

Chapitre 2

Comparaisons de deux distributions

Il s'agit de comparer les distributions d'un même caractère dans deux populations, observées sur deux échantillons. Les techniques statistiques utilisées dépendent du type de caractère étudié, qualitatif ou quantitatif, des tailles des échantillons et de s'ils sont indépendants ou non (appariés).

Pour un caractère qualitatif (à deux modalités ou plus) et des tailles d'échantillons suffisamment grandes (> 30) on utilise des tests du khi-deux (ou khi-carré χ^2) qui consistent à comparer les proportions des différentes modalités.

Pour un caractère quantitatif, lorsque les distributions sont supposées normales, il suffit pour les comparer, de comparer leurs moyennes (indice de position ou de valeur centrale) et donc de procéder à un test de comparaison de deux moyennes basé sur la loi de Student, ou lorsque les tailles des échantillons sont suffisamment grandes (> 30) d'utiliser des tests basés sur les approximations normales des moyennes empiriques.

En revanche lorsque les distributions ne peuvent pas être considérées comme normales, et en général pour de petites tailles d'échantillons (< 30), il est préférable d'utiliser des tests dits non-paramétriques (*distribution free*) qui ne font pas d'hypothèse sur la forme des distributions et consistent à comparer l'ensemble des distributions (les fonctions de répartition) ou les médianes (indice de position ou de valeur centrale) de ces distributions.

La plupart de ces techniques se généralisent à la comparaison de plus de deux distributions.

1. Variables qualitatives

Tests de comparaison de proportions

- Deux échantillons indépendants : Test du khi-deux d'homogénéité (cf Annexe, section **B**)
généralisation à plus de 2 distributions : même test
- Deux échantillons appariés : Test du khi-deux de Mac-Nemar
généralisation à plus de 2 distributions : Test Q de Cochran

2. Variables quantitatives

- * Lois normales : tests paramétriques, test de comparaison de deux moyennes (cf Annexe, section **A**)
 - Deux échantillons indépendants : Test de Student (cf Annexe, section **A.1**)
généralisation à plus de 2 distributions (moyennes) : Test d'ANOVA à un facteur
 - Deux échantillons appariés : Test de Student (cf Annexe, section **A.3**)
généralisation à plus de 2 distributions (moyennes) : Test d'ANOVA à un facteur avec mesures répétées

- * Grands échantillons : tests paramétriques, test de comparaison de deux moyennes (cf Annexe, section **A**)

- Deux échantillons indépendants : Test normal (cf Annexe, section **A.2**)
généralisation à plus de 2 distributions (moyennes) : Test d'ANOVA à un facteur
- Deux échantillons appariés : Test normal (cf Annexe, section **A.3**)
généralisation à plus de 2 distributions (moyennes) : Test d'ANOVA à un facteur avec mesures répétées

Conditions d'application des tests paramétriques (cf Annexe, section **A.4**) :

- adéquation à la loi normale : Test de Kolmogorov-Smirnov, Test de Lilliefors, Test de Shapiro-Wilk, droite de Henry
- égalité de deux variances : Test de Fisher (rapport des variances)
- homogénéité des variances : Test de Bartlett (généralisation du test de Fisher), Test de Levene, Test de Brown-Forsythe, Test de Hartley, Test de Cochran

- * **Petits échantillons : tests non-paramétriques**

- Deux échantillons indépendants : Test de Wilcoxon Mann-Whitney (cf chapitre 2, section 2)
généralisation à plus de 2 distributions : Test d'ANOVA à un facteur de Kruskal et Wallis
- Deux échantillons appariés : Test des signes (cf chapitre 2, section 3.3), Test de Wilcoxon ou test des signes et rangs (cf chapitre 2, section 3.4)
généralisation à plus de 2 distributions : Test d'ANOVA de Friedman, coefficient de concordance de Kendall

Tests non paramétriques

1 Tests non paramétriques basés sur les rangs

Les tests non paramétriques de (Wilcoxon) Mann-Whitney et de Wilcoxon (ou "signes et rangs") sont basés sur les rangs des observations, par ordre croissant (de la plus petite à la plus grande valeur).

Ces tests nécessitent seulement de savoir ordonner les individus les uns par rapport aux autres (on n'a pas besoin des valeurs précises de la variable étudiée).

Rang : place occupée par une valeur dans la suite ordonnée en ordre croissant (de la plus petite à la plus grande valeur).

- exemple pour $n = 6$ observations de la variable X :

x_i	30	12	41	27	20	32	
$\text{rang}(x_i)$	4	1	6	3	2	5	les rangs vont de 1 à 6.

Propriétés des rangs

Dans un échantillon de n valeurs, les rangs vont de 1 à n .

La somme S de tous les rangs ne dépend que de n : $S = \frac{n \times (n+1)}{2}$

- calcul de S pour $n = 6$:

$$\begin{array}{rcl} S & = & 1 + 2 + 3 + 4 + 5 + 6 \\ S & = & 6 + 5 + 4 + 3 + 2 + 1 \\ \hline 2S & = & 7 + 7 + 7 + 7 + 7 + 7 = 6 \times 7 \end{array} \quad \text{donc } S = \frac{6 \times 7}{2} = 21.$$

- même raisonnement pour n :

$$\begin{array}{rcl} S & = & 1 + 2 + \dots + n \\ S & = & n + (n-1) + \dots + 1 \\ \hline 2S & = & n+1 + (n+1) + \dots + (n+1) \\ & = & n(n+1) \end{array} \quad \text{donc } S = \frac{n \times (n+1)}{2}$$

Traitement des ex aequo

En cas d'ex aequo, on attribue le rang moyen c'est à dire la moyenne des rangs qu'ils auraient eu s'ils avaient été consécutifs.

- exemple : ex aequo sur les rangs 1 et 2 :

$$\text{rang moyen } \frac{1+2}{2} = 1,5 \quad \begin{array}{c|cccc} x_i & 14 & 22 & 14 & 37 \\ \hline \text{rang}(x_i) & 1,5 & 3 & 1,5 & 4 \end{array} \quad S = \frac{4 \times 5}{2} = 10$$

- exemple : ex aequo

$$\begin{array}{l} \text{sur les rangs 2 et 3 : rang moyen } \frac{2+3}{2} = 2,5 \\ \text{et sur les rangs 4, 5 et 6 : rang moyen } \frac{4+5+6}{3} = 5 \end{array}$$

$$\begin{array}{c|cccccc} x_i & 12 & 21 & 37 & 21 & 37 & 37 \\ \hline \text{rang}(x_i) & 1 & 2,5 & 5 & 2,5 & 5 & 5 \end{array} \quad S = \frac{6 \times 7}{2} = 21$$

2 Comparaison de deux distributions sur deux échantillons indépendants

Test de Wilcoxon-Mann-Whitney

2.1 Contexte

On étudie deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$ et deux variables qui représentent le même caractère, quantitatif de loi continue. Elles sont notées : X dans $\mathcal{P}_1$ et Y dans $\mathcal{P}_2$.

On veut comparer les distributions de X et de Y .

On dispose de deux **échantillons indépendants** ; cas le plus habituel : ils ont été obtenus par tirage au sort dans deux populations différentes.

Exemple type d'utilisation de ce test : comparaison de l'efficacité de deux traitements

$\mathcal{P}_1 = \{\text{personnes sous traitement } A\}$ X désigne le résultat avec le traitement A , et
 $\mathcal{P}_2 = \{\text{personnes sous traitement } B\}$ Y représente le résultat avec le traitement B .

Exemple 1

Pour étudier l'efficacité d'un traitement contre la claustrophobie, 13 personnes atteintes de claustrophobie ont été réparties au hasard dans 2 groupes de 6 et 7 personnes.

Les personnes du premier groupe ont reçu un placebo et celles du second groupe le traitement. Après 15 jours de traitement, on a évalué le degré de claustrophobie des 13 personnes

placebo	5,2	5,3	5,6	6,3	7,7	8,1	
traitement	4,6	4,9	5,1	5,5	6,1	6,5	7,2

Peut-on au risque $\alpha = 5\%$, accepter l'hypothèse que le traitement est efficace ?

Contexte :

$\mathcal{P}_1 = \{\text{personnes claustrophobes sous placebo}\}$

$\mathcal{P}_2 = \{\text{personnes claustrophobes sous traitement}\}$

X = degré de claustrophobie sous placebo, dans $\mathcal{P}_1$

Y = degré de claustrophobie sous traitement, dans $\mathcal{P}_2$

Les 2 variables mesurent le même caractère et sont quantitatives continues.

