



Série de Fiches Méthodologiques en Recherche et en Rédaction Scientifique

Fiche Méthodologique n°7:

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

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

مقارنة متوسطين أو أكثر من عينات مستقلة بالاختبارات غير المعيارية

A Bezzaouha¹, A Bouamra¹, A Ammimer¹, A El Keboub¹, A Ben Abdelaziz²

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 de la Performance des Etablissements de Santé ». Université de Sousse (Tunisie)

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 Abbes, 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
Bezzaouha 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

ETUDE DE CAS

Dans une étude d'évaluation pédagogique intitulée «outil pédagogique numérique d'auto-apprentissage versus travaux dirigés par un enseignant: Essai randomisé», publiée récemment dans la revue « La Tunisie Médicale » [1], l'équipe du Professeur Dziri a rapporté les résultats d'une étude expérimentale comparant un Outil Pédagogique Numérique (OPN) dans les urgences abdominales non traumatiques aux Travaux Dirigés par l'Enseignant (TDE), sur deux groupes parallèles d'étudiants de Deuxième Année du Deuxième Cycle des Etudes Médicales (DCEM2) obtenus par tirage au sort.

Des situations cliniques d'urgences abdominales non traumatiques ont été sélectionnées parmi les urgences abdominales les plus fréquentes rencontrées lors d'une garde de chirurgie générale (appendicite aiguë, péritonite aiguë par perforation d'ulcère gastrique ou duodénal, cholécystite aiguë lithiasique, angiocholite aiguë lithiasique, pancréatite aiguë, occlusion intestinale aiguë sur bride, occlusion colique d'origine néoplasique, volvulus du côlon pelvien, infarctus mésentérique et abcès péri-néoplasique du cancer du côlon). Pour chaque dossier sélectionné, une observation clinique a été rédigée, rapportant, en plus d'un cliché d'imagerie médicale spécifique, les antécédents du patient, l'histoire de la maladie (symptomatologie), les données de l'examen physique, les données de la biologie. L'outil numérique d'auto-apprentissage a été confectionné sous forme de diaporama PowerPoint®, mis sur un CD-rom. Dans le groupe «OPN», l'étudiant avait à étudier ce CD-rom, sans l'aide du tuteur (enseignant). Pour chaque cas clinique, le temps alloué à l'étudiant était de 10 minutes pour chaque cas clinique (travail individuel). Dans le groupe «TDE », l'étudiant devait analyser les mêmes dossiers

sélectionnés pendant une durée de sept minutes. Durant les trois minutes suivantes, deux enseignants vérifiaient avec l'étudiant son interprétation des clichés d'imagerie médicale, le diagnostic posé et lui assuraient un «feedback». Le critère de jugement principal était l'évaluation de la compétence acquise par l'étudiant à l'aide du score total obtenu par l'étudiant à l'Epreuve Clinique Objective en Stations Multiples (ECOSM), composé de 14 cas cliniques concernant les mêmes pathologies parmi celles sélectionnées pour les cas cliniques de l'«OPN » et des « TDE ». Le taux de succès a été défini par un score total supérieur ou égal à 210 points (14 stations de 10 points chacune soit un total de 140 points pour chaque étudiant et par correcteur, le score total pour les trois correcteurs était de 420 points, 210 points étant la moitié de 420 points).

Un total de 27 étudiants: 14 étudiants dans le groupe «OPN » et 13 étudiants dans le groupe «TDE» ont été inclus dans cet essai pédagogique. La moyenne du score total pour tous les étudiants était de 230 ± 52 points [extrêmes : 71,5 - 318,5]. Le score total a été de $254,1 \pm 39,2$ dans le groupe «OPN» versus $203,6 \pm 52,4$ dans le groupe «TDE». Suite à la comparaison du score total pour le groupe «OPN» au score total pour le groupe «TDE», en utilisant le test «U» de Mann-Whitney, les auteurs ont constaté une différence statistiquement significative ($p < 10^{-2}$) en termes d'acquisition de compétence pour l'interprétation de l'imagerie médicale dans les urgences abdominales non traumatiques, entre les deux groupes. Cette étude a conclu que «l'OPN a permis à l'étudiant de DCEM2 d'acquérir une compétence en matière d'interprétation de l'imagerie médicale dans les urgences abdominales non traumatiques avec un taux supérieur comparativement aux TDE».

