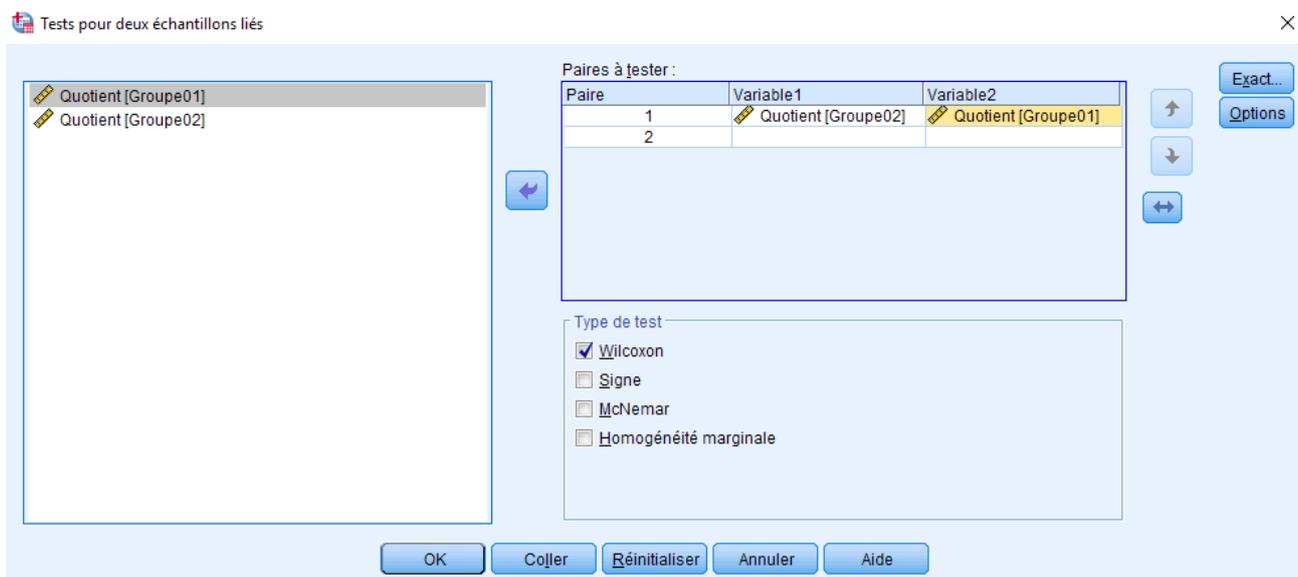


Figure 4. Procédure de réalisation du test de Wilcoxon sur le logiciel statistique SPSS