2.2 Hypothèses de test et risque

Sous l'hypothèse nulle H_0 , les variables X et Y sont distribués de la même manière (les deux traitements ont la même efficacité)

$$H_0 : X \text{ et } Y \text{ ont la même loi} \quad \text{ou} \quad H_0 : X \equiv Y$$

pour l'hypothèse alternative H_1 , selon l'hypothèse de recherche envisagée,

soit l'un des deux traitements est plus efficace que l'autre, **alternative unilatérale**

les valeurs de X peuvent être globalement supérieures à celles de Y

c-à-d que la distribution de X est à droite de celle de Y

$$H_1 : X \succ Y$$

ou bien les valeurs de X peuvent être globalement inférieures à celles de Y

c-à-d que la distribution de X est à gauche de celle de Y

$$H_1 : X \prec Y$$

soit les traitements ont des efficacités différentes, **alternative bilatérale**

les valeurs de X sont globalement différentes de celles de Y

X et Y n'ont pas la même loi (sans orientation)

$$H_1 : X \neq Y$$

Exemple 1

Hypothèses et risque $\alpha = 5\%$

Sous H_0 on suppose que traitement et placebo ont la même efficacité, les degrés de claustrophobie sous placebo X et sous traitement Y sont globalement identiques.

Le traitement est efficace si les degrés de claustrophobie sous traitement Y sont inférieurs aux degrés sous placebo $X : Y \prec X$ ou $X \succ Y$.

D'où les hypothèses à tester $\begin{cases} H_0 : \text{les valeurs de } X \text{ sont globalement égales à celles de } Y \\ H_1 : \text{les valeurs de } X \text{ sont globalement supérieures à celles de } Y \end{cases}$

ou $\begin{cases} H_0 : X \text{ et } Y \text{ ont la même distribution} \\ H_1 : \text{ la distribution de } X \text{ est à droite de celle de } Y \end{cases}$

Le test s'écrit de manière équivalente $\begin{cases} H_0 : X \equiv Y \\ H_1 : X \succ Y \end{cases}$ test unilatéral, au risque $\alpha = 5\%$

2.3 Observations

On dispose de 2 échantillons tirés au hasard de manière indépendante dans les 2 populations. On note :

$\mathcal{E}_1$ l'échantillon de taille n_1 issu de $\mathcal{P}_1$ et x_i les mesures de $\mathcal{E}_1$,

$\mathcal{E}_2$ l'échantillon de taille n_2 issu de $\mathcal{P}_2$ et y_i les mesures de $\mathcal{E}_2$

n représente la taille totale des 2 échantillons : $n = n_1 + n_2$

Exemple 1

On dispose de 2 échantillons indépendants :

$\mathcal{E}_1$ de taille $n_1 = 6$ issu de $\mathcal{P}_1$ et $\mathcal{E}_2$ de taille $n_2 = 7$ issu de $\mathcal{P}_2$.

Au total, nous avons $n = 6 + 7 = 13$ individus.

$\mathcal{E}_1$ est appelé "groupe témoin" et $\mathcal{E}_2$ "groupe expérimental".

2.4 Analyse descriptive des données

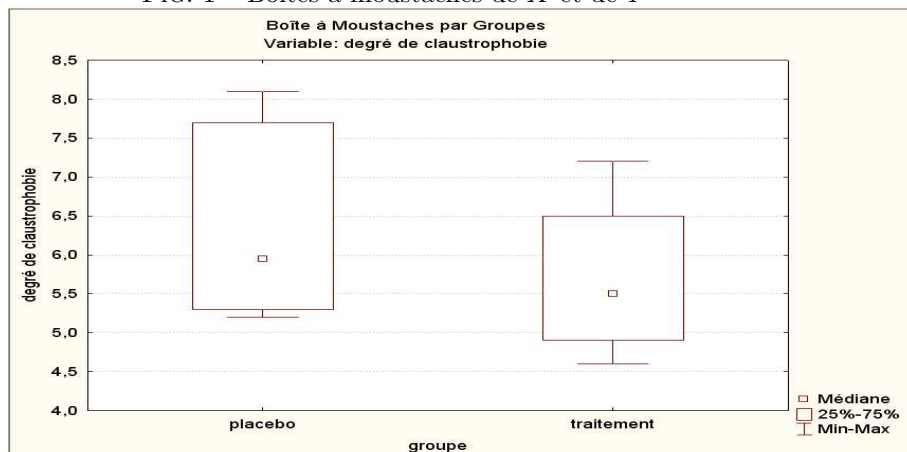
Histogrammes et boîtes à moustaches comparatifs des données observées de X et de Y permettent de situer les deux distributions l'une par rapport à l'autre et de comparer visuellement indices de position (médiane) et de dispersion (intervalle inter-quartiles).

Exemple 1

Les médianes observées de X et de Y valent resp. $m_X = 5,95$ (milieu entre 5,6 et 6,3) et $m_Y = 5,5$ (valeur observée de Y de rang 4) : elles sont proches (pour l'ensemble des 2 groupes la médiane observée vaut $m = 5,6$ valeur observée de (X, Y) de rang 7).

Cependant la dispersion observée de X est plus grande que celle de Y et la distribution observée de X est décalée à droite par rapport à celle de Y (valeurs de X globalement plus grandes que celles de Y) : les degrés de claustrophobie observés sont globalement plus élevés sous placebo que sous traitement (Figure 1). Il faut néanmoins faire un test pour confirmer ou infirmer la présence de ce décalage dans les populations.

FIG. 1 – Boîtes à moustaches de X et de Y



2.5 Statistiques de test

On définit tout d'abord les statistiques de Wilcoxon pour 2 échantillons indépendants notées W_x et W_y puis les statistiques de Mann-Whitney notées U_x et U_y qui en découlent, d'utilisation plus simple.

• Principe

Sous H_0 : X et Y ont la même loi, ou $H_0 : X \equiv Y$

les deux échantillons ne forment qu'un seul échantillon tiré d'une seule population.

Si on range par ordre croissant l'ensemble des n valeurs (les 2 échantillons confondus) les rangs de X et de Y sont équivalents.

Exemple 1

Ici, $n = 13$: les rangs dans l'interclassement de X et de Y , notés $\text{rang}(x, y)$ vont de 1 à 13.

x_i	5,2	5,3	5,6	6,3	7,7	8,1	
$\text{rang}(x, y)$	4	5	7	9	12	13	$w_x = 50$
y_i	4,6	4,9	5,1	5,5	6,1	6,5	7,2
$\text{rang}(x, y)$	1	2	3	6	8	10	11
							$w_y = 41$

Remarques :

- En cas d'ex aequo, on leur attribue leur rang moyen.
- On ne supprime jamais d'observations dans ce test.

• Statistiques de Wilcoxon W_x et W_y pour 2 échantillons indépendants

W_x = somme des rangs de X et W_y = somme des rangs de Y

W_x et W_y sont des variables quantitatives discrètes.

Propriété : $W_x + W_y = \frac{n(n+1)}{2}$

en effet, cette somme correspond à celle de tous les rangs des individus, du 1^{er} au $n^{\text{ème}}$, elle vaut donc $1 + 2 + \dots + n = \frac{n(n+1)}{2}$.

Exemple 1

Statistiques de Wilcoxon W_x et W_y

Les valeurs observées sont

pour $W_x : w_x = 4 + 5 + 7 + 9 + 12 + 13 = 50$

pour $W_y : w_y = 1 + 2 + 3 + 6 + 8 + 10 = 41$

vérification : $n = 13$, donc $\frac{n \times (n+1)}{2} = \frac{13 \times 14}{2} = 91$.

Nous avons $w_x + w_y = 50 + 41 = 91$ donc $w_x + w_y = \frac{n(n+1)}{2}$.

Remarque :

les valeurs observées de W_x et W_y sont généralement des valeurs entières, sauf en présence d'ex aequo.

• Domaines de variation de W_x et W_y

- Pour la statistique W_x

Au minimum : les n_1 valeurs de X ont les rangs les plus faibles, donc de 1 à n_1 (les x_i prennent les n_1 premières places) alors : $w_x = 1 + 2 + \dots + n_1 = \frac{n_1(n_1+1)}{2}$.

