

Tests paramétriques et tests non paramétriques

A. Tests Paramétriques

Lorsque les données sont issues d'une distribution paramétrée, et dans ce cas les caractéristiques des données peuvent être résumées à l'aide de paramètres estimés sur l'échantillon.

La moyenne et la variance, suffisent pour caractériser complètement la distribution.

B. Tests non Paramétriques

Ne font aucune hypothèse sur la distribution sous-jacente des données.

- Lorsque les données sont quantitatives, les tests non paramétriques transforment les valeurs en rangs. "Test de rangs".

- Et lorsque les données sont qualitatives, seuls les tests non paramétriques sont utilisables.

- Comme les tests non paramétriques ne font aucune hypothèse sur les distributions des données ces tests élargissent le champ d'application des procédures statistiques, et ils sont même puissants lorsque ces hypothèses sont compatibles avec les données.

La puissance de tests non paramétrique est généralement plus faible que leur équivalent en paramétrique. La philosophie de ces tests est généralement d'appliquer une transformation sur les données avant d'effectuer le calcul. Dans le cas, le principe est de remplacer les observations par leur rang, après les avoir classées par ordre croissant.

Exemple ① Soient les observations suivantes
10 - 20 - 13 - 45 - 12 - 50

On les remplace par les rangs:

On a:

10	12	13	20	45	50
1	2	3	4	5	6

Et les observations:

9	10	9	7	8	8	7	9	10	14
1	2	3	4	5	6	7	8	9	10
7	7	8	8	9	9	9	10	10	14
1.5	1.5	3.5	3.5	6	6	6	8.5	8.5	10

Rang de 7: $\frac{1+2}{2} = 1.5$; Rang de 8: $\frac{3+4}{2} = 3.5$

Rang de 9: $\frac{5+6+7}{3} = 6$; Rang de 14: 10

Les tests non paramétriques

① Comparaison des moyennes de deux échantillons indépendants: Test de Mann-Whitney

} H_0 : Les deux distributions sont semblables.

} H_1 : Les deux distributions ne sont pas semblables.

② Classer/dot des observations des deux échantillons.

③ On donne des rangs à ces observations.

④ On calcule les sommes des rangs de chacun des deux échantillons (W_x et W_y).

$$⑤ W_x = n_1 \times n_2 + \frac{n_1(n_1+1)}{2} - W_x$$

$$W_y = n_1 \times n_2 + \frac{n_2(n_2+1)}{2} - W_y$$

$$⑥ U_{\text{cal}} = \min(W_x, W_y)$$

⑦ Conclusion: On note par n_0 la taille la plus petite des deux échantillons.

1^{er} cas: Si ($n_0 \leq 20$), On compare U_{cal} à la valeur U_5 lue dans la table de Mann-Whitney en fonction de n_1 et n_2 et ne que $\alpha = 5\%$ ou 1% .

Si $U_{\text{cal}} \leq U_5$, alors H_0 est rejetée.

2^{ème} cas si $n_0 > 20$, $Z_{\text{cal}} = \frac{|U_{\text{cal}} - (n_1 \times n_2)/2|}{\sqrt{n_1 \times n_2 \left(\frac{n_1 + n_2 + 1}{12} \right)}}$

On compare Z_{col} à la valeur (1,96) ou (2,58)

Si $Z_{\text{col}} \geq 1,96$ (ou 2,58), alors H_0 est rejetée.

② Comparaison entre deux échantillons dépendants (liés): Test-W de Wilcoxon

① On calcule la différence: $(x_i, y_i) \rightarrow d_i = y_i - x_i$

② Classement de d_i par valeurs absolues croissantes

③ On élimine les ($d_i = 0$), et on assigne les d_i

④ On donne un rang au $|d_i|$

$\left. \begin{array}{l} H_0: \text{Les deux distributions sont semblables} \\ H_1: \text{Les deux distributions ne sont pas} \\ \text{semblables} \end{array} \right\}$

⑤ $W_{\text{col}} = \min(N, p)$ où p est la somme des rangs positifs et N est la somme des rangs négatifs.

2.2 $W_{\text{col}} = \frac{n(n+1)}{2}$ / n est le nombre de $d_i \neq 0$

⑥ Conclusion:

1^{er} cas Si ($n \leq 20$), on compare W_{col} à la valeur W_s donnée par la table de Wilcoxon (n, α), et on a

Si $W_{\text{col}} < W_s$ alors H_0 est rejetée.

2^{ème} Cas : Si $(n > 20)$; $Z_{\text{wil}} = \frac{|W_{\text{wil}} - \frac{n(n+1)}{4}|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$

Si $Z_{\text{wil}} > 1.96$ (2.58), alors H_0 est rejetée.

Exemple

x_i	y_i	d_i	$ d_i $	Rang
128	110	18	18	7
112	119	24	24	8
114	116	-2	2	1
212	205	7	7	4
107	96	11	11	5
116	104	12	12	6
131	128	-3	3	2
148	148	0	0	1
131	126	5	5	3

$$N = 9 : (x_i, y_i)$$

$$n = 8 ; d_i \neq 0$$

① Test de Wilcoxon

H_0 : Les deux distributions sont semblables

$$② N = 1 + 2 = 3$$

$$P = 3 + 4 + 5 + 6 + 7 + 8 = 33$$

$$③ W_{\text{wil}} = \min(N/P) = \min(3, 33) = 3$$

$$\text{vérification : } 3 + 33 = \frac{8 \times 9}{2}$$

Conclusion : $W_s = W_{\text{wil}} = 4$
(0.05; 8)

$W_{\text{wil}} < W_s \Rightarrow H_0$ est rejetée au risque $\alpha = 5\%$.
donc, il y a une différence significative au
risque $\alpha = 5\%$.

③ Comparaison entre K ($K \geq 3$) moyennes
des échantillons indépendants

— Test de Kruskal-Wallis

C'est l'anova à 1 facteur non paramétrique

① On remplace l'ensemble de toutes les observations
dans toutes les échantillons par leur rang

② On fait la somme r_i des rangs de chaque
colonne

③ On calcule la statistique H_{col}

$$H_{col} = \left(\frac{12}{N(N+1)} \sum_{i=1}^K \frac{r_i^2}{n_i} \right) - 3(N+1)$$

④ Conclusion:

1^{er} cas Si $n_i \geq 5$, On compare H_{col} à la
valeur $\chi^2_{(K-1)}$; Si $H_{col} \geq \chi^2_{(K-1)}$ alors H_0 est rejetée.

2^{ème} cas Si l'un ou moins des effectifs
 n_i est petit, on compare H_{col} à la valeur
 H_{col} de la table de Kruskal-Wallis.
Si $H_{col} \leq H_{col}$ alors H_0 est rejetée.

Exemple

Traitement

	Témoin	+2 % Glucose	+2 % Fructose	+1 % glucose +1 % fruc	+2 % saccharose
Poids	58	57	58	75	62
de	59	58	61	67	66
	58	60	56	70	65
	61	59	58	75	63
Penicillium	57	62	57	65	64
	56	60	56	71	62
	58	60	61	67	65
	57	57	60	67	65
	57	59	57	76	62
	59	61	58	68	67
Effectifs (N _i)	10	10	10	10	10
Moyennes	58	59,3	59,2	70,1	64,1
Variances	2	2,41	3,16	14,29	2,79

H_0 : Les 5 distributions sont semblables.

H_1 : " " " " " ne " pas "

On remplace le tableau ci-dessus par un autre tableau des rangs, puis on calcule H_{calc} .