Quizz

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

- a. L'outil pédagogique numérique
- b. Le succès à l'Epreuve clinique
- c. Le score obtenu à l'Epreuve clinique

Question 2. Le test U de Mann-Whitney sert à comparer:

- a. Deux moyennes

- b. Au moins trois moyennes
- c. Deux ou plusieurs pourcentages

Question 3. Les auteurs ont préféré le test U de Mann-Whitney au test t de Student parce que :

- a. Les variances des deux groupes à comparer sont significativement différentes
- b. Le score obtenu à l'Epreuve clinique ne suit probablement pas une loi normale
- c. Le test U est plus puissant que le test t

INTRODUCTION

Un investigateur est confronté au problème de la comparaison de deux ou plusieurs moyennes lorsqu'il cherche à étudier une relation entre une variable numérique (quantitative) et une variable qualitative comportant deux ou plusieurs catégories d'individus. Les groupes d'individus indépendants l'un de l'autre sont à distinguer de groupes appariés (sujets pris comme leur propre témoin, groupes appariés selon une caractéristique quelconque dans une étude cas-témoins...). Un investigateur peut être intéressé, par exemple, par la comparaison de la durée moyenne d'hospitalisation chez les sujets victimes d'accidents de la route et de la durée moyenne d'hospitalisation chez les sujets victimes de chutes. Les deux groupes de sujets sont indépendants, ils constituent les deux modalités (accident de la route, chute) de la variable qualitative qui pourrait être appelée «type d'accident». Celle-ci est la variable qui sert de base à la comparaison car elle identifie les groupes de l'étude à comparer, elle est souvent appelée variable dépendante ou même facteur d'étude. Un troisième groupe de sujets peut être rajouté si l'investigateur est intéressé par la comparaison de trois moyennes, ou même un quatrième groupe pour comparer quatre moyennes... La durée d'hospitalisation est la variable numérique dont on cherche à étudier la relation avec le type d'accident. Cette variable numérique, lorsqu'il s'agit de comparer deux ou plusieurs moyennes, est la variable indépendante appelée aussi critère de jugement, cible ou variable de réponse.

LES TESTS NON PARAMÉTRIQUES: QUOI ?

Les deux tests statistiques exposés dans cette fiche sont le test « U » de Mann-Whitney pour comparer deux moyennes et le test « H » de Kruskal-Wallis pour comparer plus de deux moyennes. Ces deux tests sont dits non paramétriques car 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 paramétrique classique « t » de Student ou plus de deux moyennes avec le test « F » paramétrique classique de l'analyse de la variance, les tests de Mann-Whitney et de Kruskal-Wallis peuvent les remplacer.

LES TESTS NON PARAMÉTRIQUES: POURQUOI ?

Les deux tests de Mann-Whitney et de Kruskal-Wallis sont applicables à des variables quantitatives et qualitatives ordinales (données transformées en rangs). Les deux tests devraient être préférés au test « t » paramétrique de Student et au test paramétrique de l'analyse de la variance en présence de valeurs extrêmes dans les distributions rendant les moyennes peu représentatives des groupes. Par ailleurs, les deux tests s'appliquent à toutes les distributions, quelle que soit leur forme. Le nombre d'observations par groupe devrait au moins être égal à deux, même si des effectifs au moins égaux à 5 sont souhaitables pour permettre notamment à la statistique « H » du test de Kruskal-Wallis de suivre une loi de Chi-Deux si le nombre de groupes à comparer est inférieur à quatre.

Mais à tailles égales d'échantillons, le test « U » de Mann-Whitney est un peu moins puissant que le test paramétrique « t » de Student, lorsque la référence à la loi normale est possible. De même, la puissance du test « H » de Kruskal-Wallis atteint 95% de celle de l'analyse de variance paramétrique lorsque la référence à la loi normale est possible. Si la référence à la loi normale pouvait être acceptée (distribution à peu près symétrique de la variable numérique) ou démontrée (utilisation du test non paramétrique de Kolmogorov Smirnov avec le même logiciel SPSS), l'utilisation des tests paramétriques, pour comparer des moyennes, devrait être préférée. Mais au moindre doute sur les conditions d'application des tests paramétriques (distribution normale du caractère numérique essentiellement et égalité des variances des groupes dans une moindre mesure), l'utilisation des tests non paramétriques permet de relever le défi et de comparer des moyennes dans tous les cas.

LES TESTS NON PARAMÉTRIQUES: COMMENT ?

A. Test "U" de Mann-Whitney

Considérons l'exemple suivant où l'on compare deux groupes de sujets accidentés vis à vis de la durée d'hospitalisation. Les données sont contenues au tableau 1. La moyenne de la durée d'hospitalisation chez les victimes d'accidents de la route est 15.5 jours tandis qu'elle est de 11.4 jours chez les victimes de chutes. La durée d'hospitalisation n'a habituellement pas une allure

symétrique et ne peut donc être supposée suivre une loi normale (durées faibles très nombreuses suivies par des durées plus longues d'effectif de plus en plus petit). Le test « U » de Mann-Whitney permet de comparer les deux moyennes.

