

Généralités sur les tests

Tests de signe, de Wilcoxon et de Mann-Whitney

Table des matières

1 Généralités sur les tests	1
1.1 Rappels : déroulement d'un test	1
1.2 Exemple d'un test paramétrique	1
2 Statistique de rangs	2
3 Tests de signe et tests de Wilcoxon pour deux échantillons appariés	3
4 Tests de Mann Whitney pour deux échantillons indépendants	3
5 TP 2	4

1 Généralités sur les tests

1.1 Rappels : déroulement d'un test

1. Préciser les hypothèses à tester H_0 et H_1
2. Chercher la statistique du test $h(X)$ (X étant le vecteur des observations). Déterminer sa loi ou sa loi asymptotique sous H_0 . Cette loi doit être **libre** des paramètres inconnus.
3. Préciser l'allure de la zone de rejet de l'hypothèse H_0 (cette zone de rejet dépend de l'hypothèse H_1) : cela revient à voir le comportement de $h(X)$ sous H_1).

ensuite

4. Calcul de la zone de rejet R_α en fonction d'un niveau α déterminé par le statisticien (typiquement $\alpha = 0.05$).
5. Observation de la réalisation $h(x)$ de $h(X)$ et détermination de son appartenance ou non dans la zone de rejet.

ou bien

- 4bis. Observation de la réalisation $h(x)$ de $h(X)$.
- 5bis. Calcul de la p -valeur $\hat{\alpha}(x)$ associée à cette observation $h(x)$ et comparaison à niveau α fixé par un responsable.

Conclusion

6. Conservation ou rejet de H_0 . Commentaires sur le crédit d'accorder ou pas aux résultats en fonction de la p -valeur et de la puissance du test.

1.2 Exemple d'un test paramétrique

On suppose avoir des observations $X = (X_1, \dots, X_n)$ indépendantes et identiquement distribuées selon la loi normale $\mathcal{N}(\theta, 1)$ pour un paramètre $\theta \in \mathbb{R}$ inconnu. On veut tester $H_0 : \theta \geq 1$ contre $H_1 : \theta < 1$.

1. Justifier pourquoi un test de niveau (et même de taille) α est donné par

$$\phi_\alpha(X) = I_{\bar{X}_n < k_\alpha}$$

2. Montrer que $k_\alpha = 1 + q_\alpha/\sqrt{n}$ où q_α est le quantile d'une loi (libre des observations) que l'on déterminera.
3. Donner l'expression de la p -valeur du test associée à l'observation X .
4. On suppose que $n = 4$ et que l'on observe les réalisations

$$x_1 = 0.15, x_2 = 0.45, x_3 = 0.9, x_4 = 0.5$$

Conclure au niveau $\alpha = 0.1$. Calculer la p -valeur associée à l'observation (x_1, x_2, x_3, x_4) .

2 Statistique de rangs

Exercice 1.

Soit $X = (X_1, \dots, X_n)$ n variables aléatoires réelles iid. On rappelle que le vecteur des rangs R_X est une permutation de $\{1, \dots, n\}$ telle que pour tout $i \in \{1, \dots, n\}$ $X_i = X_{(R_X(i))}$.

1. Donner une condition portant sur les (X_i) sous laquelle le vecteur rang R_X est unique.
2. Dans la suite on supposera que cette condition est remplie. Soit σ une permutation de $\{1, \dots, n\}$. Montrer que les 3 événements suivants sont équivalents :

$$\begin{aligned} (R_X = \sigma) \\ (X_1 = X_{(\sigma(1))}, \dots, X_n = X_{(\sigma(n))}) \\ (X_{\sigma^{-1}(j)} = X_{(j)}, 1 \leq j \leq n) \\ (X_{\sigma^{-1}(1)} < X_{\sigma^{-1}(2)} < \dots < X_{\sigma^{-1}(n)}) \end{aligned}$$