Au maximum : à l'inverse, les n_1 valeurs de X ont les rangs les plus élevés, et par conséquent, les y_i occupent les rangs les plus faibles donc de 1 à n_2 (les y_i prennent les n_2 premières places et les x_i les n_1 dernières places) alors : $w_y = \frac{n_2(n_2+1)}{2}$ et au maximum $w_x = \frac{n(n+1)}{2} - \frac{n_2(n_2+1)}{2}$.

Domaine de variation de W_x : $\left\{ \frac{n_1(n_1+1)}{2}, \dots, \frac{n(n+1)}{2} - \frac{n_2(n_2+1)}{2} \right\}$

Exemple 1

Pour $n_1 = 6$ et $n_2 = 7$ le minimum de W_x vaut $\frac{n_1(n_1+1)}{2} = \frac{6 \times 7}{2} = 21$ et le maximum de W_x est $\frac{13 \times 14}{2} - \frac{7 \times 8}{2} = 91 - 28 = 63$.

W_x varie de 21 à 63 : son domaine de variation (définition) est $\{21, 22, \dots, 63\}$.

- Pour la statistique W_y

Un raisonnement identique pour les mesures y_i nous conduit au domaine de variation de W_y :

$$\left\{ \frac{n_2(n_2+1)}{2}, \dots, \frac{n(n+1)}{2} - \frac{n_1(n_1+1)}{2} \right\}$$

Exemple 1

Pour $n_1 = 6$ et $n_2 = 7$ le minimum de W_y vaut $\frac{n_2(n_2+1)}{2} = \frac{7 \times 8}{2} = 28$ et le maximum $\frac{n(n+1)}{2} - \frac{n_1(n_1+1)}{2} = 91 - 21 = 70$ donc W_y varie de 28 à 70.

Lorsque $n_1 \neq n_2$ les statistiques W_x et W_y ont des domaines de variations différents.

Sous H_0 , elles n'ont donc pas tout à fait la même distribution et ne jouent pas un rôle symétrique, d'où des difficultés d'utilisation.

Pour simplifier nous allons utiliser les statistiques de Mann-Whitney.

• Statistiques de Mann-Whitney U_x et U_y

On se base sur W_x et W_y mais on fait en sorte que les domaines de variation de U_x et U_y commencent à 0. On définit :

$$U_x = W_x - \frac{n_1 \times (n_1 + 1)}{2} \quad \text{et} \quad U_y = W_y - \frac{n_2 \times (n_2 + 1)}{2}$$

U_x et U_y sont des variables quantitatives discrètes.

Exemple 1

Statistiques de Mann-Whitney U_x et U_y

Les valeurs observées sont

$$\text{pour } U_x : u_x = 50 - \frac{6 \times 7}{2} = 50 - 21 = 29$$

$$\text{pour } U_y : u_y = 41 - \frac{7 \times 8}{2} = 41 - 28 = 13$$

Le domaine de variation de U_x va de $21 - 21 = 0$ jusqu'à $63 - 21 = 42$ qui est en fait égal à $n_1 \times n_2 = 6 \times 7 = 42$.

Le domaine de variation de U_y va de $28 - 28 = 0$ jusqu'à $70 - 28 = 42$ qui est aussi égal à $n_1 \times n_2 = 42$.

• Domaine de variation de U_x et U_y

U_x et U_y ont le même domaine de variation : $\{0, 1, \dots, n_1 \times n_2\}$

- U_x et W_x varient dans le même sens, de même U_y et W_y .

- U_x et U_y varient en sens contraire de manière à ce que leur somme soit toujours constante, égale à :

$$U_x + U_y = n_1 \times n_2$$

Exemple 1

Vérification : $u_x + u_y = 29 + 13 = 42$ et $n_1 \times n_2 = 6 \times 7 = 42$, donc $u_x + u_y = n_1 \times n_2$.

• Interprétation de U_x et U_y

U_x est le nombre de fois où une valeur de X est supérieure à une valeur de Y , ou le nombre de fois où un rang de X est supérieur à un rang de Y . De même pour U_y .

2.6 Lois des statistiques de Mann-Whitney sous H_0 et niveau de signification du test

Sous H_0 les statistiques U_x et U_y ont la même distribution, notée U , définie sur $\{0, 1, \dots, n_1 \times n_2\}$ symétrique autour de la moyenne (et milieu) $\frac{n_1 \times n_2}{2}$.

• Loi exacte des statistiques de Mann-Whitney sous H_0

Pour des échantillons de petites tailles ($n_1 \leq 8$ et $n_2 \leq 8$) les probabilités cumulées $P_{H_0} [U \leq u]$ sont données dans les tables de Mann-Whitney.

Puisque $u_x + u_y = n_1 \times n_2$ on déduit que u_x et u_y sont symétriques par rapport au milieu sur la distribution de U et donc que : $P_{H_0} [U_x \geq u_x] \stackrel{\text{sym}}{=} P_{H_0} [U_y \leq u_y] = P_{H_0} [U \leq u_y]$.

On pourra ainsi trouver le niveau de signification α_{obs} ou *p-valeur*, soit en lisant directement les tables, soit en utilisant au préalable la symétrie de la distribution de U .

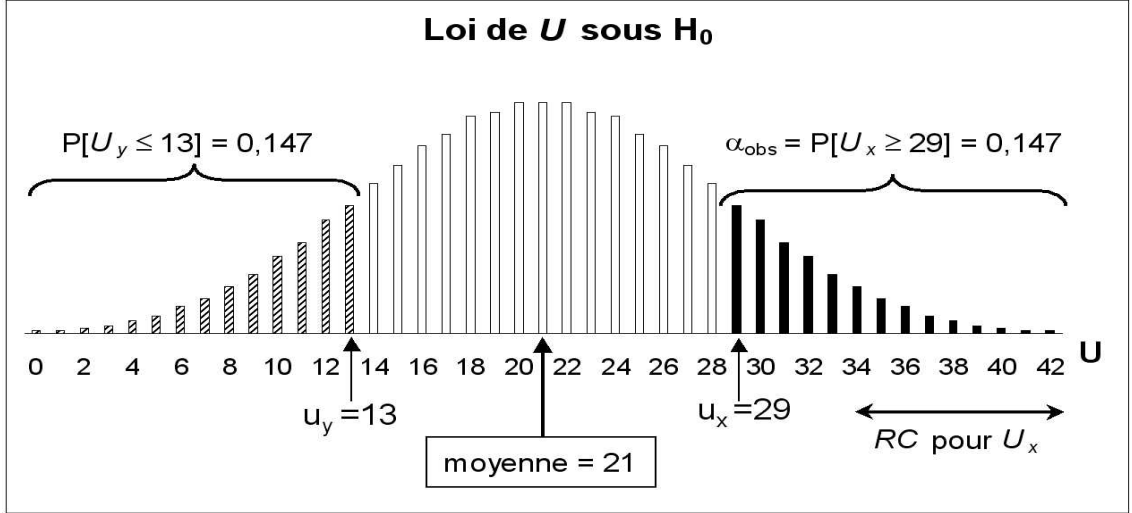
Exemple 1

Loi exacte des statistiques de Mann-Whitney sous H_0

Sous $H_0 : X \equiv Y$

- U_x et U_y ont la même distribution, définie sur $\{0, 1, \dots, 6 \times 7 = 42\}$ symétrique autour de la moyenne (et milieu) $\frac{n_1 \times n_2}{2} = \frac{42}{2} = 21$, représentée par son diagramme en bâtons (Figure 2).
 - les x_i et les y_i sont mélangés de manière équilibrée donc les rangs faibles (forts) apparaissent en quantités similaires dans les 2 échantillons.
- Alors les valeurs de U_x et U_y sont proches, situées près du centre de la distribution.

FIG. 2 – Diagramme en bâtons de la loi de U sous H_0 pour $n_1 = 6$ et $n_2 = 7$



Région critique et niveau de signification du test

Sous $H_1 : X \succ Y$ les valeurs de X occupent des rangs élevés ce qui implique de grandes valeurs de W_x et donc de U_x , et à l'opposé de petites valeurs de W_y et donc de U_y . La région critique du test RC est à droite de la distribution de U_x (à gauche pour U_y).

