

Test non paramétrique

- Test des rangs signés de Wilcoxon -

Dr. Sacha Varone

Objectif

Savoir effectuer et interpréter un test sur la médiane

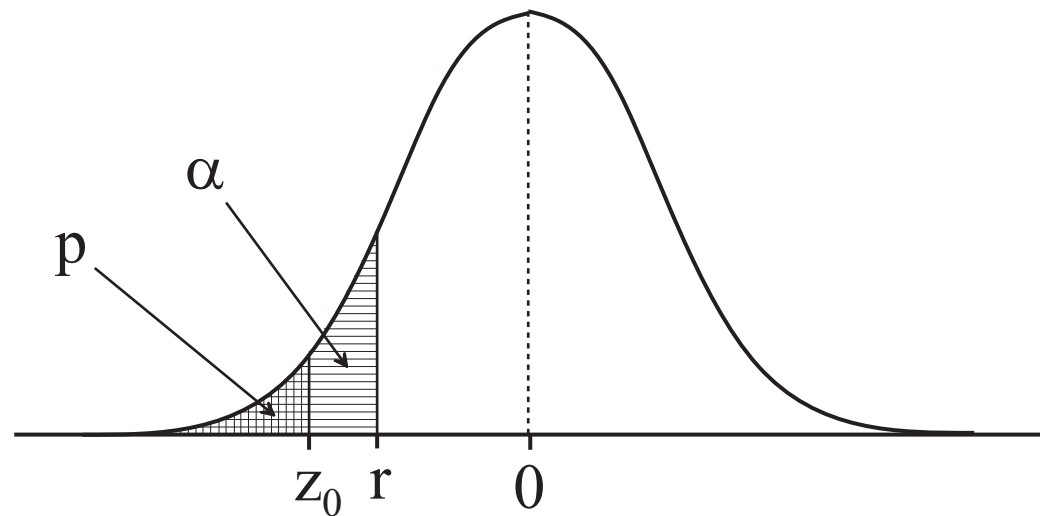
Rappels

Test de la médiane

Rappels

Méthode de la p -valeur

La p -valeur, appelée *niveau (ou degré) de signification observé*, est la probabilité d'observer l'échantillon réellement utilisé sachant que l'hypothèse nulle H_0 est vraie.



[4.] Rejeter H_0 si $p\text{-valeur} < \alpha$

Résumé

Rappels

p-valeur

Résumé

Test de la médiane

estimé	hypothèse	distribution
μ	σ^2 connu	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
	σ^2 inconnu, distr. normale	$T_{(n-1)} = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim \mathcal{T}_{n-1}$
π	$n\pi \geq 5$ et $n(1 - \pi) \geq 5$	$Z = \frac{\bar{P} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim \mathcal{N}(0, 1)$
σ^2	μ connu, distr. normale	$Q_{(n)} = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$
	μ inconnu, distr. normale	$Q_{(n-1)} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Test de la médiane

Tests non paramétriques

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Que faire si les hypothèses sur la distribution ne sont pas respectées ?

Tests non paramétriques

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Que faire si les hypothèses sur la distribution ne sont pas respectées ?

Test non paramétrique :

- moins restrictif
- mais moins puissant

Principe

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Principe :

Médiane =

valeur de séparation d'une population en deux parties de même taille.

Si la médiane hypothétique de la population est trop éloignée de la médiane de l'échantillon, alors H_0 rejeté.

-> Test des rangs signés de Wilcoxon

Exemple

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Une étude a été faite sur les salaires en début de carrière des diplômés de l'école "Jess Aitou", afin de savoir si le salaire médian est supérieur à 35000 Euros, à un seuil $\alpha = 0.05$. Les données récoltées sont les salaires en euros :

36400 38500 27000 35000 29000
40000 52000 34000 38900 41000

Exemple

Rappels

Test de la médiane

Tests non paramétriques

Principe

Exemple

Statistique

Grand échantillon

Une étude a été faite sur les salaires en début de carrière des diplômés de l'école "Jess Aitou", afin de savoir si le salaire médian est supérieur à 35000 Euros, à un seuil $\alpha = 0.05$. Les données récoltées sont les salaires en euros :