Tableau 1. Répartition de 9 victimes de chutes et de 13 victimes d'accidents de la route selon la durée d'hospitalisation

	Durée d'hospitalisation (jours)												
Chutes	19	3	6	45	11	10	2	4	3				
Accidents de la route	9	4	14	10	15	2	8	55	5	30	30	7	12

La base de données sur SPSS, après la création d'une variable numérique pour la durée d'hospitalisation et d'une variable nominale pour le type d'accident, est la suivante (figure n°1). On clique sur Analyse > Tests non paramétriques > Boite de dialogue ancienne

version > 2 échantillons indépendants dans le menu principal, comme indiqué ci – dessous (figure n°2). Après avoir introduit respectivement les deux variables « Durée d'hospitalisation » et « Groupe » dans les deux espaces « Liste des variables à tester » et « Variable de regroupement », vous cliquez sur le bouton « Définir des groupes... », vous définissez les deux groupes (ici 1 et 2), vous cliquez sur poursuivre sans oublier de cocher le test « U de Mann-Whitney » (figure n°3). Les résultats suivants sont donnés par le logiciel après avoir cliqué sur OK (figure n°4). Il ressort des sorties du logiciel que le test «U » de Mann-Whitney = 42.5, p = 0.285 > 0.05 (situation bilatérale où chacune des deux moyennes pouvait être supérieure à l'autre). Par conséquent, les deux moyennes ne diffèrent significativement pas entre elles au seuil de signification 5%.

	groupe	DuréeHospital	var	var	var	var	var
1	Chutes	19					
2	Chutes	3					
3	Chutes	6					
4	Chutes	45					
5	Chutes	11					
6	Chutes	10					
7	Chutes	2					
8	Chutes	4					
9	Chutes	3					
10	Accidents de la route	9					
11	Accidents de la route	4					
12	Accidents de la route	14					
13	Accidents de la route	10					
14	Accidents de la route	15					
15	Accidents de la route	2					
16	Accidents de la route	8					
17	Accidents de la route	55					
18	Accidents de la route	5					
19	Accidents de la route	30					
20	Accidents de la route	30					
21	Accidents de la route	7					
22	Accidents de la route	12					

Figure 1. Capture d'écran, sur le logiciel SPSS, de la base des données avec les deux variables (types d'accidents, durée d'hospitalisation)