Si on choisit de travailler avec U_x :

$$\alpha_{obs} = P_{H_0}[U_x \geq u_x] = P_{H_0}[U_x \geq 29] \stackrel{\text{sym}}{=} P_{H_0}[U \leq u_y] = P_{H_0}[U \leq 13] = 0,147$$

(cf table de Mann-Whitney)

- Région critique et niveau de signification du test

Test unilatéral

$$\begin{array}{lll} H_1 : X \succ Y & RC \text{ à droite pour } U_x \text{ (à gauche pour } U_y) & \alpha_{obs} = P_{H_0}[U_x \geq u_x] \stackrel{\text{sym}}{=} P_{H_0}[U \leq u_y] \\ H_1 : X \prec Y & RC \text{ à gauche pour } U_x & \alpha_{obs} = P_{H_0}[U_x \leq u_x] \end{array}$$

Test bilatéral

$$\begin{array}{lll} H_1 : X \neq Y & RC \text{ aux 2 extrémités de } U_x \text{ (ou de } U_y) & \alpha_{obs} = 2 \times P_{H_0}[U \leq u_{min}] \\ & & \text{où } u_{min} = \min(u_x, u_y) \end{array}$$

- Approximation normale des statistiques de Mann-Whitney sous H_0

Pour des tailles d'échantillons suffisamment grandes ($n_1 > 8$ ou $n_2 > 8$) la distribution de U sous H_0 est approximée par une loi normale de moyenne $\mu(U) = \frac{n_1 \times n_2}{2}$ et de variance $var(U) = \frac{n_1 \times n_2 \times (n+1)}{12}$ où $n = n_1 + n_2$.

$$\text{La variable centrée et réduite : } Z = \frac{U - \mu(U)}{\sqrt{var(U)}} = \frac{U - \frac{n_1 \times n_2}{2}}{\sqrt{\frac{n_1 \times n_2 \times (n+1)}{12}}} \approx \mathcal{N}(0, 1)$$

$$\text{Si on choisit de travailler avec } U_x \text{ la valeur observée de } Z : z_{obs} = \frac{u_x - \mu(U)}{\sqrt{var(U)}} \left(= -\frac{u_y - \mu(U)}{\sqrt{var(U)}} \right)$$

- Région critique et niveau de signification du test

Si on choisit de travailler avec U_x (dans le cas contraire, pour U_y le signe de Z et les inégalités sont inversés)

Test unilatéral

$$\begin{aligned}
H_1 : X \succ Y & \quad RC \text{ à droite pour } U_x \text{ et pour } Z & \alpha_{obs} = P_{H_0} [U_x \geq u_x] \simeq P_{H_0} [Z \geq z_{obs}] \\
H_1 : X \prec Y & \quad RC \text{ à gauche pour } U_x \text{ et pour } Z & \alpha_{obs} = P_{H_0} [U_x \leq u_x] \simeq P_{H_0} [Z \leq z_{obs}]
\end{aligned}$$

Test bilatéral

$$H_1 : X \neq Y \quad RC \text{ aux 2 extrémités de } U_x \text{ et de } Z \quad \alpha_{obs} \simeq 2 \times P_{H_0} [Z \geq |z_{obs}|]$$

Exemple 1

Approximation normale sous H_0

Si on approxime la loi de U par la loi normale de moyenne $\mu(U) = \frac{n_1 n_2}{2} = \frac{6 \times 4}{2} = 12$ et de variance

$$var(U) = \frac{n_1 n_2 (n_1 + 1)}{12} = \frac{6 \times 4 \times 14}{12} = 28$$

$$\text{la statistique de test } Z = \frac{U - \mu(U)}{\sqrt{var(U)}} \stackrel{\text{approx}}{\sim} \mathcal{N}(0, 1)$$

(mais cette approximation est peu justifiée ici puisque n_1 et $n_2 \leq 8$).

On choisit de travailler avec U_x donc la valeur observée de Z :

$$z_{obs} = \frac{u_x - \mu(U)}{\sqrt{var(U)}} = \frac{29 - 12}{\sqrt{28}} = \frac{17}{\sqrt{28}} \simeq 3,22857$$

Région critique et niveau de signification du test

$$\alpha_{obs} \simeq P[Z \geq 3,22857] = 1 - P[Z \leq 3,22857] = 1 - 0,9973451 = 0,0026549$$

(cf Table de la loi $\mathcal{N}(0, 1)$).

2.7 Décision et conclusion

- Règle de décision basée sur α_{obs}

- si $\alpha_{obs} > \alpha$ on conserve H_0 (on ne valide pas H_1) au seuil α et au risque de 2^{de} espèce β inconnu
- si $\alpha_{obs} \leq \alpha$ on rejette H_0 en faveur de H_1 (on valide H_1) au risque α et au niveau de signification (*p-valeur*) α_{obs}

- Conclusion

Exemple 1

Décision et conclusion

loi exacte : $\alpha_{obs} = 14,7\% > \alpha = 5\%$ donc on conserve H_0 et on ne valide pas H_1 au risque β

approximation normale : $\alpha_{obs} \simeq 12,7\% > \alpha = 5\%$ donc on conserve H_0 au risque β .

NB : dans ce cas l'approximation normale étant peu justifiée puisque n_1 et $n_2 \leq 8$, on préférera donc conclure à partir de la décision découlant de la loi exacte.

On ne peut pas conclure à l'efficacité du traitement contre la claustrophobie, au seuil $\alpha = 5\%$ et au risque β inconnu.

Remarque : le test de Student de comparaison de deux moyennes pour deux échantillons indépendants sous les conditions de normalité de X et de Y (tests de normalité non significatifs) et d'égalité des variances (tests d'égalité non significatifs) donnerait $t_{obs} = 1,09499$ et $\alpha_{obs} = P_{H_0} [T \geq t_{obs}] = P_{H_0} [T \geq 1,09499] = 0,158459 \simeq 15,8\%$ d'après la loi de Student T_{11} et induirait la même conclusion. Cependant, étant donné les petites tailles d'échantillons, les tests vérifiant la validité des conditions d'application du test de Student sont peu puissants pour détecter un écart aux conditions, leurs résultats sont peu fiables ; il est donc préférable de se référer au test non paramétrique présenté ici.

2.8 Remarques

Les variables X et Y étant continues, la probabilité de trouver deux valeurs identiques (ex aequo) est nulle. En pratique, ils sont produits par les arrondis ou les erreurs de mesure :

- les formules (donc les tables) changent en présence d'ex aequo. Il est important de les utiliser uniquement s'il n'y a que "peu" d'ex aequo.
- des corrections pour prendre en compte les ex aequo sont disponibles et programmées dans les logiciels.

Résultats obtenus avec STATISTICA

STATISTICA calcule les valeurs observées des statistiques de Wilcoxon W_x et W_y et de la statistique $U_{min} = \min(U_x, U_y)$.

STATISTICA ne calcule que les p -valeurs bilatérales; pour obtenir le niveau de signification d'un test unilatéral il faut diviser la valeur donnée par 2.

STATISTICA calcule systématiquement les p -valeurs bilatérales exacte et basée sur l'approximation normale avec correction de continuité sur la variable U_x , c'est à dire :

$$z_{obs}^C = \begin{cases} \frac{u_x - \mu(U) - 0,5}{\sqrt{\text{var}(U)}} & \text{si } u_x - \mu(U) > 0 \\ \frac{u_x - \mu(U) + 0,5}{\sqrt{\text{var}(U)}} & \text{si } u_x - \mu(U) < 0 \end{cases}$$

sans tenir compte des tailles des échantillons. Pour de petites tailles (n_1 ou $n_2 \leq 8$) il est préférable d'interpréter la p -valeur exacte, à moins qu'il y ait beaucoup d'ex aequo.

Cette correction de continuité consiste à diminuer la valeur obtenue avec l'approximation normale afin d'augmenter le niveau de signification pour rendre le test plus conservateur.

En présence d'ex aequo, on utilise la valeur de l'approximation normale corrigée pour en tenir compte (Z ajusté) calculée par STATISTICA, et sa p -valeur associée.

Exemple 2

Les niveaux de compétences manuelles des élèves de deux écoles maternelles A et B ont été mesurées pour 6 élèves de l'école A et 8 de l'école B, tirés au sort parmi les élèves de grande section de chaque école

école A	20	12	0	2	6	92	
école B	31	14	15	10	21	19	28 22

Peut-on, au risque $\alpha = 10\%$, accepter l'hypothèse que les niveaux de compétences manuelles sont différents dans les deux écoles ?

Contexte

$\mathcal{P}_1 = \{\text{élèves de l'école A}\}$ $X = \text{niveau de compétences manuelles dans } \mathcal{P}_1$

$\mathcal{P}_2 = \{\text{élèves de l'école B}\}$ $Y = \text{niveau de compétences manuelles dans } \mathcal{P}_2$

Les 2 variables mesurent le même caractère et sont quantitatives continues.

