Master 1 Mathématiques Appliquées et Statistique **2021-2022**
parcours MSS et CMI

TP1 - Statistique Non Paramétrique

Tests de Wilcoxon / Mann et Whitney et de Kruskal-Wallis

Exercice 1 **Tests statistiques de Wilcoxon / Mann et Whitney**

Soient les échantillons indépendants

$X = (4.5, 6.6, 7, 6.7, 3.9, 5.7, 5.2)$,

$Y = (4.2, 5.3, 6.4, 5.1, 4.8)$,

Dans un premier temps, on va tester l'homogénéité (*test bilatéral*) des échantillons X et Y avec le test exact à la main, puis comparer le résultat avec celle de la fonction *wilcox.test* dans R.

(a) Calculer à la main l'observation de la statistique W des rangs du test de Wilcoxon.

(b) Conclure le test pour alpha=0,05, à l'aide d'une table statistique.

(c) Comparer avec le résultat de la fonction wilcox.test de R.

Histoire : Le test de Wilcoxon / Mann et Whitney a été proposé par Frank Wilcoxon en 1945 et détaillé par Henry Mann et Donald Ransom Whitney en 1947. Dans leurs recherches, différentes statistiques de test s'emploient mais elles se basent toutes sur les rangs.

Soient les échantillons $(X_1, ..., X_{n_1})$ et $(Y_1, ..., Y_{n_2})$, avec $n_2 \leq n_1$:

- La statistique W de Wilcoxon : $W = \sum_{i2=1}^{n_2} R_{i2}$

- La statistique U de Mann et Whitney : $U = \sum_{i=1}^{n_2} \#\{X_{j1} : X_{j1} < Y_{i2}, j1 = 1, ..., n_1\}$

Leur relation : $U = W - n_2(n_2 + 1)/2$.

Celle utilisée dans R est U. Etant donée l'observation de la W précedante, on a donc l'observation de la statistique U

$\texttt{U} = \texttt{W} - \texttt{n2} * (\texttt{n2} + \texttt{1})/\texttt{2}$

Complèter les argument dans la fonction wilcox.test ci-dessous à l'aide de la page

https ://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/wilcox.test

puis effectuer le test exact de wilcoxon / mann et whitney.

wilcox.test(Y,X,alternative = ,exact=)

NB : on doit placer le petit échantillon en premier, ici Y.

(d) On va maintenant tester l'homogénéité (*test bilatéral*) de X et Y avec le test asymptotique.

Générer un échantillon de taille $n_1 = 80$ de variables aléatoires i.i.d. selon une loi continue de votre choix. En générer un second de taille $n_2 = 50$ selon une autre loi continue. Tester l'homogénéité de ces deux échantillons à l'aide du test de Mann-Whitney-Wilcoxon.

(e) calculer W à la main, son espérance EW et sa variance VW sous H0.

(f) Calculer à la main les valeurs critiques pour W à $\alpha = 0,05$ (rappeler la loi asymptotique de W sous H0) et conclure le test.

(g) Calculer la p-valeur sans correction de continuité de Yates.
Rappel : Pour le test bilatéral, la p-valeur est donnée par

$$p = 2min\{P(T \geq t|H_0), P(T \leq t|H_0)\}$$

où T est la statistique de test et t est sa valuer observée avec l'échantillon. Si de plus la fonction de répartition est symétrique par rapport à **zero** sous H0 :

$$p = P(|T| \geq |t||H_0) \tag{0.1}$$

(h) Retrouver tous ces résultats avec R et la fonction wilcox.test.

(i) Comparer à un test de Student (test sur les moyennes, modèle gaussien) en supposant les conditions d'applications remplies (indépendance, normalité, homoscédasticité).

Exercice 2 **Tests statistiques de Kruskal-Wallis**

Etant donnés les échantillons X et Y de l'exercice 1, nous considérons un autre échantillon d'observations iid
$Z = (7.2, 6.8, 5.6, 5.9, 8.5)$, et on suppose que les 3 échantillons X, Y, Z sont indépendents. On va tester globalement l'homogénéité des trois échantillons comme suit.

(a) Calculer avec R les rangs moyens des échantillons X, Y et Z quand ils sont mélangés, puis calculer la statistique de Kruskal-Wallis.

(b) Test exact (alpha=0,05) : comparer la valeur de KW avec une table, voir par exemple si-dessous, et conclure quant le test de Kruskal-Wallis
`https://www.dataanalytics.org.uk/critical-values-for-the-kruskal-wallis-test/#grp-6-9`

(c) Test asymptotique (alpha=0,05) : calculer la valeur critique avec la loi asymptotique de la statistique de Kruskal-Wallis et conclure le test de Kruskal-Wallis

(d) Comparer le résultat du test asymptotique avec le test effectué dans R ci-dessous

(e) On considère maintenant les 3 échantillons (indépendants) suivants :
X = (4.5, 6.6, 7, 6.7, 3.9, 5.7, 5.2),
Y =(4.2,5.3,6.4,5.2, 4.8),
Z = (7.7, 7.3, 6.1, 6.4,9.0).
Vérifier que l'on rejette H0 avec la fonction Kruskal-Wallis de R au niveau $\alpha = 0.05$. Ainsi au moins deux des 3 échantillons n'ont pas la même médiane. Que peut-on faire pour préciser cette réponse ? Proposer des idées (sans les effectuer explicitement faute de temps).