L2 - Psychologie 2019-2020 - Deux exercices sur les tests de normalité et tests non paramétriques

Novembre 2019

Tous les tests demandés sont à effectuer au niveau d'erreur 5%.

Exercice 1. Le tableau suivant contient les résultats de deux groupes :

Groupe 1	7	18	9	9	18	27	12	10	32	6	37
Groupe 2	10	13	13	15	17	10	10	15	4	24	

1. Effectuer un test de normalité pour chacun des deux groupes.
2. Effectuer un test non-paramétrique de comparaison des deux groupes.

Exercice 2. Le tableau suivant donne une mesure de l'état de dépression d'un groupe de 10 patients, avant et après 3 mois de traitement :

AVANT: X	296	376	309	222	150	316	321	447	220	375
APRÈS: Y	175	329	238	60	271	291	364	402	70	335
D										

Décider, à l'aide d'un test non paramétrique de comparaison au risque d'erreur 5%, si les résultats sont significativement différents avant et après le traitement.

Corrigé de l'Exercice 1. 1. Effectuer un test de normalité pour chacun des deux groupes.

GROUPE 1.

Hypothèses. H_0 : les résultats du groupe 1 sont distribués selon une loi normale. H_1 : les résultats du groupe 1 ne sont pas distribués selon une loi normale.

Statistique du test. Sous l'hypothèse nulle, la variable aléatoire $W = \dfrac{\sum a_i d_i}{ns^2} \hookrightarrow$ Loi de Shapiro-Wilk, où les a_i sont les coefficients de Shapiro-Wilk, les d_i sont les différences emboîtées aléatoires, n est la taille de l'échantillon et s son écart-type. Ici, $n = 11$, $s = 10.223$, donc $W = \dfrac{\sum a_i d_i}{1149.6363}$.

Région critique K_α ($\alpha = 0.05$). D'après la table pour $n = 11$, $w_\alpha = 0.855$. Donc $K_\alpha = \{W \leq 0.855\}$.

Décision. On a, après avoir rangé les valeurs : 6,7,9,9,10,12,18,18,27,32,37

	37	32	27	18	18	
Valeurs rangées	6	7	9	9	10	12
Différences emboîtées d_i	31	25	18	8	8	

Donc $w_e = 0.866 > 0.855$, $w_\alpha \notin K_\alpha$. Au niveau d'erreur $\alpha = 0.05$, on conserve donc l'hypothèse de normalité des résultats du Groupe 1.

GROUPE 2.

Hypothèses. H_0 : les résultats du groupe 2 sont distribués selon une loi normale. H_1 : les résultats du groupe 2 ne sont pas distribués selon une loi normale.

Statistique du test. Sous l'hypothèse nulle, la variable aléatoire $W = \dfrac{\sum a_i d_i}{ns^2} \hookrightarrow$ Loi de Shapiro-Wilk, où les a_i sont les coefficients de Shapiro-Wilk, les d_i sont les différences emboîtées aléatoires, n est la taille de l'échantillon et s son écart-type. Ici, $n = 10$, $s = 5.0289$, donc $W = \dfrac{\sum a_i d_i}{252.899}$.

Région critique K_α ($\alpha = 0.05$). D'après la table pour $n = 10$, $w_\alpha = 0.845$. Donc $K_\alpha = \{W \leq 0.845\}$.

Décision. On a, après avoir rangé les valeurs : 4,10,10,10,13,13,15,15,17,24

	24	17	15	15	13
Valeurs rangées	4	10	10	10	13
Différences emboîtées d_i	20	7	5	5	0

Donc $w_e = 0.945 > w_\alpha$. Au niveau d'erreur $\alpha = 0.05$, on conserve l'hypothèse de normalité pour le Groupe 2.

2. Effectuer un test non-paramétrique de comparaison des deux groupes. On procède au test de comparaison de Mann-Whitney.

Hypothèses. H_0 : les deux distributions ne diffèrent pas significativement. H_1 : les deux distributions diffèrent significativement.