Hypothèses de test et risque

$\begin{cases} H_0 : X \text{ et } Y \text{ ont la même distribution} \\ H_1 : \text{les } X \text{ sont globalement différents des } Y \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : X \equiv Y \\ H_1 : X \not\equiv Y \end{cases} \quad \text{test bilatéral, } \alpha = 10\%$

Observations

On dispose de 2 échantillons indépendants :

$\mathcal{E}_1$ de taille $n_1 = 6$ issu de $\mathcal{P}_1$ et $\mathcal{E}_2$ de taille $n_2 = 8$ issu de $\mathcal{P}_2$. Au total, $n = 6 + 8 = 14$.

Statistiques de test

Statistiques de Wilcoxon W_x et W_y pour 2 échantillons indépendants.

On range par ordre croissant l'ensemble des $n = 14$ valeurs : les rangs vont de 1 à 14.

x_i	20	12	0	2	6	92	
$\text{rang}(x, y)$	9	5	1	2	3	14	$w_x = 34$
y_i	31	14	15	10	21	19	28 22
$\text{rang}(x, y)$	13	6	7	4	10	8	12 11 $w_y = 71$

$W_x = \text{somme des rangs de } X$ et $W_y = \text{somme des rangs de } Y$.

Les valeurs observées sont :

$$w_x = 9 + 5 + 1 + 2 + 3 + 14 = 34$$

$$w_y = 13 + 6 + 7 + 4 + 10 + 8 + 12 + 11 = 71$$

vérification : la somme $w_x + w_y = 34 + 71 = 105$.

Or $n = 14$, donc $\frac{n(n+1)}{2} = \frac{14 \times 15}{2} = 105$ d'où $w_x + w_y = \frac{n(n+1)}{2}$.

Statistiques de Mann-Whitney U_x et U_y

Les valeurs observées sont :

$$u_x = w_x - \frac{n_1 \times (n_1 + 1)}{2} = 34 - \frac{6 \times 7}{2} = 34 - 21 = 13$$

$$u_y = w_y - \frac{n_2 \times (n_2 + 1)}{2} = 71 - \frac{8 \times 9}{2} = 71 - 36 = 35$$

U_x et de U_y variables quantitatives discrètes définies sur : $\{0, 1, \dots, n_1 \times n_2 = 6 \times 8 = 48\}$
vérification : $u_x + u_y = 13 + 35 = 48$ est égale à $n_1 \times n_2 = 6 \times 8 = 48$, donc $u_x + u_y = n_1 \times n_2$.

Lois sous H_0 , région critique et niveau de signification du test

– loi exacte

Sous H_0 , U_x et U_y ont la même loi définie sur $\{0, 1, \dots, 48\}$ et symétrique de moyenne $\frac{n_1 n_2}{2} = 24$.

Région critique et niveau de signification

Sous $H_1 : X \neq Y$ RC aux deux extrémités du domaine de U_x (ou de U_y).

$\alpha_{obs} = 2 \times P_{H_0} [U \leq u_{min}] = 2 \times P_{H_0} [U \leq 13] = 2 \times 0,091 = 0,182 = 18,2\%$ (table de Mann-Whitney)

– approximation normale

Sous H_0 , U_x et U_y suivent la même loi, approximativement normale

de moyenne $\mu(U) = \frac{n_1 n_2}{2} = 24$ et de variance $var(U) = \frac{6 \times 8 \times 15}{12} = 60$

et la variable centrée réduite $Z = \frac{U-24}{\sqrt{60}}$ suit approximativement une loi normale $\mathcal{N}(0, 1)$

(approximation peu justifiée ici puisque n_1 et $n_2 \leq 8$).

Valeur observée de Z

On choisit de travailler avec U_x : $z_{obs} = \frac{u_x - \mu(U)}{\sqrt{var(U)}} = \frac{13-24}{\sqrt{60}} = -1,420094$

Région critique et niveau de signification

Sous $H_1 : X \neq Y$ RC aux deux extrémités du domaine de U_x (ou de U_y) et donc de Z .

$\alpha_{obs} = 2 \times P_{H_0} [Z \geq |z_{obs}|] = 2 \times P [Z \geq |-1,420094|] = 2 \times (1 - P [Z \leq 1,420094])$
 $= 2 \times (1 - 0,922210) = 2 \times 0,077790 = 0,15558 \simeq 15,6\%$ (cf table de la loi $\mathcal{N}(0, 1)$)

Décision et conclusion

loi exacte : $\alpha_{obs} = 18,2\% > \alpha = 10\%$ donc on conserve H_0 , on ne valide pas H_1 au seuil $\alpha = 10\%$ et au risque de 2^{de} espèce β .

approximation normale : $\alpha_{obs} = 15,6\% < \alpha = 10\%$ même décision que précédemment.

NB : dans ce cas l'approximation normale étant peu justifiée puisque n_1 et $n_2 \leq 8$, on préférera donc conclure à partir de la décision découlant de la loi exacte.

Les niveaux de compétences manuelles ne sont pas différents dans les deux écoles, au seuil $\alpha = 10\%$ et au risque de 2^{de} espèce β inconnu.

Remarque : le test de Student de comparaison de deux moyennes pour deux échantillons indépendants n'est pas applicable sur ces données puisque la condition de normalité de X n'est pas vérifiée (tests de normalité Lilliefors ($p < 0,05$) et Shapiro-Wilk ($p = 0,004$) significatifs à 5%) et celle d'égalité des variances ne l'est pas non plus (test du rapport des variances significatif à 5% ($p = 0,000539$)).

Exemple 3

On a noté les poids de naissance de bébés (en kg) suivant si la mère avait reçu des soins prénataux depuis le 1er trimestre de grossesse (expérimental) ou le 3ème (en routine) pour resp. 10 et 8 bébés

3 ^{ème} trimestre	1,68	3,83	3,11	2,76	1,70	2,79	1,40	2,66			
1 ^{er} trimestre	3,05	2,77	2,94	3,38	4,90	2,81	2,80	3,21	3,08	2,95	

Peut-on, au risque $\alpha = 5\%$, accepter l'hypothèse que les soins prénataux précoces sont bénéfiques sur le poids de naissance du bébé ?

Contexte

$\mathcal{P}_1 = \{\text{bébés dont les mères ont des soins prénataux au 3^{ème} trimestre de grossesse}\}$

$\mathcal{P}_2 = \{\text{bébés dont les mères ont des soins prénataux au 1^{er} trimestre de grossesse}\}$

$X = \text{poids de naissance du bébé dans } \mathcal{P}_1$ $Y = \text{poids de naissance du bébé dans } \mathcal{P}_2$

Les 2 variables mesurent le même caractère et sont quantitatives continues.

Hypothèses de test et risque

$\begin{cases} H_0 : X \text{ et } Y \text{ ont la même distribution} \\ H_1 : \text{les } X \text{ sont globalement inférieurs aux } Y \end{cases}$ ou $\begin{cases} H_0 : X \equiv Y \\ H_1 : X < Y \end{cases}$ test unilatéral, $\alpha = 5\%$

Observations

On dispose de 2 échantillons indépendants :

$\mathcal{E}_1$ "groupe témoin" de taille $n_1 = 8$ issu de $\mathcal{P}_1$ et $\mathcal{E}_2$ "groupe expérimental" de taille $n_2 = 10$ issu de $\mathcal{P}_2$ d'où $n = n_1 + n_2 = 18$.

Statistiques de test

Statistiques de Wilcoxon puis de Mann-Whitney pour 2 échantillons indépendants.

On calcule les rangs des individus dans l'ensemble des $n = 18$ individus : les rangs vont de 1 à 18.