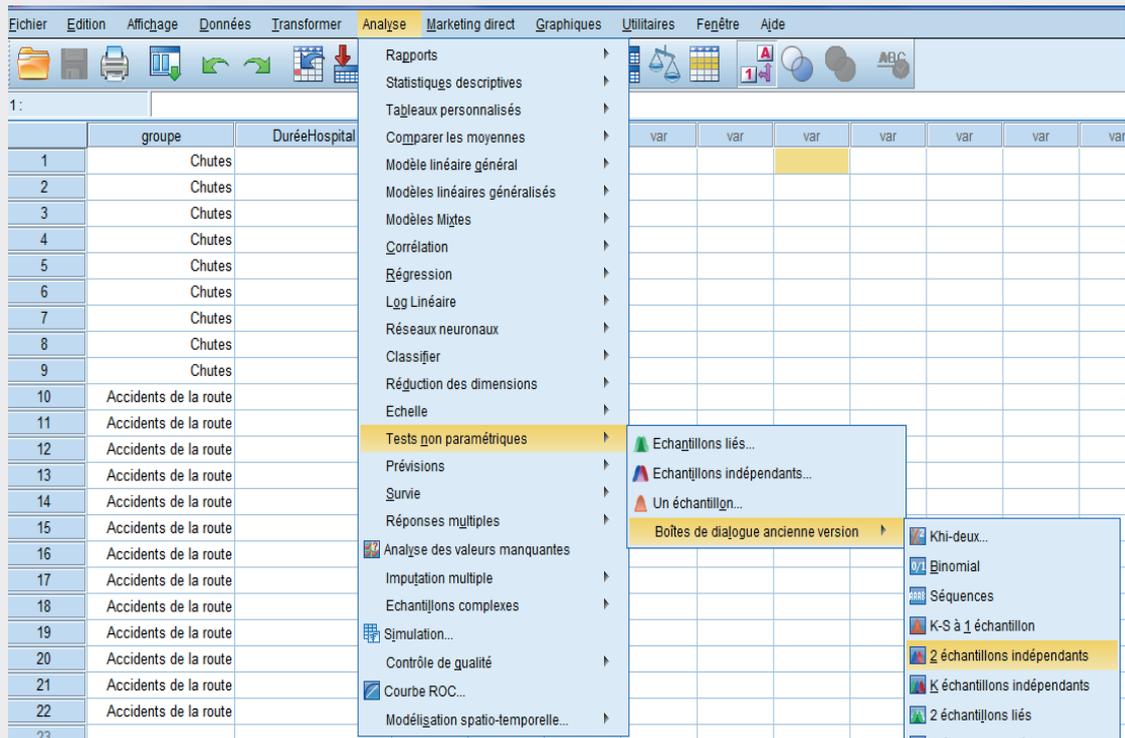


Figure 2. Chemin de la commande, sur le logiciel SPSS, de comparaison des deux échantillons indépendants, par des tests non paramétriques.

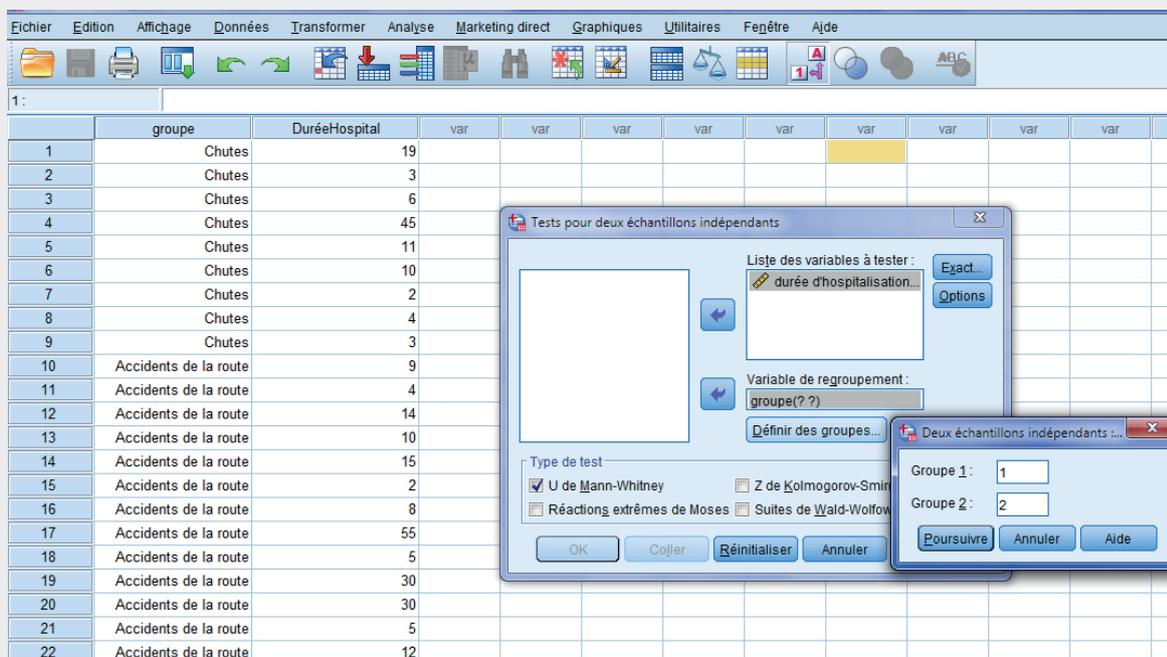


Figure 3. Procédure de réalisation du test U de Mann Withney sur le logiciel statistique SPSS

→ Tests non paramétriques

Test de Mann-Whitney

		Rangs		
	groupe de l'accident	N	Rang moyen :	Somme des rangs
durée d'hospitalisation en jours	Chutes	9	9,72	87,50
	Accidents de la route	13	12,73	165,50
	Total	22		

Tests statistiques^a

	durée d'hospitalisation en jours
U de Mann-Whitney	42,500
W de Wilcoxon	87,500
Z	-1,070
Sig. asymptotique (bilatérale)	,285
Sig. exacte [2*(sig. unilatérale)]	,292 ^b

a. Variable de regroupement : groupe de l'accident

b. Non corrigé pour les ex aequo.

Figure 4. Sorties des résultats de test « U » de Mann Withney sur le logiciel SPSS

B. Test “H” de Kruskal-Wallis

Reprenons les données du tableau 1 auxquelles on a rajouté les données sur les accidents par corps étrangers. L'ensemble de ces données est consigné au tableau 2.

