Exercice 1:

Une entreprise pharmaceutique désire tester l'effet de 2 somnifères (Sopo et Dodo). Le médecin chargé de l'étude vous apporte les mesures (nombre d'heures de sommeil additionnel par rapport à la moyenne habituelle) effectuées sur 20 individus.

Individu	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sopo	**-0.5**	**-0.1**	**0.9**	**1**	**1.5**	**1.8**	**1.9**	**3.1**	**4.5**	**4.8**

Individu	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Dodo	**-1.3**	**00**	**0.1**	**1.4**	**1.2**	**0.4**	**1.1**	**2.1**	**2.3**	**3.2**

On veut tester l'égalité des effets des 2 somnifères.

1) Définissez le test que vous pouvez faire dans ce cas, en posant les conditions nécessaires, ainsi que l'hypothèse nulle et l'hypothèse alternative. Donnez la région critique du test.
   Solution
   On pose
   - X la variable nombre d'heures de sommeil additionnel par rapport à la moyenne habituelle chez les personnes qui prennent le somnifère Sopo,
   - Y la variable nombre d'heures de sommeil additionnel par rapport à la moyenne habituelle chez les personnes qui prennent le somnifère Dodo .

On dispose de deux échantillons indépendants, un issu de X et l'autre issu de Y.

On peut appliquer le test de la somme des rangs de Wilcoxon (ou le test de MannWhitney) avec les conditions : X et Y sont des v.a indépendantes absolument continues de fonctions de répartition$F$ et G identiques à une translation prés .

Hypothèse $H_0: F = G$, Hypothèse $H_1: F \neq G$

(Hypothèse $H_0: M_1 = M_2$, Hypothèse $H_1: M_1 \neq M_2$)

Région critique : $W = Min(W_1, W_2) \leq w_\alpha$où $W_i$est la somme des rangs des observations du groupe$i$, $i = 1,2$.

2) Effectuez le test pour un seuil de $\alpha = 5\%$. Concluez.

z_i	-0.5	-0.1	0.9	1	1.5	1.8	1.9	3.1	4.5	4.8	-1.3	00	0.1	1.4	1.2	0.4	1.1	2.1	2.3	3.2
$z_{(i)}$	-1.3	-0.5	-0.1	0	0.1	0.4	0.9	1	1.1	1.2	1.4	1.5	1.8	1.9	2.1	2.3	3.1	3.2	4.5	4.8
$rg(z_i)$	2	3	7	8	12	13	14	17	19	20	1	4	5	11	10	6	9	15	16	18

$$w_1 = 2 + 3 + 7 + 8 + 12 + 13 + 14 + 17 + 19 + 20 = 115$$

$$w_2 = \frac{20(21)}{2} - w_1 = 95$$

Pour $\alpha = 5\%, n = m = 10$, on lit$w_{0.05} = 78$.

$w = \min(w_1, w_2) = 95 > w_{0.05}$ donc on accepte H_0.

3) En présentant les résultats de l'étude au médecin qui vous a chargé de la mener, vous réalisez que vous avez mal compris et qu'il s'agit en fait d'une expérience où chacun des 10 individus a testé chacun des 2 somnifères, ce qui donne le tableau suivant :

Individu	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sopo	**-0.5**	**-0.1**	**0.9**	**1,0**	**1.5**	**1.8**	**1.9**	**3.1**	**4.5**	**4.8**
Dodo	**-1.3**	**00**	**0.1**	**1.4**	**1.2**	**0.4**	**1.1**	**2.1**	**2.3**	**3.2**

Ainsi on veut tester de nouveau l'égalité de l'effet des 2 somnifères. Effectuez le test au seuil $\alpha = 5\%$. Donnez toutes les étapes du test.

Solution

Comme les deux échantillons proviennent d'un même groupe de malades, et que chaque malade a testé chacun des deux somnifères, alors le test adéquat sera le test de signe pour données appariées ;

On a un n-échantillon $((X_1, Y_1) \ldots (X_n, Y_n))$, issu du couple de v.a$(X, Y)$avec $= 10$.

On détermine les différences $D_i = X_i - Y_i$, $i = 1, \ldots, 10$, on suppose que les D_i sont i.i.d. de fonction de répartition F_D continue, et on effectue un test de signe sur la médiane M_D de la différence.

Le but est de tester l'égalité de l'effet des deux somnifères pris par les 10 malades au seuil α, alors on pose $D = X - Y$ et on teste :

$$H_0 : M_D = 0 \quad vs \quad H_1 : M_D \neq 0$$

Individu	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$\boldsymbol{d_i = x_i - y_i}$	**0.8**	**-0.1**	**0.8**	**-0.4**	**0.3**	**1.4**	**0.8**	**1,0**	**2.2**	**1.6**
Signe (d_i-0)	+	-	+	-	+	+	+	+	+	+

La statistique du test de signe est :

$$T = \min(T^+; T^-)$$

avec $T^+ = \sum_{i=1}^{10} \mathbf{1}_{\{D_i > 0\}}$ et $T^- = \sum_{i=1}^{10} \mathbf{1}_{\{D_i < 0\}}\}$ qui suivent sous H_0 une loi $B(10, 0.5)$.