3 ^{ème} trimestre	1,68	3,83	3,11	2,76	1,70	2,79	1,40	2,66		
$\text{rang}(x, y)$	2	17	14	5	3	7	1	4	$w_x = 53$	
1 ^{er} trimestre	3,05	2,77	2,94	3,38	4,90	2,81	2,80	3,21	3,08	2,95
$\text{rang}(x, y)$	12	6	10	16	18	9	8	15	13	11
									$w_y = 118$	

On définit les statistiques de Wilcoxon :

$$W_x = \text{somme des rangs de } X \quad \text{et} \quad W_y = \text{somme des rangs de } Y$$

dont les valeurs observées sont

$$w_x = 2 + 17 + 14 + 5 + 3 + 7 + 1 + 4 = 53$$

$$w_y = 12 + 6 + 10 + 16 + 18 + 9 + 8 + 15 + 13 + 11 = 118$$

$$\text{vérification : } w_x + w_y = \frac{n(n+1)}{2} \quad \text{puisque } w_x + w_y = 53 + 118 = 171 \quad \text{et} \quad \frac{n(n+1)}{2} = \frac{18 \times 19}{2} = 171.$$

On définit les statistiques de Mann-Whitney : U_x et U_y

dont les valeurs observées sont

$$u_x = w_x - \frac{n_1(n_1+1)}{2} = 53 - \frac{8 \times 9}{2} = 53 - 36 = 17$$

$$u_y = w_y - \frac{n_2(n_2+1)}{2} = 118 - \frac{10 \times 11}{2} = 118 - 55 = 63$$

$$\text{vérification : } u_x + u_y = n_1 \times n_2 \quad \text{puisque } u_x + u_y = 17 + 63 = 80 \quad \text{et} \quad n_1 n_2 = 8 \times 10 = 80.$$

U_x et U_y quantitatives discrètes sur $\{0, \dots, 80\}$.

Lois sous H_0 région critique et niveau de signification du test

– loi exacte

Sous H_0 , U_x et U_y ont la même loi définie sur $\{0, 1, \dots, 80\}$ et symétrique de moyenne $\frac{n_1 n_2}{2} = 40$.

Région critique et niveau de signification

Sous $H_1 : X \prec Y$ les valeurs de X occupent des rangs faibles d'où de petites valeurs de W_x et de U_x (à l'opposé, de grandes valeurs de W_y et de U_y) ; RC se situe à gauche du domaine de U_x (à droite pour U_y).

On choisit de travailler avec U_x .

$$\alpha_{obs} = P_{H_0} [U_x \leq u_x] = P_{H_0} [U \leq 17] = 0,0217 = 2,17\% \quad (\text{table de Mann-Whitney})$$

– approximation normale

Sous H_0 , U_x et U_y suivent la même loi, approximativement normale

de moyenne $\mu(U) = \frac{n_1 n_2}{2} = 40$ et de variance $\text{var}(U) = \frac{8 \times 10 \times 19}{12} = 95$

et la variable centrée réduite $Z = \frac{U-40}{\sqrt{95}}$ suit approximativement une loi normale $\mathcal{N}(0, 1)$.

Valeur observée de Z

$$\text{On choisit de travailler avec } U_x : \quad z_{obs} = \frac{u_x - \mu(U)}{\sqrt{\text{var}(U)}} = \frac{17-40}{\sqrt{95}} = -2,04$$

Région critique et niveau de signification

Sous $H_1 : X \prec Y$ RC se situe à gauche du domaine de U_x et donc de Z .

$$\alpha_{obs} = P[Z \leq z_{obs}] = P[Z \leq -2,04] = 1 - P[Z \leq 2,04] = 1 - 0,9793 = 0,0207 \quad (\text{Table de la loi } \mathcal{N}(0, 1))$$

Décision et conclusion

loi exacte : $\alpha_{obs} = 2,17\% < \alpha = 5\%$ donc on rejette H_0 en faveur de H_1 (on valide H_1) au risque $\alpha = 5\%$.

approximation normale : $\alpha = 5\% > \alpha_{obs} = 2,07\%$ donc on rejette H_0 en faveur de H_1 (on valide H_1) au risque $\alpha = 5\%$ (même conclusion).

Les soins prénataux donnés au 1er trimestre de grossesse augmentent le poids de naissance du bébé, au risque $\alpha = 5\%$ et au niveau de signification $\alpha_{obs} = 2,17\%$.

NB : dans ce cas l'approximation normale peut être justifiée puisque $n_2 = 10 > 8$, mais quand on dispose de la p -valeur calculée à partir de la loi exacte on préférera conclure à partir de la décision en découlant. On peut remarquer qu'ici les deux p -valeurs sont très proches et conduisent à la même décision.

En présence d'ex aequo, l'approximation normale avec correction pour les ex aequo peut donner des résultats plus fiables que la loi exacte qui n'en tient pas compte.

Remarque : le test de Student de comparaison de deux moyennes pour deux échantillons indépendants n'est pas applicable sur ces données puisque la condition de normalité de Y n'est pas vérifiée (tests de normalité Lilliefors ($p < 0,05$) et Shapiro-Wilk ($p = 0,0002$) significatifs à 5%).

3 Comparaison de deux distributions pour deux échantillons appariés

3.1 Contexte

Il s'agit d'un test portant sur deux échantillons appariés (ou pairés, ou appareillés) de deux variables X et Y représentant le même caractère quantitatif continu,

soit issus d'une même population $\mathcal{P}$, et dans ce cas les mesures de X et de Y sont faites sur les mêmes individus,

soit issus d'une population $\mathcal{P}$ composée de paires d'individus (2 individus) aussi semblables que possible :

- personnes de la même famille (paires de jumeaux, (père, fils), ...)
- ou bien, on définit des "variables d'appariement" (sexe, âge, durée ou gravité de la maladie, ...).

Afin de comparer les distributions de X et Y , on cherche à contrôler les facteurs "connus" et "inconnus" qui jouent un rôle dans la différence entre les variables X et Y , autres que le facteur étudié, qui pourraient être des "facteurs de confusion".

Exemples types d'utilisation de ce test :

* comparaison de type avant-après (un traitement, une thérapie, un régime, ...)

Pour chaque sujet, X représente le résultat avant traitement et Y le résultat après traitement.

Pour détecter un changement dû à ce traitement :

en hypothèse nulle, on suppose qu'il n'y a pas de changement, les valeurs de X sont globalement égales à celles de Y $H_0 : X \equiv Y$

en alternative,

il y a une amélioration, **alternative unilatérale**

les valeurs de X sont globalement plus élevées que celles de Y $H_1 : X \succ Y$
ou plus faibles $H_1 : X \prec Y$

ou une modification (sans préciser l'orientation) **alternative bilatérale**

les valeurs de X sont globalement différentes de celles de Y $H_1 : X \neq Y$

La "paire" est constituée par un seul individu et l'appariement est idéal : le sujet est pris comme son "propre témoin" (ou "contrôle"). Ceci permet d'augmenter la puissance en diminuant la variabilité des résultats : en effet, la variabilité des résultats d'un même individu ("intra-sujet") est plus faible que celle d'individus différents ("inter-sujets").

* comparaison de l'efficacité de deux traitements

X désigne le résultat avec le traitement A et Y le résultat avec le traitement B .

Pour comparer l'efficacité des traitements :

en hypothèse nulle, on suppose que les deux traitements ont la même efficacité $H_0 : X \equiv Y$

en alternative, selon l'hypothèse de recherche envisagée

un traitement est plus performant que l'autre, **alternative unilatérale**

$H_1 : X \succ Y$ ou $H_1 : X \prec Y$

ou ils ont simplement des efficacités différentes (sans orientation) **alternative bilatérale**

$H_1 : X \neq Y$

Si le sujet est son propre témoin, chaque sujet reçoit le traitement A et le traitement B dans un ordre qui peut-être aléatoire, ou prédéterminé si l'on veut étudier un "effet ordre" du traitement.

L'appariement permet de contrôler les facteurs connus (ou inconnus) pour influencer l'évolution de la maladie, autres que le traitement (âge, intensité, durée, ...).

Il permet de limiter le "biais de confusion" qui consisterait à conclure à l'existence d'une différence d'efficacité entre les traitements alors qu'elle n'est due qu'à un (ou des) facteur(s) de confusion.

On dispose donc de deux échantillons appariés (en paires d'individus "jumeaux") de taille n' .

Principe pour des échantillons appariés :

on travaille avec la variable "différence" notée $D = X - Y$ (ou $Y - X$).

Le test de comparaison de 2 distributions sur 2 échantillons appariés est le test de comparaison de leur différence à 0.

On peut utiliser :

- soit le **test du signe (ou des signes) sur la différence** D en comparant la médiane de la différence D à la valeur théorique $M_0 = 0$,
- soit le **test des signes et rangs de Wilcoxon sur la différence** D (D supposée symétrique).

Comme leur nom l'indique, ces deux tests utilisent *les signes* de la différence D .

3.2 Observations

On dispose de 2 échantillons appariés de même taille n' issu de $\mathcal{P}$.

On note :

- $\mathcal{E}_1$ l'échantillon de taille n' et $(x_1, \dots, x_{n'})$ les mesures de X ,
- $\mathcal{E}_2$ l'échantillon de taille n' et $(y_1, \dots, y_{n'})$ les mesures de Y .