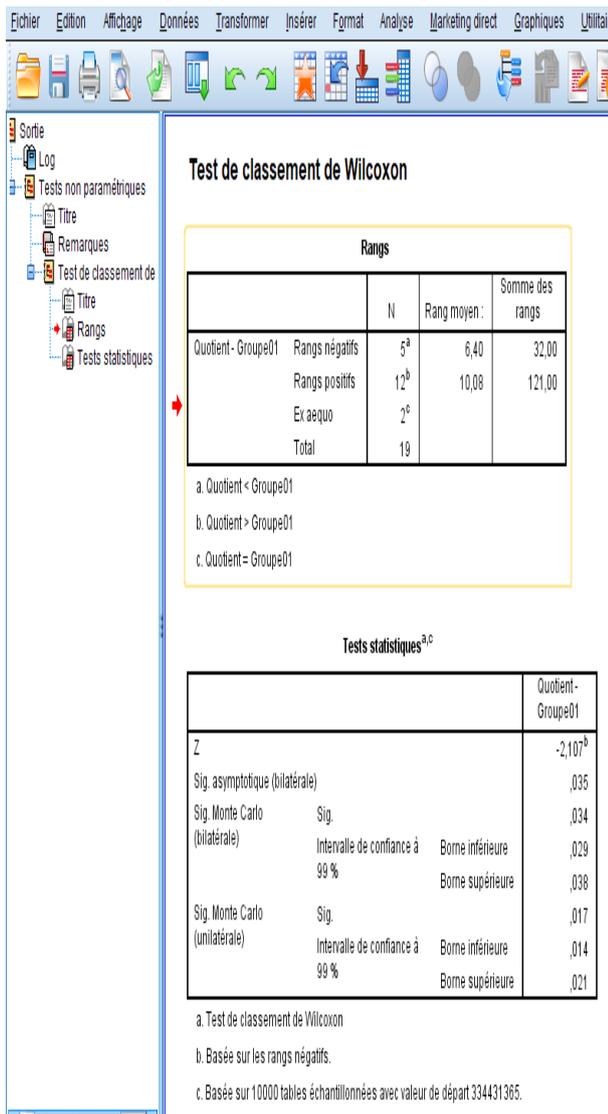


Figure 5. Sorties des résultats du test de Wilcoxon sur le logiciel SPSS

B. Test de Friedman

Le tableau 2 contient les appréciations accordées à un graphe par trois examinateurs qui ont corrigé tour à tour les mêmes 20 copies d'examens. L'appréciation devait être formulée de façon qualitative : complet, partiellement complet, incomplet.

Tableau 2 : Appréciation qualitative d'un graphe d'examen sur 20 copies par trois examinateurs différents

(C: complet, PC : partiellement complet, I : incomplet)

Copie	Examinateur		
	A	B	C
1	I ^a	I	I
2	PC ^b	PC	PC
3	PC	C ^c	C
4	PC	C	C
5	C	C	C
6	C	C	C
7	I	C	PC
8	C	C	C
9	PC	C	C
10	C	C	C
11	C	C	C
12	C	C	PC
13	I	PC	I
14	PC	C	C
15	PC	PC	PC
16	PC	C	C
17	I	PC	I
18	PC	C	C
19	C	I	C
20	PC	C	PC

L'appréciation est une variable qualitative ordinale puisque ses modalités sont soumises à un ordre. En accordant la valeur «1» à la modalité C, la valeur «2» à la modalité PC et la valeur «3» à la modalité I, on peut appréhender si les trois examinateurs donnent, en moyenne, la même appréciation. Pour comparer les trois moyennes des appréciations, on utilise le test de Friedman puisque ces moyennes concernent les mêmes copies (groupes liés) et non pas des échantillons indépendants. Le test de Friedman peut être utilisé même lorsque le critère de jugement est une variable qualitative ordinale comme dans l'exemple. La base de données sur SPSS, après la création de trois variables numériques pour les trois examinateurs A, B et C (figure n°6), est celle de la figure 7. On clique sur Analyse > Tests non paramétriques > Boite de dialogue ancienne version > K échantillons liés comme indiqué ci – dessous (figure n°8). Il faut introduire ensuite les trois variables par leur libellé « Appréciation par A », « Appréciation par B » et « Appréciation par C » dans

« Tests pour plusieurs échantillons liés », on clique sur OK sans oublier de cocher le test de « Wilcoxon » (figure n°9). Les résultats suivants sont donnés (figure n°10). La statistique du test de Friedman est un $c^2 : c^2 = 10.76, dl = 2, p = 0.005$. Ce qui importe comme toujours, c'est la valeur du « p ». Les trois moyennes sont donc significativement différentes. En moyenne, c'est l'examineur B qui

est le plus enclin à accorder l'appréciation complet au graphe des 20 copies. Si on n'avait pas tenu compte de l'appariement et si on avait effectué le test de Kruskal-Wallis de comparaison de plusieurs moyennes sur des échantillons indépendants (fiche n°7), on aurait conclu à tort à l'absence de différence significative entre les trois moyennes ($c^2 = 4.51, dl = 2, p = 0.10$).

Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle	
1	ExamineurA	Numérique	8	0	appréciation par A	(1, Complet)...	Aucune	8	Droite	Nominal	Entrée
2	ExamineurB	Numérique	8	0	appréciation par B	(1, Complet)...	Aucune	8	Droite	Nominal	Entrée
3	ExamineurC	Numérique	8	0	appréciation par C	(1, Complet)...	Aucune	8	Droite	Nominal	Entrée

Figure 6. Capture d'écran de vue des variables pour trois groupes appariés sur le logiciel SPSS

	ExamineurA	ExamineurB	ExamineurC	var
1	3	3	3	
2	2	2	2	
3	2	1	1	
4	2	1	1	
5	1	1	1	
6	1	1	1	
7	3	1	2	
8	1	1	1	
9	2	1	1	
10	1	1	1	
11	1	1	1	
12	1	1	2	
13	3	2	3	
14	2	1	1	
15	2	2	2	
16	2	1	1	
17	3	2	3	
18	2	1	1	
19	1	3	1	
20	2	1	2	
21				
22				
23				