La région critique du test : $T \leq t^*{}_\alpha$

On a $T^+(obs) = t^+ = 8$, $T^-(obs) = t^- = 2$ et $T(\text{obs}) = t = \min(t^+, t^-) = t^- = 2$.

De la table on lit $t^*{}_{0,05} = 1\ pour\ n = 10$.

Comme $t > t^*{}_{0,05}$ alors on accepte H_0 (égalité de l'effet des 2 somnifères)

Avec la p-valeur:

$$\alpha_0 = 2Min\left(P_{H_0}(T^+ \geq 8), P_{H_0}(T^+ \leq 8)\right) = 2Min\left(1 - F_{B(10,½)}(7), F_{B(10,½)}(8)\right)$$
$$= 2Min(1 - 0.9453, 0.9893)$$
$$= 0.1094 > 0.05$$

Donc on accepte H_0 .

Remarque : On peut appliquer aussi le test des rangs signés.

Exercice 2:

Un fabricant d'un certain type de tubes de radio affirme que la médiane de leur durée de vie est égale à 70 heures. On soupçonne qu'elle est moindre. Un échantillon de 12 tubes a été testé et on a obtenu les résultats suivants :

62.4	53.3	57.8	69	49.3	70.7	76.8	55.3	66.8	69.1	73.4	64.3

1) Préciser les hypothèses nulle et alternative (à tester au niveau $\alpha = 0.05$).
2) On propose d'appliquer le test des signes
   i) Donner la statistique du test et préciser sa loi sous l'hypothèse nulle ?
   ii) Donner la région critique. Conclure
   iii) Calculer la p-valeur.

Solution

Soit X la durée de vie d'un tube de radio. On suppose que X est absolument continue de médiane M.

- Hypothèses du test

Hypothèse nulle: $H_0: M = 70$,

Hypothèse alternative: $H_1: M < 70$.

- Statistique du test: $T^+ = \sum_{i=}^{12} 1_{\{X_i > 70\}}$

  Sous H_0, $T^+ \rightsquigarrow B(12, \frac{1}{2})$.

- Région critique: $T^+ \leq t_\alpha$ avec $P_{H_0}(T^+ \leq t_\alpha) \leq \alpha$.

  De la table du test des signes, on lit $t_{0.05} = 2$.

  D'où la région critique : $T^+ \leq 2$.

x_i	62.4	53.3	57.8	69	49.3	70.7	76.8	55.3	66.8	69.1	73.4	64.3
Signe($x_i - 70$)	−	−	−	−	−	+	+	−	−	−	+	−

  On a $T^+{}_{obs} = t^+ = 3 > 2$ d'où on accepte H_0.

- La p-valeur est donnée par $P_{H_0}(T^+ \leq 3) = 0.073 > 0.05$.

3) On veut maintenant appliquer le test des rangs signés de Wilcoxon.
   i) Dresser le tableau complet nécessaire à la réalisation du test.
   ii) Déterminer la statistique du test et donner sa moyenne et sa variance sous H_0 .
   iii) Déterminer la région critique. Conclure.

Solution

On suppose de plus que la loi de X est symétrique.

On considère les différences $D_i = X_i - 70$ et $R_i^+ = rg(|D_i|)$, $i = \overline{1, \dots, 12}$.

- Tableau nécessaire à la réalisation du test des rangs signés de Wilcoxon

x_i	62.4	53.3	57.8	69	49.3	70.7	76.8	55.3	66.8	69.1	73.4	64.3
$d_i = x_i - 70$	-7.6	-16.7	-12.2	-1	-20.7	0.7	6.8	-14.7	-3.2	-0.9	3.4	-5.7
$z_i = \|d_i\|$	7.6	16.7	12.2	1	20.7	0.7	6.8	14.7	3.2	0.9	3.4	5.7
$z_{(i)}$	0.7	0.9	1	3.2	3.4	5.7	6.8	7.6	12.2	14.7	16.7	20.7
$rg(z_i)$	8	11	9	3	12	1	7	10	4	2	5	6
Signe(d_i)	−	−	−	−	−	+	+	−	−	−	+	−

- Statistique du test : $W^+ = \sum_{i=1}^{12} R_i^+ 1_{\{D_i>0\}}$
  Sous $H_0, on\ a$

$$E(W^+) = \frac{n(n+1)}{4} = \frac{12(12+1)}{4} = 39$$

$$V(W^+) = \frac{n(n+1)(2n+1)}{24} = 162.5$$

- Région critique : $\{W^+ \leq w_\alpha\}$ avec $P_{H_0}(W^+ \leq w_\alpha) \leq \alpha$
  De la table on lit $w_{0.05} = 17$.
  La région critique est donc : $\{W^+ \leq 17\}$.
  On a $W^+{}_{obs} = w^+$=1+7+5=13<17 donc on rejette H_0.