3. Montrer que pour une permutation τ de $\{1, \dots, n\}$ les deux vecteurs (X_1, \dots, X_n) et $(X_{\tau(1)}, \dots, X_{\tau(n)})$ ont la même loi.
4. En déduite que $\mathbb{P}(R_X = \sigma)$ est une constante ne dépendant pas de σ et par conséquent

$$\mathbb{P}(R_X = \sigma) = \frac{1}{n!}.$$

5. R_X étant le vecteur $(R_X(1), R_X(2), \dots, R_X(n))$. Montrer que la loi suivie par chaque marginale $R_X(i)$ est la loi uniforme sur l'ensemble $\{1, \dots, n\}$. En déduire que

$$\mathbb{E}(R_X(i)) = \frac{n+1}{2}, \text{Var}(R_X(i)) = \frac{n^2-1}{12}.$$

6. Donner la loi du couple $(R_X(i), R_X(j))$ pour $i \neq j$. En déduire

$$\text{Cov}(R_X(i), R_X(j)) = -\frac{n+1}{12}$$

7. Que vaut $\sum_{i=1}^n \phi(R_X(i))$ pour une fonction mesurable ϕ donnée ? Peut-on déduire un résultat analogue pour $\sum_{i=1}^n \phi_i(R_X(i))$ pour une suite de fonctions mesurables ϕ_i , $1 \leq i \leq n$?
8. Montrer que $X^* = (X_{(1)}, \dots, X_{(n)})$ et R_X sont deux vecteurs indépendants.

On rappelle que

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

3 Tests de signe et tests de Wilcoxon pour deux échantillons appariés

Exercice 2.

Soient (U, V) un couple de variable aléatoires. On suppose que (U, V) et (V, U) ont la même loi. On pose $X = U - V$.

1. Montrer que la loi de X est symétrique.
2. On suppose que $\mathbb{P}(U = V) = 0$. Montrer dans ce cas que $\mathbb{P}(X \geq 0) = \mathbb{P}(X \leq 0) = 1/2$.

Exercice 3.

Soient X_1, \dots, X_n des va i.i.d. de loi symétrique. On suppose que $\mathbb{P}(|X_i| = |X_j|) = 0$ pour $i \neq j$. On définit :

$$W_n^+ = \sum_{i=1}^n R_{|X|}(i) I_{X_i > 0}, \quad W_n^- = \sum_{i=1}^n R_{|X|}(i) I_{X_i < 0}$$

$R_{|X|}$ étant le vecteur rang associé aux $(|X_i|)_{1 \leq i \leq n}$.

1. Montrer qu'il existe une permutation σ de $\{1, \dots, n\}$ telle qu'on ait presque sûrement,

$$W_n^+ = \sum_{j=1}^n j I_{X_{\sigma(j)} > 0}$$

2. Soit Y_1, \dots, Y_n une suite de va iid de loi de Bernoulli de paramètre $1/2$. Montrer que W_n^+ et $\sum_{j=1}^n j Y_j$ ont la même loi.
3. En déduire que $\mathbb{E}(W_n^+) = n(n+1)/4$ et que $\text{Var}(W_n^+) = n(n+1)(2n+1)/24$.
4. Montrer que W_n^+ et W_n^- ont la même loi.
5. Que vaut $W_n^+ + W_n^-$?
6. Montrer que la loi de W_n^+ est symétrique autour de son espérance.

4 Tests de Mann Whitney pour deux échantillons indépendants

Exercice 4.

Soit $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_p)$ deux échantillons **indépendants** de lois diffuses (c'est-à-dire $\mathbb{P}(X_i = x) = \mathbb{P}(Y_j = x) = 0$ pour tous x). On suppose que X et Y ont la même loi. On considère $\{X, Y\}$ l'échantillon global et on note par R_1, \dots, R_n les rangs des va X_i dans l'échantillon global et par S_1, \dots, S_p ceux de (Y_i) dans l'échantillon global. On note par,

$$W_X = \sum_{i=1}^n R_i - \frac{n(n+1)}{2}.$$

1. Montrer que W_X est une variable aléatoire discrète à valeurs dans $\{0, \dots, np\}$
2. Montrer que $\mathbb{E}(W_X) = np/2$ et que $\text{Var}(W_X) = np(n+p+1)/12$.
3. Montrer que, presque sûrement, $W_X = \sum \sum_{i,j} I_{Y_j < X_i}$.
4. On définit de manière analogue $W_Y = \sum_{i=1}^p S_i - p(p+1)/2$. Montrer que W_Y et W_X ont la même loi.
5. Montrer que, presque sûrement,

$$W_X + W_Y = np.$$

6. Montrer que la loi de W_X est symétrique autour de son espérance.

Exercice 5.