Tableau 2. Répartition de 13 victimes d'accidents de la route, de 9 victimes de chutes et de 8 victimes de corps étrangers selon la durée d'hospitalisation

Route	9	4	14	10	15	2	8	55	5	30	30	7	12
Chute	19	3	6	45	11	10	2	4	3				
Corps étranger	1	2	1	2	4	3	7	16					

Nous nous proposons donc de comparer ces trois groupes vis-à-vis de la durée d'hospitalisation. En d'autres termes, nous comparons globalement entre elles, les trois moyennes de durée d'hospitalisation: 15.5 jours pour les accidents de la route, 11.4 jours pour les chutes et 4.5 jours pour les corps étrangers. L'analyse de variance par rangs de Kruskal-Wallis est applicable à cette situation puisque la durée d'hospitalisation ne peut être supposée suivre une loi normale.

On clique sur Analyse> Test non paramétriques > Boîte de dialogue ancienne version > K échantillons indépendants dans le menu principal, comme indiqué ci – dessous (figure n°5). Après avoir introduit respectivement les deux variables « Durée d'hospitalisation » et « Groupe » dans les deux espaces « Liste des variables à tester » et « Variable de regroupement », vous cliquez sur le bouton « Définir plage », vous définissez la plage selon le nombre

de groupes (ici de 1 à 3), vous cliquez sur « poursuivre » et cocher le test « H » de Kruskal-Wallis (figure n°6). Après avoir cliqué sur OK, le logiciel donne les résultats suivants (figure n°7). Dans cet exercice, la valeur du

test « H » de Kruskal-Wallis = 7.065, $p = 0.029 < 0.05$. Les trois moyennes diffèrent entre elles globalement et significativement au seuil de signification 5%.

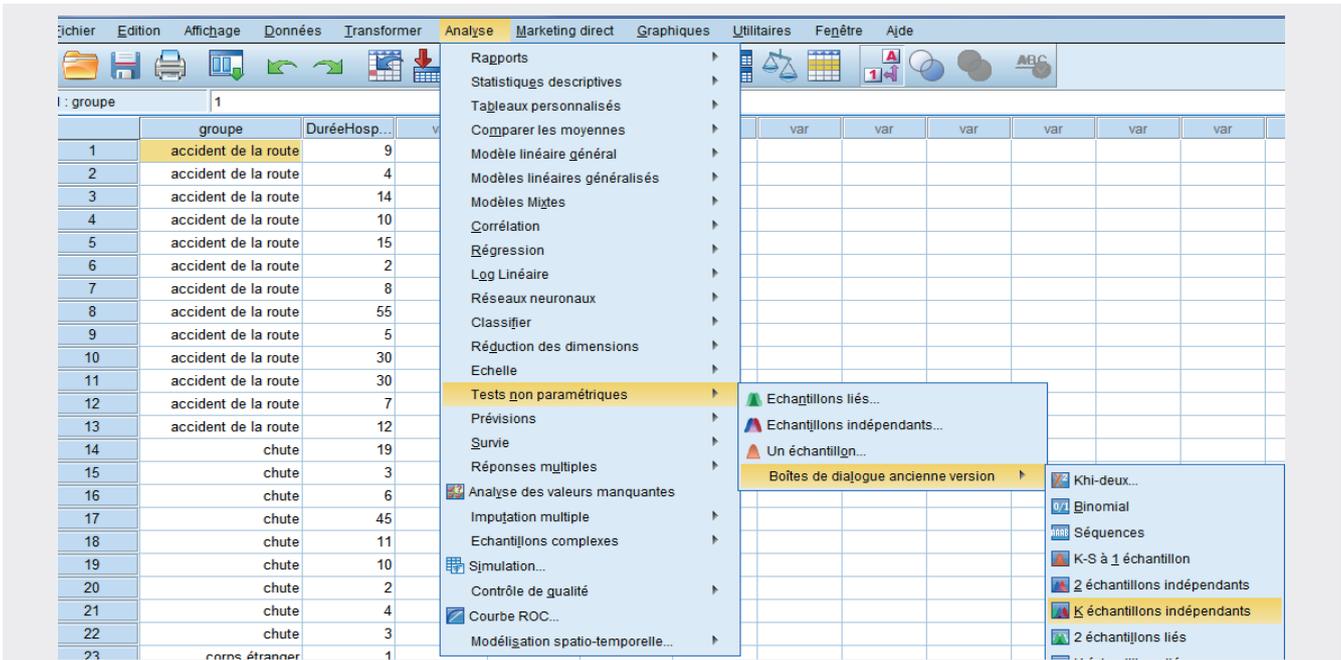


Figure 5. itinéraire de la commande de tests non paramétriques sur le logiciel SPSS