Exercice 3 :

On a chargé un médecin de répondre à la question suivante : Aspirine (AAS) diminue –elle l'espérance de vie chez les patients asthmatiques ? Ce médecin a récolté des données selon les critères suivants : individu asthmatique est décédé de manière naturelle au cours des 5 dernières années. Les informations retenues sont l'âge au décès et si l'aspirine a été recommandée au patient (oui=O et non=N). Le tableau suivant présente un échantillon aléatoire des milliers de réponses obtenues. La distribution des données est aléatoire.

Age du décès	AAS
45.6	O
45.85	O
48.45	O
48.63	O
48.74	N
49.6	N
51.4	O
60.86	N
52.06	O
53.16	N
54	O
65.16	N
56.93	N
57.38	O
57.94	N
67.96	N
58.24	O
69.7	N
51.48	O
51.56	O
55.18	O
55.32	N
57.8	O
58.59	O
58.63	N

58.89	O
59.18	O
59.24	O
60.53	O
64.86	N
65.81	N
67.72	O
68.8	N
69.58	N
72.66	N

1) Définissez le test que vous pouvez faire dans ce cas, en posant les conditions nécessaires, ainsi que l'hypothèse nulle et l'hypothèse alternative. Donnez la statistique et la région critique du test.
2) Effectuez le test pour un seuil de $\alpha = 5\%$. Concluez.

Solution

1) Le test qu'on peut effectuer dans ce cas est le test de la somme des rangs de Wilcoxon qui est basé sur les rangs des individus des échantillons $(X_1 \cdots X_n)$ et $(Y_1 \cdots Y_m)$ issus respectivement de X et Y où

- X : l'âge du décès d'un individu asthmatique n'ayant pas pris de l'aspirine
- Y : l'âge du décès d'un individu asthmatique ayant pris de l'aspirine

Les conditions nécessaires sont :

Xet$Y$ sont absolument continues de fonction de répartition F et G (resp) tq :

$$F(x) = G(x - \theta), \forall x.$$

Si μ_1 et μ_2 désignent respectivement l'âge moyen de décès d'un individu n'ayant pas pris de l'aspirine et d'un individu ayant pris de l'aspirine alors on teste

$$H_0 : \mu_1 = \mu_2 vs H_1 : \mu_1 > \mu_2.$$

La statistique du test :

Les tailles des échantillons sont $n = 16$ et $m = 19$ (resp) considérées comme assez grandes $(n > 8 \ et \ m > 8)$. On peut effectuer le test en utilisant la statistique

$$Z = \frac{W_1 - E(W_1)}{\sqrt{var(W_1)}} \rightsquigarrow N(0,1)$$

avec$W_1 = \sum_{i=1}^{n+m} iD_i$ et $D_i = \begin{cases} 1 \ si Z_{(i)} \ estun X \\ 0 \ si Z_{(i)} estun Y \end{cases} i = \overline{1, n+m}$

Et $\left(Z_{(1)}, \ldots ; Z_{(n+m)}\right)$ est l'échantillon ordonné des deux échantillons $(X_1 \cdots X_n)$ et $(Y_1 \cdots Y_m)$ regroupés. Autrement dit est la somme des rangs des X_i dans $\left(Z_{(1)}, \ldots ; Z_{(n+m)}\right)$.

La région critique du test

$$\{Z \geq Z_{1-\alpha}\} \quad \text{tq} \quad P_{H_0}(Z \geq Z_{1-\alpha}) = \alpha.$$

2) On souhaite tester au seuil $\alpha = 0.05$ $H_0: \mu_1 = \mu_2 vs H_1: \mu_1 > \mu_2$. On regroupe les deux échantillons et on réordonne les 35 observations par ordre croissant, on obtient

i	z_i	$z_{(i)}$
1	45.6	45.6
2	45.85	45.85
3	48.45	48.45
4	48.63	48.63
5	48.74	48.74
6	49.6	49.6
7	51.4	51.4
8	60.86	51.48
9	52.06	51.56
10	53.16	52.06
11	54	53.16
12	65.16	54
13	56.93	55.18
14	57.38	55.32
15	57.94	56.93
16	67.96	57.38
17	58.24	57.8
18	69.7	57.94
19	51.48	58.24
20	51.56	58.59
21	55.18	58.63
22	55.32	58.89
23	57.8	59.18
24	58.59	59.24
25	58.63	60.53
26	58.89	60.86
27	59.18	64.86
28	59.24	65.16
29	60.53	65.81
30	64.86	67.72
31	65.81	67.96
32	67.72	68.8
33	68.8	69.58
34	69.58	69.7
35	72.66	72.66

- Les observations du deuxième échantillon (Y)
- Les observations du premier échantillon (X)

$W_{1(obs)} = w_1 = 5 + 6 + 11 + 14 + 15 + 18 + 21 + 26 + 27 + 28 + 29 + 31 + 32 + 33 + 34 + 35 = 365.$

Si H_0est vraie alors :

$E(W_1) = \frac{n(n+m+1)}{2} = \frac{16(36)}{2} = 288 \text{et} Var(W_1) = \frac{nm(n+m+1)}{12} = 912.$

$Z_{obs} = z = \frac{w_1 - E(W_1)}{\sqrt{var(W_1)}} = \frac{365-288}{\sqrt{912}} = 2.54 \text{avec } z_{0.95} = 1.645\ .$

Comme $z > 1.645$ alors on rejetteH_0 au seuil 0.05.