- On calcule la variable différence pour chaque paire $d_i = x_i - y_i$ (ou $y_i - x_i$) :
suppression des individus pour lesquels la valeur de la différence D est nulle
on dispose d'un échantillon $\mathcal{E}$ de taille n de D et $(d_1, \dots, d_n)$ représentent les valeurs observées non nulles de D .
- On construit la variable auxiliaire "signe" notée $signe(D)$ qualitative à 2 modalités : '+' ou '-'

$$\begin{cases} signe(d_i) = + & \text{si } d_i = x_i - y_i > 0 \text{ ou si } x_i > y_i \\ signe(d_i) = - & \text{si } d_i = x_i - y_i < 0 \text{ ou si } x_i < y_i \end{cases}$$
de paramètre $p = \begin{cases} \text{proportion de '+'} \\ \text{proportion de valeurs de } D > 0 \\ \text{proportion de valeurs de } X > Y \end{cases}$ inconnue dans $\mathcal{P}$.

Exemple 4

On cherche à savoir si dans les familles, les aînés ont tendance à être plus indépendants que leurs cadets. Lors d'une étude on a procédé à l'évaluation sur une échelle d'indépendance en 25 points, de 9 aînés et du frère ou de la sœur qui suit directement chacun des aînés. On a obtenu les résultats suivants

score de l'aîné	8	12	14	15	5	13	18	20	17
score du puîné	9	10	14	12	9	8	12	10	10

Peut-on répondre à la question, au risque $\alpha = 0,05$?

Contexte

$\mathcal{P} = \{\text{paire}=(\text{aîné, puîné}) \text{ d'une même famille}\}$

$X = \text{score d'indépendance de l'aîné}$ $Y = \text{score d'indépendance du puîné}$

Les 2 variables mesurent le même caractère (score d'indépendance) et sont quantitatives continues.

Hypothèses et risque $\alpha = 5\%$

On souhaite tester si les aînés sont plus indépendants, donc si les scores X des aînés sont globalement supérieurs aux scores Y des puînés, noté $X \succ Y$.

ce qui s'écrit $\begin{cases} H_0 : X \equiv Y \\ H_1 : X \succ Y \end{cases}$ alternative unilatérale, au risque $\alpha = 5\%$.

On définit la variable auxiliaire "différence" : $D = X - Y$.

C'est une variable quantitative continue de médiane M inconnue.

Observations

On dispose de deux échantillons appariés de taille $n' = 9$. On calcule les différences $d_i = x_i - y_i$. Une différence est nulle, donc on enlève cette valeur de l'échantillon et on continue à travailler avec $n = 8$. On dispose d'un échantillon $\mathcal{E}$ de la variable différence D de taille $n = 8$.

On construit ensuite la variable auxiliaire "signe" : $signe(D)$ qualitative dichotomique '+' ou '-' pour les $n = 8$ valeurs observées de D non nulles.

score de l'aîné	x_i	8	12	14	15	5	13	18	20	17
score du puîné	y_i	9	10	14	12	9	8	12	10	10
différence	$d_i = x_i - y_i$	-1	2	0	3	-4	5	6	10	7
	$signe(d_i)$	-	+	X	+	-	+	+	+	+

Analyse descriptive des données

La médiane observée de la différence des scores $m = 3$ (valeur observée de rang 5 car $n' = 9$) : est-elle significativement supérieure à 0 ?

Par ailleurs la boîte à moustache des valeurs observées de D indique une distribution symétrique : $\min(D) = -4$, $\max(D) = 10$, $1^{er} \text{ quartile} = 0$ et $3^{ème} \text{ quartile} = 6$.

3.3 Test du signe pour deux échantillons appariés

Ce test de comparaison d'une médiane M inconnue à une valeur théorique M_0 s'applique à une variable quantitative continue. Nous utilisons ici un cas particulier de ce test pour une valeur théorique nulle $M_0 = 0$ et pour une variable représentant la différence D entre deux autres variables, notées précédemment X et Y .

Le nom du test vient du fait qu'il utilise des signes '+' et '-' au lieu de mesures quantitatives.

Il peut être appliqué à des problèmes où les mesures quantitatives ne sont pas possibles (ou disponibles), mais où l'on sait seulement s'il y a augmentation ou diminution.

3.3.1 Hypothèses de test et risque

On considère la variable différence $D = X - Y$ quantitative continue de médiane M , inconnue dans la population $\mathcal{P}$.

Les hypothèses se traduisent sur la médiane de la variable D .

L'hypothèse nulle	$H_0 : X \equiv Y$	se traduit par	$H_0 : M = 0$
l'alternative bilatérale	$H_1 : X \not\equiv Y$	s'écrit	$H_1 : M \neq 0$
ou les alternatives unilatérales	$H_1 : X \succ Y$	devient	$H_1 : M > 0$ à droite,
et	$H_1 : X \prec Y$		$H_1 : M < 0$ à gauche.

On est donc ramené à un test de comparaison d'une médiane (de la différence) à la valeur théorique (de référence) 0, au même risque α .

Remarque : $M \neq M_X - M_Y$ où M_X est la médiane de X et M_Y celle de Y .

Exemple 4

Hypothèses de test et risque

Les hypothèses du test de $H_0 : X \equiv Y$ contre $H_1 : X \succ Y$ se traduisent sur la médiane M de la différence des scores d'indépendance, par

$$\begin{cases} H_0 : M = 0 \\ H_1 : M > 0 \end{cases} \text{ test unilatéral droit, au risque } \alpha = 5\%.$$

En effet, H_1 signifie que les valeurs de X sont globalement supérieures aux valeurs de Y , donc que les différences $D = X - Y$ sont globalement positives et ainsi la médiane de la variable D est positive.

• Principe du test du signe

On utilise la variable auxiliaire "signe" pour traduire les hypothèses précédentes sur la proportion p de signes '+' inconnue dans la population.

Sous $H_0 : M = 0$ la médiane de D étant nulle, il y aura globalement autant '+' que de '-'

donc $H_0 : M = 0$ se traduit par $H_0 : p = \frac{1}{2}$

En revanche sous $H_1 : M > 0$ (la vraie médiane M est > 0) du fait de ce décalage on enregistre plus de valeurs de D positives, donc plus de la moitié de signes '+'

d'où $H_1 : M > 0$ se traduit par $H_1 : p > \frac{1}{2}$

On s'est ramené à un test de comparaison d'une proportion à la valeur théorique $p_0 = \frac{1}{2}$ $\begin{cases} H_0 : p = \frac{1}{2} \\ H_1 : p > \frac{1}{2} \end{cases}$

où $p = \begin{cases} \text{proportion de '+'} \\ \text{proportion de valeurs de } D > 0 \\ \text{proportion de valeurs de } X > Y \end{cases}$ test unilatéral droit, au risque α .

Un raisonnement similaire permet de déduire que

sous $H_1 : M < 0$ (la vraie médiane M est < 0) du fait de ce décalage on enregistre plus de valeurs '-' et moins de valeurs '+'

donc $H_1 : M < 0$ se traduit par $H_1 : p < \frac{1}{2}$ unilatérale gauche,

et que $H_1 : M \neq 0$ se traduit par $H_1 : p \neq \frac{1}{2}$ bilatérale.

Ainsi, le test du signe est le test de comparaison de la proportion p de signes '+', à $\frac{1}{2}$ ($= 0,5 = 50\%$).

Test unilatéral droit

$$\text{test de } \begin{cases} H_0 : X \equiv Y \\ H_1 : X \succ Y \end{cases} \quad \text{ou de } \begin{cases} H_0 : M = 0 \\ H_1 : M > 0 \end{cases} \quad \text{ou de } \begin{cases} H_0 : p = \frac{1}{2} \\ H_1 : p > \frac{1}{2} \end{cases}$$

Test unilatéral gauche

$$\text{test de } \begin{cases} H_0 : X \equiv Y \\ H_1 : X \prec Y \end{cases} \quad \text{ou de } \begin{cases} H_0 : M = 0 \\ H_1 : M < 0 \end{cases} \quad \text{ou de } \begin{cases} H_0 : p = \frac{1}{2} \\ H_1 : p < \frac{1}{2} \end{cases}$$

Test bilatéral

$$\text{test de } \begin{cases} H_0 : X \equiv Y \\ H_1 : X \neq Y \end{cases} \quad \text{ou de } \begin{cases} H_0 : M = 0 \\ H_1 : M \neq 0 \end{cases} \quad \text{ou de } \begin{cases} H_0 : p = \frac{1}{2} \\ H_1 : p \neq \frac{1}{2} \end{cases}$$

Exemple 4

On considère la variable $\text{signe}(D)$ qualitative à 2 modalités '+' et '-'
de paramètre p = proportion de signes '+'.