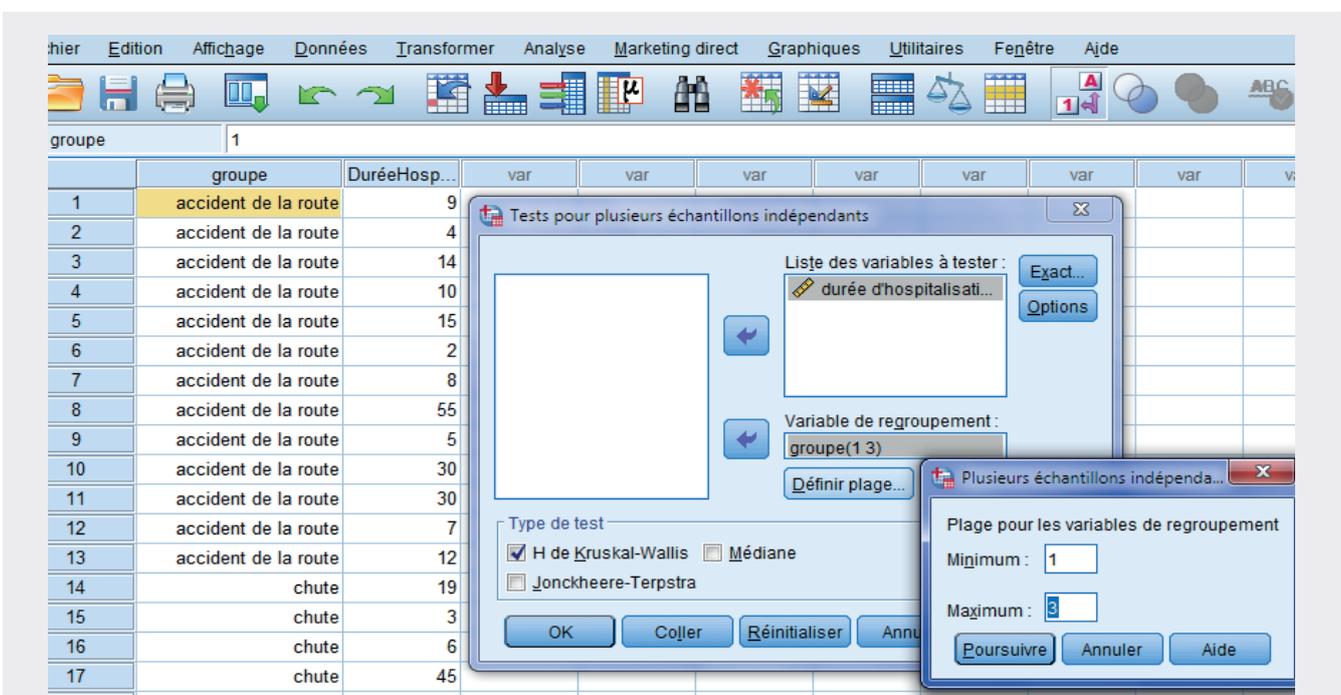


Figure 6. Processus du choix du test H de Kruskal-Wallis, sur le logiciel SPSS

→ Tests non paramétriques

Test de Kruskal-Wallis

Rangs			
	groupe de l'accident	N	Rang moyen :
durée d'hospitalisation en jours	accident de la route	13	19,42
	chute	9	15,67
	corps étranger	8	8,94
	Total	30	

Tests statistiques^{a,b}

	durée d'hospitalisation en jours
Khi-deux	7,065
ddl	2
Sig. asymptotique	,029

a. Test de Kruskal Wallis

b. Variable de regroupement :
groupe de l'accident

Figure 7. Sorties des résultats du test H de Kruskal-Wallis, sur le logiciel SPSS

C. Application des tests paramétriques aux mêmes exemples

On peut se poser la question de savoir si l'application des tests paramétriques aux mêmes exemples aboutit aux mêmes conclusions formulées ci-dessus. Si on applique le test « t » de Student aux données du tableau 1, le logiciel SPSS donne les résultats suivants (figure n°8). On peut remarquer que le test « F » de Levene sur l'égalité des variances ne montre pas de différence significative entre les deux variances ($F = 0.112$; $p = 0.742$). Une des conditions pour appliquer le test « t » de Student est satisfaite. Cependant, même si cette condition n'était pas satisfaite (hypothèse de variances inégales), le logiciel SPSS compare les deux moyennes avec un autre test « t » paramétrique, celui de Welch. Le test « t » de Welch est une adaptation du test « t » de Student, il compare deux moyennes de deux échantillons indépendants lorsque les variances des deux groupes sont inégales tout en maintenant la référence à la loi normale. Le dénominateur n'est pas une estimation commune des deux variances et le nombre de degrés de liberté, estimé par l'équation de Welch, n'est pas égal à la somme des deux effectifs

des deux groupes moins deux comme pour le test « t » de Student. Dans notre exemple, il convient de regarder plutôt à la ligne « Hypothèse de variances égales ». On retrouve la même conclusion énoncée par l'exécution du test « U » de Mann-Whitney, la différence entre les deux moyennes n'est pas significative au risque 5%: $t = -0.645$; $p = 0.526$.