Conclusion : l'aspirine diminue l'espérance de vie des personnes asthmatiques.

Exercice 4:

On dispose de deux échantillons de tailles respectives 10 et15 :

Echantillon 1 (X)	80	100	90	110	125	130	70	75	71	83					
Echantillon 2 (Y)	100	120	80	140	130	160	115	120	73	88	135	125	128	95	87.

Tester au seuil 5% si les deux échantillons proviennent de la même population. Quelle hypothèse faites-vous ?

1. **Test de la somme des rangs de Wilcoxon**
   On suppose que X et Y sont deux v.a. continues de fonction de répartition respectives F et G identiques à une translation prés$(\boldsymbol{F(x) = G(x-\theta), \forall x})$.
   On veut tester $H_0: F = G$ v.s $H_1: F \neq G$ $\quad (H_0: M_1 = M_2 v.s H_1: M_1 \neq M_2)$

On réordonne les 25 observations par ordre croissant, on obtient

i	z_i	$z_{(i)}$	rang de z_i	d_i
1	80	70	5.5	1
2	100	71	12.5	1
3	90	73	10	0
4	110	75	14	1
5	125	80	18.5	0
6	130	80	21.5	1
7	70	83	1	1
8	75	87	4	0
9	71	88	2	1
10	83	90	7	1
11	100	95	12.5	0
12	120	100	16.5	0
13	80	100	5.5	1
14	140	110	24	1
15	130	115	21.5	0
16	160	120	25	0
17	115	120	15	1
18	120	125	16.5	1
19	73	125	3	0
20	88	128	9	0
21	135	130	23	0
22	125	130	18.5	1
23	128	135	20	0
24	95	140	11	0
25	87	160	8	0

- Les observations du premier échantillon (X)
- Les observations du deuxième échantillon (Y)

- les ex-aequos

avec$d_i = \begin{cases} 1\ si Z_{(i)}\ est un X \\ 0\ si Z_{(i)} est un Y \end{cases} \overline{i = 1, n+m}$, $n = 10$, $m = 15$.

La somme des rangs des individus du premier groupe est :

$w_1 = 5.5+12.5+10+14+18.5+21.5+1+4+2+7 = 96$.

La somme des rangs des individus du deuxième groupe est :

$$w_2 = \frac{(n+m)(n+m+1)}{2} - w_1 = \frac{25(26)}{2} - 96 = 229$$

Ou bien :

w_2 =12.5+16.5+5.5+24+21.5+25+15+16.5+3+9+23+18.5+20+11+8 = 229

Soit H_0: $M_1 = M_2$. Si H_0est vraie alors :

$$E(W_1) = \frac{n(n+m+1)}{2} = \frac{10(26)}{2} = 130$$

Le tableau contient 5 paires d'ex- aequo alors :

$$Var(W_1) = \frac{nm(n+m+1)}{12} - \frac{nm}{12(n+m)(n+m+1)} \sum t_l\,(t_l^2 - 1)$$
$$= \frac{150(26)}{12} - \frac{150}{12(25)(24)} \big(2(4-1)\big) \times 5$$
$$Var(W_1) = 324.375.$$

- **Approximation normale sans correction de continuité**

On a $n = 10 > 8 et m = 15 > 8$, donc on utilise la statistique $Z = \frac{W_1 - E(W_1)}{\sqrt{Var(W_1)}}$qui suit une loi normale centrée réduite.
Comme le test est bilatéral alors la région critique est de la forme : $|Z| \geq z_{1-\frac{\alpha}{2}}$.
$Z_{obs} = z = \frac{96-130}{\sqrt{324.375}} = -1.88$ et $z_{0.975} = 1.96$pour $\alpha = 0.05$. Par suite $|z| < z_{0.975}$et on accepte H_0.
Ceci est confirmé par la p-valeur :
$\alpha_0 = 2\min\big(\phi(z), 1-\phi(z)\big) = 2min\big(\phi(-1.88), 1-\phi(-1.88)\big) = 0.0602 > 0.05$.