Figure 7. Capture d'écran de la base des données avec les trois groupes appariés sur le logiciel SPSS

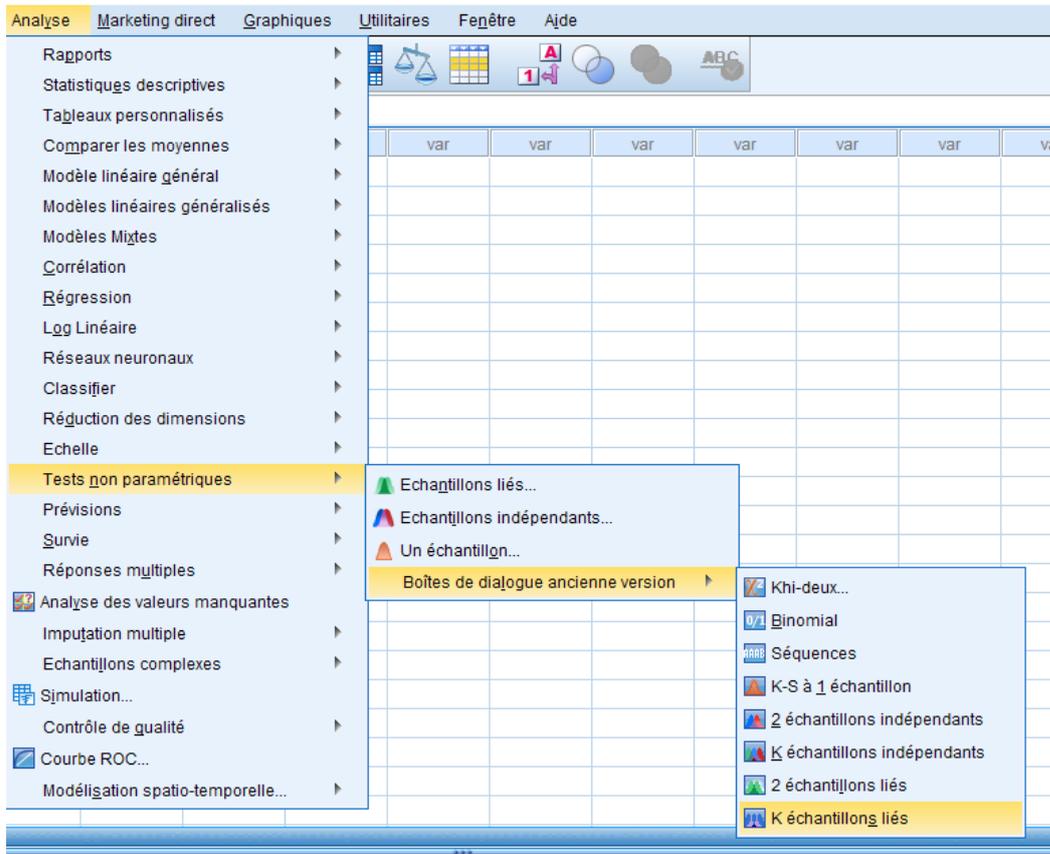


Figure 8. Chemin de la commande, sur le logiciel SPSS, de la comparaison de plusieurs moyennes par le test de Friedman