Outre l'égalité des variances, une condition plus importante pour exécuter les tests paramétriques de comparaison de moyennes est la référence à la loi normale. Le logiciel SPSS permet de tester la normalité de la distribution du caractère numérique, en l'occurrence la durée d'hospitalisation dans notre exemple. En confondant les deux groupes (accidents de la route et chutes), le logiciel SPSS donne les résultats suivants pour le test de normalité de la durée d'hospitalisation (figure n°9). Le test de normalité de Kolmogorov-Smirnov, avec une statistique de 0.239 et un p de 0.002, montre que la distribution observée de la durée d'hospitalisation diffère significativement de la distribution théorique correspondante obéissant à une loi normale. L'autre test de normalité de la figure 9, celui de Shapiro-Wilk, aboutit à la même conclusion (le test de

Shapiro-Wilk est considéré comme plus puissant que celui de Kolmogorov-Smirnov et semble plus recommandé). Notre distribution de la durée d'hospitalisation n'obéissant pas à une loi normale, il convient d'adopter le test « U » de Mann-Whitney pour comparer les deux moyennes et de rapporter la probabilité p correspondante dans un rapport scientifique.

Le logiciel SPSS, par l'exécution du test paramétrique « F » d'analyse de la variance pour comparer les trois moyennes du tableau 2, donne les résultats suivants (figure n°10). Le test d'homogénéité des variances exécuté comme option sur SPSS avec le test paramétrique d'analyse de la variance aurait montré que les trois groupes proviennent d'une population de même variance (statistique de Leven

→ Test T

Statistiques de groupe					
groupe de l'accident	N	Moyenne	Ecart type	Moyenne erreur standard	
durée d'hospitalisation en jours	Chutes	9	11,44	13,703	4,568
	Accidents de la route	13	15,46	14,780	4,099

Test des échantillons indépendants										
		Test de Levene sur l'égalité des variances		Test t pour égalité des moyennes						
		F	Sig.	t	ddl	Sig. (bilatéral)	Différence moyenne	Différence erreur standard	Intervalle de confiance de la différence à 95 %	
durée d'hospitalisation en jours	Hypothèse de variances égales	,112	,742	-,645	20	,526	-4,017	6,226	-17,005	8,971
	Hypothèse de variances inégales			-,655	18,203	,521	-4,017	6,137	-16,901	8,867

Figure 8. Sorties SPSS sur le test t appliqué aux données du tableau 1

		Statistiques	Erreur standard	
durée d'hospitalisation en jours	Moyenne	13,82	3,018	
	Intervalle de confiance à 95 % pour la moyenne	Borne inférieure	7,54	
		Borne supérieure	20,10	
		Moyenne tronquée à 5 %	12,24	
	Médiane	9,50		
	Variance	200,442		
	Ecart type	14,158		
	Minimum	2		
	Maximum	55		
	Plage	53		
	Plage interquartile	12		
Asymétrie	1,830	,491		
Kurtosis	2,924	,953		

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Moyenne pondérée (Définition 1)	durée d'hospitalisation en jours	2,00	2,30	4,00	9,50	16,00	40,50	53,50
Valeurs charnières de Tukey	durée d'hospitalisation en jours			4,00	9,50	15,00		

Tests de normalité

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
durée d'hospitalisation en jours	,239	22	,002	,765	22	,000

Figure 9. Sorties SPSS du test de normalité de la durée d'hospitalisation (les deux groupes confondus)

= 1.73 ; p = 0.196). Le test « F » paramétrique d'analyse de la variance montre, contrairement au test « H » de Kruskal-Wallis, que les trois moyennes ne diffèrent pas significativement entre elles : F = 1.868, p = 0.174. Bien évidemment, il convient de retenir dans un rapport scientifique la conclusion énoncée après l'exécution du test « H » de Kruskal-Wallis. Là aussi, le logiciel SPSS

permet de tester l'hypothèse de normalité de notre distribution relative à la durée d'hospitalisation sur les trois groupes confondus (figure 11). Le test de Kolmogorov-Smirnov montre que la durée d'hospitalisation ne suit pas une loi normale (p = 0.001), ce qui conforte l'utilisation du test « H » de Kruskal-Wallis pour comparer les trois moyennes et la conclusion correspondante.