- **Approximation normale avec correction de continuité**

$Region\ critique$: $W_1 \leq z_{\frac{a}{2}}\sqrt{Var(W_1)} - 0.5 + E(W_1)$ ou $W_1 \geq z_{1-a/2}\sqrt{Var(W_1)} + 0.5 + E(W_1)$

$$z_{0.025}\sqrt{Var(W_1)} - 0.5 + E(W_1) = 94.19; \quad z_{1-\frac{a}{2}}\sqrt{Var(W_1)} + 0.5 + E(W_1) = 165.8.$$

On remarque que $\mathbf{94.19 < w_1 < 165.8}$donc on accepte H_0. Ceci est confirmé par la p-valeur

$$\alpha_0 = 2\min\left(\phi\left(\frac{\boldsymbol{w_1} + 0.5 - E(W_1)}{\sqrt{Var(W_1)}}\right), 1 - \phi\left(\frac{\boldsymbol{w_1} - 0.5 - E(W_1)}{\sqrt{Var(W_1)}}\right)\right)$$

$$= 2\min\big(\phi(1.91), 1 - \phi(1.86)\big)$$

$$= 0.0628 \quad > 0.05.$$

2. Test de Mann-Whitney

z_i	$z_{(i)}$	
80	70	X
100	71	X
90	73	Y
110	75	X
125	80	Y
130	80	X
70	83	X
75	87	Y
71	88	X
83	90	X
100	95	Y
120	100	Y
80	100	X
140	110	X
130	115	Y
160	120	Y
115	120	Y
120	125	X
73	125	Y
88	128	Y
135	130	Y
125	130	X
128	135	Y
95	140	Y
87	160	Y

- Les observations du premier échantillon (X)
- Les observations du deuxième échantillon (Y)
- Les ex-aequo entre les deux échantillons
- Les ex-aequo dans le même échantillon

Comme on est dans le cas d'ex-aequo, la statistique du test est :

$$U_T = \sum_{i=1}^{n}\sum_{j=1}^{m} D_{ij}^{*}$$

Où $$D_{ij}^{*} = \begin{cases} 1 & si\ Y_i < X_i \\ 0 & si\ Y_i = X_i \\ -1 & si\ \ Y_i > X_i \end{cases}$$

Sous H_0; on a :$E(U_T) = 0$.

Le tableau contient 5 paires d'ex- aequo alors (sousH_0) :

$$Var(U_T) = \frac{nm(n+m+1)}{12} - \frac{nm}{12(n+m)(n+m+1)} \sum t_l\,(t_l^2 - 1)$$
$$= \frac{150(26)}{12} - \frac{150}{12(25)(24)}\big(2(4-1)\big) \times 5$$

$Var(U_T) = 324.375.$

- **Approximation sans correction de continuité**

Comme $n, msont > 8$, on peut utiliser l'approximation normale. La statistique du test Zest alors définie par

$$Z = \frac{U_T - E(U_T)}{\sqrt{var(U_T)}} \rightsquigarrow N(0,1).$$

La région critique est de la forme : $|Z| > z_{0.975}$

$$Z_{obs} = z = \frac{-59 - 0}{\sqrt{324.375}} = -3.27$$

On a $z_{0.975} = 1.96 < |z|$ donc on rejette H_0.

Ceci est confirmé par la p-value

$$\alpha_0 = 2\min\big(\phi(z), 1 - \phi(z)\big) = 2\min\big(\phi(-3.27), 1 - \phi(-3.27)\big) = 0.00107 < 0.05.$$

- **Approximation avec correction de continuité**

La région critique**:** $U_T \le z_{\alpha/2}\sqrt{Var(U_T)} - 0.5$ $\quad ou U_T \ge z_{1-\alpha/2}\sqrt{var(U_T)} + 0.5.$

$$z_{\alpha/2}\sqrt{Var(U_T)} - 0.5 = -35.80\,; \quad z_{1-\alpha/2}\sqrt{var(U_T)} + 0.5 = 35.80.$$

$Comme\ u_T = -59 < -35.80$ alors on rejette H_0.

La p-valeur est donnée par

$$a_0 = 2\min\left(1 - \Phi\left(\frac{u_T - 0.5}{\sqrt{var(U_T)}}\right), \Phi\left(\frac{u_T + 0.5}{\sqrt{var(U_T)}}\right)\right) = 2\min\big(1 - \Phi(-3.30), \Phi(-3.24)\big) = 0.0011952$$
$< 0.05.$

<u>Exercice 5 :</u> Comparaison des hauteurs des arbres de trois types de forêts : test de Kruskal-Wallis ($\alpha = 5\%$)

Dans trois types de forêts, on a mesuré les hauteurs respectivement de 13, 14 et 10 peuplements, choisis au hasard et indépendamment. On désire tester l'hypothèse d'égalité des distributions des hauteurs des arbres des trois forêts. Les observations sont données dans la table suivante :

Type 1	Type 2	Type 3
23.4	**22.5**	**18.9**
24.4	**22.9**	**21.1**
24.6	**23.7**	**21.2**
24.9	**24**	**22.1**
25	**24.4**	**22.5**
26.2	**24.5**	**23.6**
26.3	**25.3**	**24.5**
26.8	**26**	**24.6**
26.8	**26.2**	**26.2**
26.9	**26.4**	**26.7**
27	**26.7**	
27.6	**26.9**	
27.7	**27.4**	
	28.5	

Solution$K = 3, = 37\ , n_1 = 13\ ,\ n_2 = 14\ , n_3 = 10$

- Les rangs des 37 observations sont :