Statistique du test. $n_1 = 11$, $n = 10$. On note S_1 et S_2 la somme des rangs de distributions aléatoires de chacun des deux groupes. Alors, sous l'hypothèse nulle, les variables

$$U_1 = S_1 - \frac{n_1(n_1+1)}{2} = S_1 - 66 \text{ et } U_2 = S_2 - \frac{n_2(n_2+1)}{2} = S_2 - 55$$

suivent la loi de Mann-Whitney.

Région critique K_α ($\alpha = 0.05$). Sur la table de Mann-Whitney, avec $|n_1 - n_2| = |11 - 10| = 1$, et $\min(n_1, n_2) = \min(11, 10) = 10$, on trouve $u_\alpha = 26$. Donc

$$K_\alpha = \{U \leq 26\}.$$

Décision. On doit calculer la valeur expérimentale u_e. Pour cela, il faut ranger ensemble les valeurs des deux distributions.

Groupe 1	7	18	9	9	18	27	12	10	32	6	37	Somme des rangs
Groupe 2	10	13	13	15	17	10	10	15	4	24		
Rangs du Groupe 1	3	16.5	4.5	4.5	16.5	19	10	7.5	20	2	21	$S_{1\mathrm{e}} = 124.5$
Rangs du Groupe 2	7.5	11.5	11.5	13.5	15	7.5	7.5	13.5	1	18		$S_{2\mathrm{e}} = 106.5$

$S_{1\mathrm{e}} = 124.5$, $S_{2\mathrm{e}} = 106.5$, donc $u_{1\mathrm{e}} = 124.5 - 66 = 58.5$ et $u_{2\mathrm{e}} = 106.5 - 55 = 51.5$. Donc $u_\mathrm{e} = \min(u_{1\mathrm{e}}, u_{2\mathrm{e}}) = 51.5$. Ainsi $u_\mathrm{e} = 51.5 > 26$ et $u_\mathrm{e} \notin K_\alpha$: au niveau d'erreur $\alpha = 0.05$, on conserve l'hypothèse H_0, qui affirme que les deux distributions ne sont pas significativement différentes.

Corrigé de l'Exercice 2. Il s'agit de procéder à un test *non paramétrique* de comparaison de deux échantillons *appariés*, on procède donc au ***test de Wilcoxon***.

Hypothèses. H_0 : les deux distributions ne diffèrent pas significativement. H_1 : les deux distributions diffèrent significativement.

Statistique du test sous H_0. $n' =$ nombre de différences non nulles des deux échantillons $= 10$. Pour un échantillon aléatoire, on désigne par Σ^+ la somme des rangs des valeurs absolues des différences positives et par Σ^- la somme des rangs des valeurs absolues des différences négatives. Sous l'hypohèse nulle, la variable $W = \min(\Sigma^+, \Sigma^-)$ vérifie :

$$W = \min\left(\Sigma^+, \Sigma^-\right) \hookrightarrow \text{Loi de Wilcoxon avec } n' = 10.$$

Région critique K_α ($\alpha = 0.05$). Dans la table de Wilcoxon, pour $n' = 10$, on trouve $w_\alpha = 8$. Donc

$$K_\alpha = \{W \leq 8\}.$$

Décision du test. Il s'agit de calculer w_e. Pour cela, on calcule les rangs des valeurs absolues des différences positives et négatives :

AVANT: X	296	376	309	222	150	316	321	447	220	375
APRÈS: Y	175	329	238	60	271	291	364	402	70	335
$D = X - Y$	121	47	71	162	-121	25	-43	45	150	40
Rangs de $\lvert X - Y\rvert$	7.5	5	6	10	7.5	1	3	4	9	2

Donc

$$\Sigma_\mathrm{e}^+ = 7.5 + 5 + 6 + 10 + 1 + 4 + 9 + 2 = 44.5 \text{ et } \Sigma_\mathrm{e}^- = 7.5 + 3 = 10.5\,.$$

Ainsi $w_\mathrm{e} = \min(\Sigma_\mathrm{e}^+, \Sigma_\mathrm{e}^-) = 10.5 > 8$, donc $w_\mathrm{e} \notin K_\alpha$. Au niveau d'erreur $\alpha = 0.08$, on accepte l'hypothèse nulle H_0 : les deux distributions ne sont pas significativement différentes.