Caractéristiques

durée d'hospitalisation en jours

	N	Moyenne	Ecart type	Erreur standard	Intervalle de confiance à 95 % pour la moyenne		Minimum	Maximum
					Borne inférieure	Borne supérieure		
accident de la route	13	15,46	14,780	4,099	6,53	24,39	2	55
chute	9	11,44	13,703	4,568	,91	21,98	2	45
corps étranger	8	4,50	5,043	1,783	,28	8,72	1	16
Total	30	11,33	12,994	2,372	6,48	16,19	1	55

ANOVA

durée d'hospitalisation en jours

	Somme des carrés	ddl	Carré moyen	F	Sig.
Inter-groupes	595,214	2	297,607	1,868	,174
Intragroupes	4301,453	27	159,313		
Total	4896,667	29			

Figure 10. Sorties d'analyse de la variance (ANOVA) sur le logiciel SPSS

			Statistiques	Erreur standard
durée d'hospitalisation en jours	Moyenne		11,33	2,372
	Intervalle de confiance à 95 % pour la moyenne	Borne inférieure	6,48	
		Borne supérieure	16,19	
	Moyenne tronquée à 5 %		9,67	
	Médiane		7,00	
	Variance		168,851	
	Ecart type		12,994	
	Minimum		1	
	Maximum		55	
	Plage		54	
	Plage interquartile		11	
	Asymétrie		2,123	,427
Kurtosis		4,470	,833	

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Moyenne pondérée (Définition 1)	durée d'hospitalisation en jours	1,00	2,00	3,00	7,00	14,25	30,00	49,50
Valeurs charnières de Tukey	durée d'hospitalisation en jours			3,00	7,00	14,00		

Tests de normalité

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
durée d'hospitalisation en jours	,213	30	,001	,732	30	,000

Figure 11. Sorties SPSS du test de normalité de la durée d'hospitalisation (les trois groupes confondus)

L'ESSENTIEL À RETENIR

Pour choisir le bon test en vue de comparer des moyennes sur des échantillons indépendants, on peut essentiellement retenir le cheminement suivant :

Tests statistiques pour la comparaison de moyennes sur des échantillons indépendants			
Référence à la loi normale possible			
Oui		Non	
Tests paramétriques		Tests non paramétriques	
Egalité des variances		Test U de Mann-Whitney pour comparer deux moyennes	Test H de Kruskal-Wallis pour comparer plusieurs moyennes
Oui	Non		
Test t de Student pour comparer deux moyennes	Test t de Welch pour comparer deux moyennes		
Ou			
Test F d'analyse de la variance pour comparer plusieurs moyennes			

Réponses aux Quizz

Question 1. La bonne réponse est la réponse « c ». La cible de l'étude est quantitative, ce qui a amené les auteurs à comparer deux moyennes. Si le succès était la cible, les auteurs auraient comparé deux pourcentages de succès. L'outil pédagogique numérique n'est qu'une modalité du type d'apprentissage qui est la variable qui sert de base aux comparaisons.

Question 2. Le test U de Mann-Whitney ne sert à comparer que deux moyennes. C'est la réponse « a » qui est exacte.

Question 3. Par l'utilisation d'un calculateur (comme le logiciel gratuit OpenEpi), on peut montrer que les deux variances ne sont pas significativement différentes ($p = 0.313$). L'égalité des variances est considérée comme une des conditions d'application du test t mais elle n'est pas la plus importante. Les auteurs se sont probablement rendu compte que le score à l'épreuve ne suivait pas une loi normale, ce qui les a dissuadés d'utiliser le test t qui est plus puissant que le test U. La bonne réponse est donc la réponse « b ».

CONCLUSION

La comparaison de deux ou plusieurs moyennes avec les tests non paramétriques, le test U de Mann-Whitney et le test H de Kruskal-Wallis, a des indications très larges et peut être préconisée sans modération surtout lorsque la référence à la loi normale pour le caractère numérique est abusive. La complexité des calculs ne doit en aucune façon être un prétexte pour ne pas effectuer ces tests statistiques et bien d'autres tests encore du fait de la disponibilité d'excellents logiciels. Ces outils facilitent énormément la tâche de l'investigateur qui se pose des questions précises auxquelles il veut apporter des réponses précises.

Pour en savoir plus:

1. Khalfallah M, Dougaz W, Jerraya H, Samaali I, Mazigh S, Loueslati MH, et al. Self-directed learning digital tool versus tutorials under the guidance of an educator: Randomized trial. *Tunis Med* 2019; 97(2):296-303.
2. Bezzaouha A. Tests statistiques en sciences médicales (2^{ème} édition). Alger : Office des Publications Universitaires, 2017.
3. Ancelle T. Statistique Epidémiologie (4^{ème} édition). Paris : Maloine, 2017.