Type 1	**Type 2**	**Type 3**
8	5.5	1
12.5	7	2
16.5	10	3
18	11	4
19	12.5	5.5
23	14.5	9
25	20	14.5
29.5	21	16.5
29.5	23	23
31.5	26	27.5
33	27.5	
35	31.5	
36	34	
	37	

r_i	316.5	280.5	106	
$\frac{r_i^2}{n_i}$	7705.55	5620.01	1123.6	$\sum_{i=1}^{3} \frac{r_i^2}{n_i}$ =14449.16

- La valeur de la statistique de Kruskal –Wallis non corrigée est :

$$H_{obs} = \frac{12}{N(N+1)} \sum_{i=1}^{3} \frac{r_i^2}{n_i} - 3(N+1) = \frac{12}{37(38)} 14449.16 - 3(38) = 9.3214$$

- Le facteur de correction (facteur d'ajustement) est :

$$c = 1 - \frac{\sum_{l=1}^{28} t_l(t_l^2 - 1)}{N(N^2 - 1)}$$

Les observations sont réparties en 28 ensembles d'ex-aequo (28 valeurs distinctes). On note t_l le nombre d'ex-aequo ayant le l −ième rang :

Rang l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
t_l	1	1	1	1	2	1	1	1	1	1	2	2	2	1	1	1	1	3	1	1	2	2	2

rang l	24	25	26	27	28
t_l	1	1	1	1	1

$$\sum_{l=1}^{28} t_l(t_l^2 - 1) = 7 * 2(2^2 - 1) + 3(3^2 - 1) = 66$$

$$c = 1 - \frac{66}{37(37^2 - 1)}$$

$$c = 0.998$$

- La statistique du test ajustée :

$$\widetilde{H}_{obs} = \frac{H_{obs}}{c} = \frac{9.3214}{0.998} = 9.340$$

Sous H_0, la statistique ajustée (corrigée) $\widetilde{H}$suit une loi du χ^2à 2 degrés de liberté (car les $n_i > 5$).

- La région critique est $\widetilde{H} \geq \chi^2_{1-\alpha}(2)$

De la table de χ^2, on lit $\chi^2_{0.95}(2) = 5.991$

Comme $9.340 > 5.991$ alors on rejette l'hypothèse nulle, ce résultat est confirmé par la p-valeur $P_{H_0}(\widetilde{H} \geq 9.340) = 0.00937 < 0.05$

- Il y a C_3^2 3 comparaisons à réaliser.

Le test Kruskal-Wallis est significatif au seuil α=0.05 .Pour les comparions deux à deux (qui sont au nombre 3), nous calculons le risque

$$a = \frac{\alpha}{K(K-1)} = \frac{0.05}{3(3-1)} = \frac{0.05}{6} = 0.00833$$

La quantile d'ordre $1 - a = 0.9916$ de la loi normale centrée reduite est $z_{0.9916} = 2.39$

Les moyennes conditionnelles des rangs sont données dans le tableau suivant :

r_i	316.5	280.5	106
$\bar{r}_i = \frac{r_i}{n_i}$	24.346	20.03	10.6

Les calculs requis pour les comparaisons sont résumés dans le tableau suivant :

$\bar{r}_i$	$\bar{r}_j$	$\lvert\bar{r}_i - \bar{r}_j\rvert$	$z_{0.9916}\sqrt{\frac{37(38)}{12}(\frac{1}{n_i} + \frac{1}{n_j})}$	Significatif
24.34	20.03	4.31	9.932	Non

24.34	10.6	13.74	10.85	Oui
20.03	10.6	9.43	10.71	non

Il y a un écart significatif entre les auteurs des arbres des forêts 1 et 3.

Exercice 6

On désire savoir s'il existe un éventuel lien entre la taille et le poids. Pour ce faire, on a prélevé le poids et la taille de 10 collégiens. Les résultats sont présentés dans le tableau suivant :

Collégien	1	2	3	4	5	6	7	8	9	10
Taille	1.69	1.53	1.62	1.63	1.50	1.67	1.54	1.57	1.49	1.55
Poids	77.5	55.0	76.6	62.5	58.0	72.5	58.5	70.0	67.5	61.5

1) Donner la statistique du test de Spearman.
2) Peut-on conclure qu'il existe une corrélation positive entre les deux variables. Tester au niveau α =0.05.
3) Calculer la *p*-valeur du test.
4) Répondre aux questions 2 et 3 en utilisant la distribution asymptotique de la statistique du test.

Solution:

Soient les deux variables : X taille ,Y poids d'un collégien.

1) La statistique de Spearman

$$R_s = \frac{12}{n(n^2-1)}\sum_{i=1}^{n} R_i S_i - \frac{3(n+1)}{n-1}$$

ou

$$\boldsymbol{R_S = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n} D_i{}^2}$$

$\boldsymbol{avec D_i = R_i - S_i}$**,** $\boldsymbol{R_i}$ **étant le rang de** $\boldsymbol{X_i}$ **et** $\boldsymbol{S_i}$ **étant le rang de** $\boldsymbol{Y_i}$ **;**
$\boldsymbol{(X_i, Y_i), i = 1, \dots, n}$**unn-échantillon de** $\boldsymbol{(X, Y)}$.

