

Probabilités

Licence 2 - Semestre 4

Exercices d'entraînement

Tests non paramétriques

Corrigés

Correction 1

On commence par ranger les notes par ordre croissant.

On a $x_{(i)} = \{3, 4, 7, 8, 8, 9, 10, 10, 10, 11, 12, 13, 15, 17, 18\}$

La moyenne est $\bar{x} = \frac{3 + 4 + \dots + 18}{15} = \frac{155}{15} = \frac{31}{3} \approx 10,33$.

La moyenne des carrés est $\overline{x^2} = \frac{3^2 + 4^2 + \dots + 18^2}{15} = \frac{1855}{15} = \frac{371}{3} \approx 123,67$.

On obtient donc la variance $s^2 = \frac{371}{3} - \left(\frac{31}{3}\right)^2 = \frac{152}{9} \approx 16,89$

et la variance corrigée $s_c^2 = \frac{15}{14}s^2 \approx 18,1$.

Puisque $15/2 = 7,5$, on considère les 7 premiers et les 7 derniers éléments.

$x_{(i)}$	3	4	7	8	8	9	10	...
$x_{(i)}^2$	9	16	49	64	64	81	100	...
$x_{(n+1-i)}$	18	17	15	13	12	11	10	...
d_i	15	13	8	5	4	2	2	...
a_i	0,5150	0,3306	0,2495	0,1878	0,1353	0,0880	0,0433	...
$a_i \times d_i$	7,725	4,2978	1,996	0,939	0,5412	0,176	0	

$$t^2 = \frac{1}{14} \times (7,725 + 4,2978 + 1,996 + 0,939 + 0,5412 + 0,176 + 0)^2$$

$$= \frac{1}{14} \times (15,675)^2 \approx 17,55$$

Et $w = \frac{t^2}{s_c^2} \approx 0,97$

D'après la table de Shapiro-Wilk avec $\alpha = 0,05$ et $n = 15$, on trouve $w_{\min} = 0,881$.

Puisque $w > w_{\min}$, on ne rejette pas H_0 , la population peut être supposée gaussienne à un niveau de confiance de 95%.

Correction 2

Soient μ_1 la consommation moyenne des véhicules de type A et μ_2 la consommation moyenne des véhicules de type B.

L'hypothèse nulle est $H_0 : \mu_1 = \mu_2$ et l'hypothèse alternative est $H_1 : \mu_1 \neq \mu_2$.

Type A	5,2	5,4	5,7	6	6,5	7,2	8,1		
Rang	4	5	7	9,5	12	15	16		
Type B	4,9	5,1	5,1	5,5	5,8	6	6,2	6,8	6,9
Rang	1	2,5	2,5	6	8	9,5	11	13	14

On a $n_1 = 7$ et $n_2 = 9$ soit un effectif total de $n = 7 + 9 = 16$.

De plus, $r_1 = 68,5$ et $r_2 = 67,5$.

On peut vérifier que $r_1 + r_2 = 136 = \frac{16 \times 17}{2}$.

On obtient $u_1 = \frac{7 \times 26}{2} - 68,5 = 22,5$ et $u_2 = \frac{9 \times 24}{2} - 67,5 = 40,5$.

Donc $u = 22,5$.

D'après la table de Mann et Whitney avec $\alpha = 0,05$, $n_1 = 7$ et $n_2 = 9$, on trouve $u_{\min} = 12$.

Puisque $u > u_{\min}$, on ne peut pas rejeter l'hypothèse nulle au risque α .

Correction 3

On veut effectuer le test de $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$.

Ayant de petits échantillons appariés de taille $n = 10$, et sans hypothèse de loi normale sur la variable étudiée, on réalise un test de Wilcoxon.

On calcule les différences entre les valeurs appariées. On supprime ensuite les différences nulles.

Et on note N le nombre de différences non nulles.

Sportif	1	2	3	4	5	6	7	8	9	10
Matin	75	85	83	80	80	76	82	87	81	77
Après-midi	77	85	81	80	79	80	81	88	81	80
	2	0	-2	0	-1	4	-1	1	0	3

On classe ces $N = 7$ différences non nulles par ordre croissant de valeurs absolues, et on affecte à chaque différence son rang dans ce classement. En cas d'ex-æquo, on prend un rang moyen.

différences	-1	-1	1	-2	2	3	4
rangs provisoires	1	2	3	4	5	6	7
rangs moyens	2	2	2	4,5	4,5	6	7

On calcule les sommes :

- des rangs des différences négatives : $w_{neg} = 2 + 2 + 4,5 = 8,5$.
- des rangs des différences positives : $w_{pos} = 2 + 4,5 + 6 + 7 = 19,5$

On peut vérifier que $w_{pos} + w_{neg} = 8,5 + 19,5 = 28 = \frac{7 \times 8}{2} = \frac{N(N+1)}{2}$.