36400 38500 27000 35000 29000
40000 52000 34000 38900 41000

1. La valeur de la population d'intérêt est le salaire médian en début de carrière.

Exemple

Rappels

Test de la médiane

Tests non paramétriques

Principe

Exemple

Statistique

Grand échantillon

Une étude a été faite sur les salaires en début de carrière des diplômés de l'école "Jess Aitou", afin de savoir si le salaire médian est supérieur à 35000 Euros, à un seuil $\alpha = 0.05$. Les données récoltées sont les salaires en euros :

36400 38500 27000 35000 29000
40000 52000 34000 38900 41000

1. La valeur de la population d'intérêt est le salaire médian en début de carrière.
2. Les hypothèses nulle et alternative sont :
$$H_0 \quad \tilde{\mu} \leq 35'000 \text{ euros}$$
$$H_1 \quad \tilde{\mu} > 35'000 \text{ euros}$$

Exemple

Rappels

Test de la médiane

Tests non paramétriques

Principe

Exemple

Statistique

Grand échantillon

Une étude a été faite sur les salaires en début de carrière des diplômés de l'école "Jess Aitou", afin de savoir si le salaire médian est supérieur à 35000 Euros, à un seuil $\alpha = 0.05$. Les données récoltées sont les salaires en euros :

36400 38500 27000 35000 29000
40000 52000 34000 38900 41000

1. La valeur de la population d'intérêt est le salaire médian en début de carrière.
2. Les hypothèses nulle et alternative sont :
$$H_0 \quad \tilde{\mu} \leq 35'000 \text{ euros}$$
$$H_1 \quad \tilde{\mu} > 35'000 \text{ euros}$$
3. Le niveau de signification est de 0.05

Exemple

4. 5.

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Salaire = x_i	$d_i = x_i - \tilde{\mu}$	$ d_i $	Rang	R_+	R_-
36400	1400	1400	2	2	
38500	3500	3500	3	3	
27000	-8000	8000	8		8
35000	0	0			
29000	-6000	6000	6.5		6.5
40000	5000	5000	5	5	
52000	17000	17000	9	9	
34000	-1000	1000	1		1
38900	3900	3900	4	4	
41000	6000	6000	6.5	6.5	
				$W = 29.5$	15.5

Soit n le nombre de d_i différents de 0. Il y a donc $n = 9$ rangs attribués.

La valeur critique du test est $W_\alpha = 37$

Exemple

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

4. 5. La statistique est $W = 29.5$, la valeur critique est $W_\alpha = 37$

Exemple

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

4. 5. La statistique est $W = 29.5$, la valeur critique est $W_\alpha = 37$

6. Comme

$$W = 29.5 < 37$$

l'hypothèse H_0 ne peut pas être rejetée.

Exemple

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

4. 5. La statistique est $W = 29.5$, la valeur critique est $W_\alpha = 37$

6. Comme

$$W = 29.5 < 37$$

l'hypothèse H_0 ne peut pas être rejetée.

7. Nous ne pouvons donc pas conclure que le salaire médian en début de carrière dépasse 35000 euros.

Construction de la statistique

La statistique de test W suit la loi tabulée pour le test des rangs signés de Wilcoxon.

1. Calculer les différences d_i entre chaque valeur et la médiane postulée $\tilde{\mu}$
2. Calculer la valeur absolue des différences précédentes : $|d_i|$
3. Déterminer le rang pour chacune des valeurs $|d_i|$, en ne tenant pas compte des valeurs nulles.
Si des observations ont la même valeur $|d_i|$, alors affecter le rang moyen de ces observations.
4. Calculer la statistique W qui est la somme des rangs dont les d_i sont positifs.

Rejet de H_0 si $W > W_\alpha =$ valeur critique.

Rappels

Test de la médiane

Tests non
paramétriques

Principe

Exemple

Statistique

Grand échantillon

Échantillon de grande taille

Rappels

Test de la médiane

Tests non paramétriques

Principe

Exemple

Statistique

Grand échantillon

En pratique lorsque $n > 20$,
Statistique de test

$$z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

où

n = taille de l'échantillon

W = somme des rangs positifs R_+ suit une loi normale centrée réduite.