On a noté l'âge d'apparition d'une maladie cardiaque chez un certain nombre de sujets. On souhaite tester si l'âge moyen d'apparition de la maladie est le même chez les hommes et chez les femmes. On dispose des résultats suivants :

Hommes	19	26	30	23	17	20
Femmes	21	25	29	31	33	

Modéliser le problème. Effectuer un test de Mann-Whitney.

5 TP 2

Exercice 6.

La quantité de bactéries par cm^3 de lait provenant de huit vaches différentes est estimée après la traite et 24 heures plus tard. On se demande s'il existe un accroissement significatif du nombre de bactéries par cm^3 au cours du temps.

Vache	Juste après la traite	24 h après la traite
1	12000	11 000
2	13000	20 000
3	21500	31 000
4	17000	28 000
5	15000	26 000
6	22000	30 000
7	11000	16 000
8	21000	29 000

Répondre à la question par un test du signe et par un test de Wilcoxon.

Exercice 7.

On veut comparer les taux de résistances de plaies soignées soit par un pansement, soit par des points de suture. On obtient les données suivantes sur 10 souris, 40 jours après que des incisions aient été faites sur leur dos, traitées les unes avec un pansement et les autres avec des points de suture.

Souris	1	2	3	4	5	6	7	8	9	10
Pansement	659	984	397	574	447	479	676	761	647	577
Points de suture	452	587	460	787	351	277	234	516	577	513

Comment réaliser la comparaison souhaitée ? En utilisant la fonction `wilcox.test` du logiciel R, quelle est la p -valeur du test effectué ?

Exercice 8.

On veut tester l'efficacité d'un nouveau médicament contre les migraines. On dispose d'un échantillon de 18 personnes sujettes aux migraines à qui on fournit une quantité égale de pilules correspondant au nouveau traitement (A) et de pilules d'aspirine (B). On demande aux patients de choisir une des deux pilules lors d'une crise et lorsqu'ils ont utilisé toutes les pilules, on leur demande de juger quel type de pilule (A ou B) a été le plus efficace. Sur les 18 patients, 12 déclarent que le nouveau traitement (A) est plus efficace que l'ancien (B). Comment tester l'efficacité du nouveau traitement ? Qu'en concluez-vous au niveau $\alpha = 0.05$?

Exercice 9.

On s'intéresse à l'effet d'une dose faible de Cambendazole sur les infections des souris par la *Trichinella Spiralis*. Seize souris ont été infectées par un même nombre de larves de *Trichinella* et ensuite réparties au hasard entre deux groupes. Le premier groupe de 7 souris a reçu du Cambendazole, à raison de 10 mg par kilo, 60 heures après l'infection. Les autres souris n'ont pas reçu de traitement. Au bout d'une semaine, toutes les souris ont été sacrifiées et le nombre suivant de vers adultes ont été retrouvés dans les intestins :

souris non traitées	51	55	62	63	65	68	71	75	79
souris traitées	47	49	53	57	60	61	67		

Décrire un protocole statistique pour tester une efficacité éventuelle du Cambendazole pour le traitement des infections des souris par la *Trichinella Spiralis*. En utilisant uniquement la table de Mann-Whitney donnée ci-dessous, pour laquelle valeur de niveau de test α est-on sûr que le test de Mann-Whitney conclura à une efficacité du Cambendazole pour le traitement des infections des souris par la *Trichinella Spiralis* ?

	1	2	3	4	5	6	7	8	9
1									
2								0	0
3					0	1	1	2	2
4				0	1	2	3	4	4
5			0	1	2	3	5	6	7
6			1	2	3	5	6	8	10
7			1	3	5	6	8	10	12
8		0	2	4	6	8	10	13	15
9		0	2	4	7	10	12	15	17

Table de Mann-Whitney pour la statistique W_X pour différentes valeurs de n et p (notations du cours). Le tableau donne la plus grande valeur v telle que $P(W_X \leq v) \leq 0.025$. Chaque ligne correspond une valeur de n , chaque colonne une valeur de p .