

M1 finance: Statistiques non paramétriques

Fiche 5: Comparaison de la médiane entre deux échantillons

11 décembre 2014

Exercice 1 : Révision : taille des élèves

On mesure la taille des élèves d'une classe de 3ème et on veut répondre à différentes questions. Pour chacune de ces situations, formulez H_0 et H_1 , d'abord par une phrase, ensuite mathématiquement. Précisez ensuite quel(s) test(s) on peut appliquer (le cas échéant, précisez les hypothèses supplémentaires que l'on doit faire). Donnez la formule théorique de la statistique de test et la forme de la zone de rejet.

1. On veut savoir si la taille suit une loi normale.
2. On suppose que cette mesure suit une loi normale. On veut savoir si la moyenne μ vaut 150cm.
3. On veut savoir si les filles sont plus grandes que les garçons (en moyenne).
4. On veut savoir si les garçons sont stochastiquement plus grands que les filles.
5. On veut savoir si les élèves ont grandi ou non entre le début et la fin de l'année scolaire.
6. On veut savoir si la taille suit une loi normale $\mathcal{N}(140, 100)$.
7. On veut savoir si la médiane vaut 140 cm.
8. On veut savoir si la moyenne vaut 150 cm.

Exercice 2 : Démographie

On mesure l'évolution de la population (en %) dans 9 cantons ruraux et 7 cantons urbains d'Alabama.

Évolution de la population dans les cantons ruraux : 1.1, -21.7, -16.3, -11.3, -10.4, -7.0, -2.0, 1.9, 6.2

Évolution de la population dans les cantons urbains : -2.4, 9.9, 14.2, 18.4, 20.1, 1.2, 70.4.

1. Quel(s) test(s) peut-on appliquer à ces données ?
2. On suppose que ces deux échantillons suivent la même loi, avec peut-être un décalage ($F_R(x - \theta) = F_U(x)$). On cherche c tel que $\theta \geq c$ à 95%.

Exercice 3 : Test d'un médicament

On voudrait savoir si un nouveau médicament est plus efficace qu'un placebo dans le traitement d'une certaine maladie. On trouve 10 patients souffrant de la même maladie, on donne à 5 patients un placebo, et à 5 autres le nouveau médicament. On mesure pour chaque patient la concentration dans le sang d'une protéine. Plus la concentration de cette protéine est faible, mieux c'est pour le patient. On observe :

Pour les patients traités avec le nouveau médicament : 3.2, 2.1, 2.3, 1.2, 1.5

Pour les patients qui n'ont que le placebo : 3.4, 3.5, 4.1, 1.7, 2.2

1. Quel(s) test(s) peut-on utiliser ici ?
2. On choisit d'utiliser le test de la médiane. Appliquer le test avec un niveau $\alpha = 5\%$. Quelle est la p -valeur ?
3. On souhaite maintenant appliquer le test de Wilcoxon somme et rang. Appliquer ce test avec un niveau $\alpha = 5\%$. Quelle est la p -valeur ?

Exercice 4 : Test de la médiane

1. Pour $m = 5$ et $n = 6$, donner la distribution complète de U .

2. Calculer la fonction de répartition de U .
3. On décide de rejeter si U est trop petit. Donner les zones de rejet à 10% et 5%.

Exercice 5 : Tests de Wilcoxon

Soient (Z_1, Z_2, \dots, Z_N) des variables aléatoires indépendantes et symétriques. La fonction de répartition F_X est continue. On pose $\begin{cases} X_i = Z_i & \text{si } Z_i > 0 \\ Y_i = |Z_i| & \text{si } Z_i < 0 \end{cases}$ On obtient deux échantillons indépendants : (X_1, \dots, X_m) et (Y_1, \dots, Y_n) avec $N = m + n$.

1. Montrer que la statistique de Wilcoxon somme et rang pour les échantillons X et Y est égale à la statistique de Wilcoxon signe et rang pour l'échantillon Z .
2. Soit l'échantillon $Z = (3, 4, -2, -5, 7, 8, -10, 1, 9, 6)$. Calculer les statistiques de Wilcoxon signe et rang et de Wilcoxon somme et rang. Regarder les p -valeurs dans les tables. Comment expliquez-vous ce paradoxe?
3. Comment peut-on déduire les tables de Wilcoxon signe et rang des tables de Wilcoxon somme et rang?