On a donc $w = 8,5$ comme la plus petite des deux valeurs w_{pos} et w_{neg} .

La table de Wilcoxon avec $N = 7$ et $\alpha = 0,05$ donne $w_{\min} = 2$.

Puisque $w > w_{\min}$, donc on ne rejette pas H_0 au risque 5% : la différence des moyennes n'est pas significative. Autrement dit, pas de différence significative entre les performances des sportifs le matin ou l'après midi.

Correction 4

On veut effectuer le test de $H_0 : \mu_1 = \mu_2 = \mu_3$ (égalité des moyennes des ventes des commerciaux) contre $H_1 = \overline{H_0}$ c'est-à-dire au moins deux de ces moyennes ne sont pas égales.

Ayant de petits échantillons indépendants et sans hypothèse de loi normale sur la variable étudiée, on réalise un test de Kruskal et Wallis.

On affecte à chaque valeur son rang dans le classement par ordre croissant de l'ensemble des valeurs observées des k échantillons

						Total
Vendeur A :	2	2	5	4	1	
Rangs	6	6	14,5	12,5	3	42
Vendeur B :	3	0	1	2	4	
Rangs	9,5	1	3	6	12,5	32
Vendeur C :	5	1	3	3	3	
Rangs	14,5	3	9,5	9,5	9,5	46

On a donc $r_1 = 42$, $r_2 = 32$ et $r_3 = 46$.

De plus $n_1 = 5$, $n_2 = 5$ et $n_3 = 5$ soit $n = 15$.

On remarquera que $r_1 + r_2 + r_3 = 120 = \frac{15 \times 16}{2} = \frac{n \times (n+1)}{2}$.

D'où $h = \frac{12}{15 \times 16} \left(\frac{42^2}{5} + \frac{32^2}{5} + \frac{46^2}{5} \right) - 3 \times 16 \approx 1,04$.

Grâce à la table de Kruskal et Wallis avec $\alpha = 0,05$ et les n_i on trouve la valeur plafond $h_{\max} = 7,72$.

Comme $h < h_{\max}$, on ne peut pas rejeter H_0 au risque $\alpha = 0,05$: il n'y a pas de différence significative du nombre moyen de ventes.

Correction 5

Sans hypothèse sur les lois, on utilise le test de Spearman.

										Total
Margarine	4,7	4,6	4,4	4,3	4,1	4,2	4,2	4,2	4,1	
Rangs x'_i	9	8	7	6	1,5	4	4	4	1,5	45
Divorces	7	6,5	5,3	5,2	4	4,6	4,5	4,2	3,7	
Rangs y'_i	9	8	7	6	2	5	4	3	1	45
x_i^2	81	64	49	36	2,25	16	16	16	2,25	282,5
y_i^2	81	64	49	36	4	25	16	9	1	285
$x_i \times y_i$	81	64	49	36	3	20	16	12	1,5	282,5

On a donc $E(x') = 5 = E(y')$, $E(x'^2) = \frac{282,5}{9} \approx 31,39$ et $E(y'^2) = \frac{285}{9} = 31,67$.

On obtient $V(x') \approx 6,389$ et $V(y') \approx 6,67$.

De plus, $\text{cov}(x', y') = E(x' \times y') - E(x') \times E(y') = \frac{282,5}{9} - 5 \times 5 = 6,389$.

Et $r_S = \frac{\text{cov}(x', y')}{\sqrt{V(x') \times V(y')}} \approx 0,979$.

Puisque $n = 9 < 10$, grâce à la table de Spearman avec $\alpha = 0,05$, on trouve $r_{\max} = 0,68$.

Puisque $|r_S| < r_{\max}$, on ne peut pas rejeter H_0 au risque $\alpha = 0,05$: il y a une corrélation significative entre la consommation de margarine aux U.S. et le taux de divorce dans le Maine sur la période considérée.

On remarquera que l'on a la même conclusion si $\alpha = 0,01$!