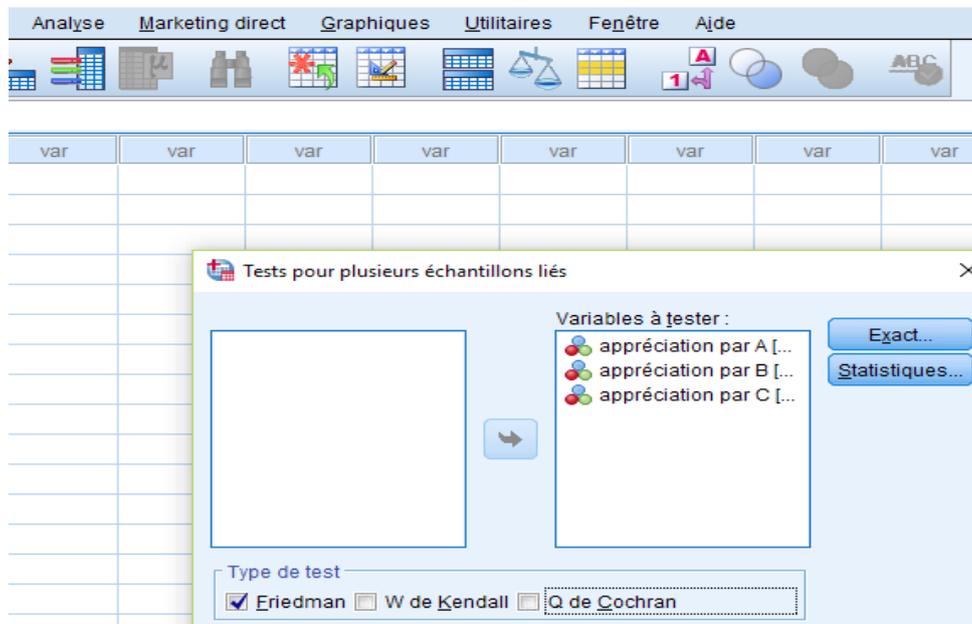


Figure 9. Procédure de réalisation du test de Friedman sur le logiciel statistique SPSS

Tests non paramétriques

Statistiques descriptives

	N	Moyenne	Ecart type	Minimum	Maximum
appréciation par A	20	1,85	,745	1	3
appréciation par B	20	1,40	,681	1	3
appréciation par C	20	1,55	,759	1	3

Test de Friedman

Rangs

	Rang moyen :
appréciation par A	2,38
appréciation par B	1,68
appréciation par C	1,95

Tests statistiques^a

N	20
Khi-deux	10,757
ddl	2
Sig. asymptotique	,005

a. Test de Friedman

Figure 10. Sorties des résultats du test de Friedman sur le logiciel SPSS

L'essentiel à retenir

Pour choisir le bon test en vue de comparer des moyennes sur des échantillons appariés, on peut retenir le cheminement suivant [2,3]:

Tests statistiques pour la comparaison de moyennes sur des échantillons appariés		
Deux moyennes		Plusieurs moyennes
Référence à la loi normale possible		Test non paramétrique de Friedman
Oui	Non	
Test «t» de Student pour groupes liés	Test non paramétrique de Wilcoxon	

Réponses aux Quizz

Question 1.

La cible de l'étude est le critère de jugement ou la variable de réponse, c'est donc la MIC ou la MBC. Ces variables sont les caractères quantitatifs pour lesquelles on veut tester séparément la liaison avec la variable qualitative servant de base aux comparaisons. Cette variable qualitative est le facteur d'étude dont les modalités sont le BC, la CHX et la TET associés ou non au LPE. La bonne réponse est donc la réponse b.

Question 2.

Le test de Wilcoxon ne sert qu'à comparer deux moyennes déterminées sur des échantillons appariés. Si on veut comparer au moins trois moyennes déterminées sur des échantillons appariés, on utilise de test de Friedman. Pour comparer deux moyennes sur des échantillons indépendants, on peut utiliser le test t de student, le test de Welch ou le test U de Mann-Whitney. Pour comparer plus de deux moyennes sur des échantillons indépendants, on peut utiliser le test « F » d'analyse de la variance ou le test « H » de Kruskal-Wallis selon que la référence à la loi normale de la distribution de la variable cible est possible ou non. La réponse exacte est donc la réponse a.

Question 3.

C'est le test de Friedman qui est utilisé dans cette situation. Les deux autres tests proposés sont utilisés pour comparer des moyennes déterminées sur des échantillons indépendants. La réponse exacte est donc la réponse c.

Pour en savoir plus

1. Zmantar T, Ben Slama R, Fdhila K, Kouidhi B, Bakhrouf A, Chaieb K. Modulation of drug resistance and biofilm formation of *Staphylococcus aureus* isolated from the oral cavity of Tunisian children. *Braz J Infect Dis* 2017; 21(1):27-34.
2. Bezzaoucha A. Tests statistiques en sciences médicales (2ème édition). Alger : Office des Publications Universitaires, 2017.
3. Ancelle T. Statistique Epidémiologie (4ème édition). Paris : Maloine, 2017.

CONCLUSION

La comparaison de deux ou plusieurs moyennes déterminées sur des groupes liés (appariés) ne pose pas de problème particulier avec les excellents logiciels dont on dispose et qui permettent d'éviter de lourds calculs. La comparaison de trois moyennes ou plus, déterminées sur des groupes liés, est possible avec le test non paramétrique de Friedman. La comparaison de deux moyennes sur des groupes appariés est possible, soit avec le test paramétrique « t » de Student pour séries appariées, soit avec le test non paramétrique de Wilcoxon.