2) On veut tester

$$\boldsymbol{H_0 : X \text{ et } Y \text{ sont indépendantes}}$$
$$\boldsymbol{H_1 : X \text{ et } Y \text{ sont positivement dépendantes}}$$

- On calcule le coefficient de corrélation des rangs de Spearman :

Taille(m)	**Poids (kg)**	$\boldsymbol{r_i}$	$\boldsymbol{s_i}$	$\boldsymbol{r_i s_i}$	$\boldsymbol{d_i = r_i - s_i}$	$\boldsymbol{d_i^2}$
1.69	**77.5**	**10**	**10**	**100**	**0**	**0**
1.53	**55**	**3**	**1**	**3**	**2**	**4**
1.62	**76.6**	**7**	**9**	**63**	**-2**	**4**
1.63	**62.5**	**8**	**5**	**40**	**3**	**9**
1.50	**58**	**2**	**2**	**4**	**0**	**0**
1.67	**72.5**	**9**	**8**	**72**	**1**	**1**
1.54	**58.5**	**4**	**3**	**12**	**1**	**1**
1.57	**70.0**	**6**	**7**	**42**	**-1**	**1**
1.49	**67.5**	**1**	**6**	**6**	**-5**	**25**
1.55	**61.5**	**5**	**4**	**20**	**1**	**1**
				$\boldsymbol{\sum_{i=1}^{10} r_i s_i = 362}$		$\boldsymbol{\sum_{i=1}^{10} d_i = 46}$

Avec la première formule :

$$r_s = \frac{12}{n(n^2-1)}\sum_{i=1}^{10} r_i s_i - \frac{3(n+1)}{n-1} = \frac{12(362)}{10(99)} - \frac{3(11)}{9} \quad = 4.38 - 3.66 = 0.72$$

Avec la deuxièmeformule :

$$r_s = 1 - \frac{6}{n(n^2-1)}\sum_{\hat{i}=1}^{10} d_i^2 = 1 - \frac{6}{10(99)}46 = 0.721$$

- On Teste au niveau$\alpha = 0.05$.

Il s'agit d'un test unilatéral à droite, la région critique est donc de la forme$\{R_s \geq r_{0.05}\}$ avec

$$P_{H_0}(R_s \geq r_{0.05}) \leq 0.05\ .$$

De la table de Spearman, on lit, pour $n = 10, r_{0.05} = 0.564$ (lecture de la table M)

Et comme $r_s > r_{0.05}$ alors on rejette H_0, c.à.d. il y a une dépendance positive entre la taille et le poids

3) La p-valeur

$$\alpha_0 = P(R_s > 0{,}72) = 0.012.$$

4) Approximation par la loi normale :

$Z = \frac{R_s}{\sqrt{Var(R_s)}} = R_s\sqrt{n-1} = R_s\sqrt{9} = 3R_s$suit une loi normale centrée réduite.

$$z = 3r_s = 0.72(3) = 2.16$$

- La région de rejet est de la forme :$Z \geq z_{1-\alpha}$
  $Commez_{1-\alpha} = z_{0.95} = 1.64$, alors$z > 1.64$ donc on rejette H_0
- La p_valeur :

  $$\alpha_0 = 1 - \Phi(3r)$$

  On calcule :

  $$\Phi(3r) = \phi(0.72(3)) = \phi(2.16) = 0.9846$$

  Par suite

  $$\alpha_0 = 0.0154$$

- Approximation par la loi de student :

  $T = \frac{R_s\sqrt{n-2}}{\sqrt{1-R_s^2}}$suit une loi de Student à $n-2 = 8$ degrés de liberté.

  $$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{0.72\sqrt{8}}{\sqrt{1-0.72^2}} = \frac{2.036}{0.69} = 2.934$$

- la région de rejet est de la forme : $T \geq t_{1-\alpha}(n-2)$.
  $t_{0.95}(8) = 1.8595$ (Lecture de la table de student)
  On a $t > 1.8595$ donc on rejette H_0
- la p valeur :

  $$\alpha_0 = 1 - F_{st(n-2)}\left(\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right)$$

  On calcule

  $$F_{st(n-2)}\left(\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right) = F_{st(8)}(2.934) = 0.99.$$

Exercice 7:

A partir de la vente de 100 postes de télévision ayant fonctionné le même nombre d'heures, pendant une année, on a pu établir le tableau suivant contenant le nombre de répartition de ces téléviseurs :

Nombre de réparations (x_i)	0	1	2	3
Nombre de téléviseurs (n_i)	61	30	7	2

Peut-on ajuster la distribution observée à la loi de Poisson de paramètre $\lambda = 1$?

Solution : ajustement à une loi de poisson

X:nombre de réparations

On veut tester$H_0: X \rightsquigarrow \mathcal{P}(1)$ v.s. $H_1: X$ ne suit pas la loi de Poisson

$$H_0:\ p_i = P(X = i) = \frac{e^{-1}}{i!}, i \geq 0.$$

i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
0	61	0.367	36.7	
1	30	0.367	36.7	
2	7	0.183	18.3	
3	2	0.061	6.1	regrouper
≥ 4	0	0.015	1.5<5	

i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
0	61	0.367	36.7	16.08
1	30	0.367	36.7	1.22
2	7	0.183	18.3	6.97
≥ 3	2	0.076	7.6	4.12

$$\sum \frac{(n_i - np_i)^2}{np_i} = 28.39$$

La statistique du test suit sous H_0 une χ^2_3 car r=0 et k=4 donc k-1=3.

Comme $\chi^2_3(0.95) = 7.815 < 28.39$ alors on rejette H_0.

Exercice 8:

On a mesuré la taille en cm de 200 étudiants et on a obtenu les résultats suivants :

Taille	[134,142[	[142,150[	[150,158[	[158,166[	[166,174[	[174,182[
Effectif	16	44	60	44	26	10

Peut-on ajuster cette répartition observée à la loi Normale $N(m, \sigma^2)$, au risque de 10% ?

Solution :

X : Taille d'un étudiant, $n = 200$.

On veut tester $H_0: X \rightsquigarrow N(m, \sigma^2)$ v.s. $H_1: X$ ne suit pas la loi normale.

Comme $m\ et\ \sigma^2$ sont inconnus, on les estime respectivement par

$$\widehat{m} = \frac{1}{n}\sum n_i x_i = \frac{1}{200}(31200) = 156$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum n_i (x_i - \widehat{m})^2$$

$$= \frac{1}{200}(16(138-156)^2 + 44(146-156)^2 + 60(154-156)^2 + 44(162-156)^2 + 26(170-156)^2 + 10(178-156)^2) = 106.72$$

oùx_i est le centre de la ième classe.

Sous $H_0, \frac{X-\widehat{m}}{\sqrt{\widehat{\sigma^2}}}$ suit une loi normale centrée réduite et

$$p_i = P(a_i \leq X < a_{i+1}) = \Phi\left(\frac{a_{i+1}-\widehat{m}}{\sqrt{\widehat{\sigma^2}}}\right) - \Phi\left(\frac{a_i-\widehat{m}}{\sqrt{\widehat{\sigma^2}}}\right)$$

$$p_1 = P(X < 134) = \Phi\left(\frac{134-156}{\sqrt{106.72}}\right) = \Phi(-2.12) = 1 - \Phi(2.12) = 1 - 0.9830 = 0.017$$

$$p_2 = P(134 \leq X < 142) = \Phi(-1.35) - \Phi(-2.12) = -\Phi(1.35) + \Phi(2.12) = 0.0715$$

$$p_3 = P(142 \leq X < 150) = \Phi(-0.58) - \Phi(-1.35) = 1 - \Phi(0.58) - 1 + \Phi(1.35) = \Phi(1.35) - \Phi(0.58) = 0.9115 - 0.7190 = 0.1925$$

$$p_4 = P(150 < X < 158) = \Phi(0.19) - \Phi(-0.58) = 05753 - 1 + 0.7190 = 0.2943$$

$$p_5 = P(158 \leq X < 166) = \Phi(0.96) - \Phi(0.19) = 0.8315 - 0.5753 = 0.2562$$

$$p_6 = P(166 \leq X < 174) = \Phi(1.74) - \Phi(0.96) = 0.9591 - 0.8315 = 0.1276$$

$$p_7 = P(174 \leq X < 182) = \Phi(2.51) - \Phi(1.74) = 0.9940 - 0.9591 = 0.0349$$

$$p_8 = P(X \geq 182) = 1 - \Phi(2.51) = 0.006$$

Classe i	x_i	n_i	$n_i x_i$	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
< 134		0	0	0.017	3.4<5	
[134,142[	138	16	2208	0.0715	14.3	regrouper
[142,150[	146	44	6424	0.1925	38.5	
[150,158[	154	60	9240	0.2943	58.86	
[158,166[	162	44	7128	0.2562	51.24	
[166,174[	170	26	4420	0.1276	25.52	
[174,182[	178	10	1780	0.0349	6.98	regrouper
≥ 182		0	0	0.006	1.2<5	

Classe i	x_i	n_i	$n_i x_i$	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
<142		16	2208	0.0885	17.7	0.163
[142,150[	146	44	6424	0.1925	38.5	0.785
[150,158[	154	60	9240	0.2943	58.86	0.022
[158,166[	162	44	7128	0.2587	51.24	1.022
[166,174[	170	26	4420	0.1276	25.52	0.009
≥ 174		10	1780	0.0409	8.18	0.404
$\sum$		200	31200	1	200	2.405

La valeur de la statistique du χ^2 : $d^2 = \sum \frac{(n_i - np_i)^2}{np_i} = 2.405$

La statistique du test suit sous H_0 une khi-deux de (6-1-2) degré de liberté χ_3^2 :

Pour $\alpha = 10\%$, $\chi_3^2(1 - 0.1) = \chi_3^2(0.9) = 6.251$.

Comme$d^2 < \chi_5^2(0.9)$, on accepte H_0la taille d'un étudiant est normalement distribuée.