Les hypothèses se traduisent pour le test du signe par :

$$\begin{cases} H_0 : p = \frac{1}{2} & \text{les proportions de signes '+' et '-' sont égales} \\ H_1 : p > \frac{1}{2} & \text{la proportion de signes '+' est plus élevée que celle des signes '-'} \end{cases}$$

test unilatéral droit, au risque $\alpha = 5\%$.

3.3.2 Statistique de test et loi sous H_0

On considère comme statistique de test

$$\begin{aligned} S_n &= \text{effectif empirique de signes '+'} \\ &= \text{nombre de signes '+' pour l'ensemble des échantillons de taille } n \end{aligned}$$

variable quantitative discrète, définie sur le domaine $\{0, 1, 2, \dots, n\}$.

La valeur observée de S_n :

$$\left. \begin{aligned} s_{obs} &= \text{nombre de signes '+'} \\ &= \text{nombre de valeurs } d_i > 0 \\ &= \text{nombre de valeurs } x_i > y_i \end{aligned} \right\} \text{observés sur l'échantillon.}$$

Exemple 4

Statistique de test

S_n = nombre de signes '+' pour tous les échantillons de taille $n = 8$

variable quantitative discrète définie sur $\{0, 1, \dots, n = 8\}$.

Sur l'échantillon $\mathcal{E}$ de taille $n = 8$ on observe $s_{obs} = 6$ signes '+' pour les différences d_i .

• **Loi exacte de la statistique de test sous H_0 : loi binomiale**

Sous $H_0 : p = \frac{1}{2}$ il y a autant de signes '+' que de signes '-',
la statistique de test S_n suit une loi binomiale de paramètres n et $\frac{1}{2}$ notée $\mathcal{B}(n, \frac{1}{2})$ symétrique, de moyenne (et milieu) $\frac{n}{2}$.

Pour de petites valeurs de n ($n \leq 30$), les tables de lois binomiales donnent les probabilités cumulées $P_{H_0} [S_n \leq s_{obs}]$ qui permettront de calculer le niveau de signification α_{obs} du test.

La symétrie de S_n sous H_0 permet de déduire que : $P_{H_0} [S_n \geq s_{obs}] \stackrel{\text{sym}}{=} P_{H_0} [S_n \leq n - s_{obs}]$.

Exemple 4

Loi exacte de la statistique de test sous H_0

Sous $H_0 : X \equiv Y$ ou $H_0 : p = \frac{1}{2}$ la statistique de test S_n suit une loi binomiale $\mathcal{B}(8, \frac{1}{2})$ définie sur $\{0, 1, \dots, n = 8\}$ symétrique autour de sa moyenne (et milieu) $\frac{n}{2} = \frac{8}{2} = 4$ représentée par son diagramme en bâtons (Figure 3).

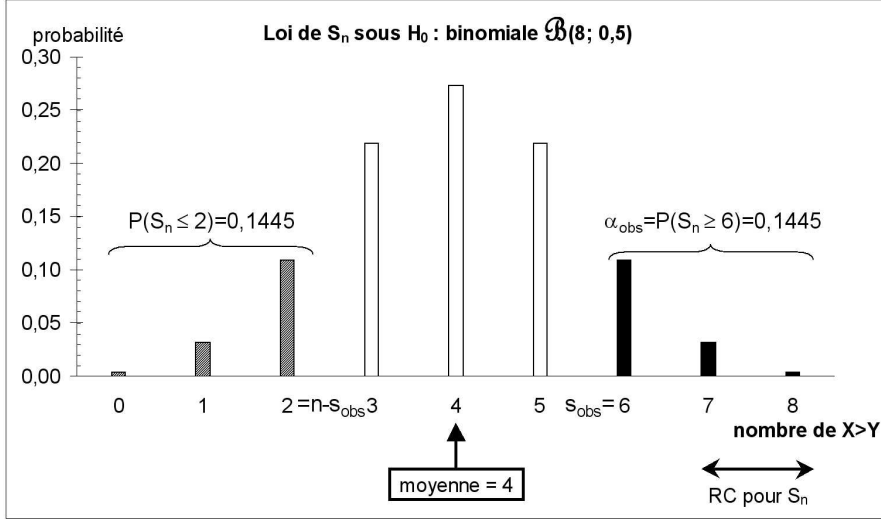
Sous H_0 , les x_i et les y_i sont mélangés de manière équilibrée donc les différences '+' et '-' apparaissent en quantités similaires dans l'échantillon.

La valeur de S_n est alors située près du centre de la distribution.

Région critique et niveau de signification du test

Sous $H_1 : X \succ Y$ les valeurs de X sont plus élevées que celles de Y ce qui implique plus de valeurs positives de $D = X - Y$ donc plus de la moitié de signes '+' (ici, plus de 4), c'est à dire une valeur

FIG. 3 – Diagramme en bâtons de la loi de S_n sous H_0 pour $n = 8 : \mathcal{B}(8, \frac{1}{2})$



de S_n à l'extrémité droite de sa distribution sous H_0 . La région critique du test RC est à droite de la distribution de S_n .

Le niveau de signification du test unilatéral droit

$$\begin{aligned}\alpha_{obs} &= P_{H_0} [S_n \geq s_{obs}] = P_{H_0} [S_n \geq 6] \stackrel{\text{sym}}{=} P_{H_0} [S_n \leq n - s_{obs}] = P_{H_0} [S_n \leq 8 - 6] \\ &= P_{H_0} [S_n \leq 2] = 0,1445 \quad (\text{cf table de la loi binomiale } \mathcal{B}(8, \frac{1}{2})).\end{aligned}$$

- Région critique et niveau de signification du test

Test unilatéral

$$H_1 : X \succ Y \quad H_1 : M > 0 \quad H_1 : p > \frac{1}{2} \quad RC \text{ à droite de } S_n \quad \alpha_{obs} = P_{H_0} [S_n \geq s_{obs}] \stackrel{\text{sym}}{=} P_{H_0} [S_n \leq n - s_{obs}]$$

$$H_1 : X \prec Y \quad H_1 : M < 0 \quad H_1 : p < \frac{1}{2} \quad RC \text{ à gauche de } S_n \quad \alpha_{obs} = P_{H_0} [S_n \leq s_{obs}]$$

Test bilatéral

$$H_1 : X \neq Y \quad H_1 : M \neq 0 \quad H_1 : p \neq \frac{1}{2} \quad RC \text{ aux 2 extrémités de } S_n \quad \alpha_{obs} = 2 \times P_{H_0} [S_n \leq s_{min}]$$

où $s_{min} = \min(s_{obs}, n - s_{obs})$

- Approximation normale de la loi binomiale sous H_0

Pour une taille d'échantillon suffisamment grande ($n \geq 30$) la distribution de S_n sous H_0 suit approximativement une loi normale de moyenne $\mu(S_n) = \frac{n}{2}$ et de variance $var(S_n) = \frac{n}{4}$.

On en déduit que la variable centrée et réduite : $Z = \frac{S_n - \mu(S_n)}{\sqrt{var(S_n)}} = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \stackrel{\text{approx}}{\sim} \mathcal{N}(0, 1)$.

La valeur observée de Z : $z_{obs} = \frac{s_{obs} - \mu(S_n)}{\sqrt{var(S_n)}} = \frac{s_{obs} - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$

- Région critique et niveau de signification du test

Test unilatéral

$$H_1 : X \succ Y \quad RC \text{ à droite pour } S_n \text{ et pour } Z \quad \alpha_{obs} = P_{H_0} [S_n \geq s_{obs}] \simeq P_{H_0} [Z \geq z_{obs}]$$

$$H_1 : X \prec Y \quad RC \text{ à gauche pour } S_n \text{ et pour } Z \quad \alpha_{obs} = P_{H_0} [S_n \leq s_{obs}] \simeq P_{H_0} [Z \leq z_{obs}]$$

Test bilatéral

$$H_1 : X \neq Y \quad RC \text{ aux 2 extrémités de } S_n \text{ et de } Z \quad \alpha_{obs} \simeq 2 \times P_{H_0} [Z \geq |z_{obs}|]$$

Exemple 4

Approximation normale sous H_0

Si on approxime la loi de S_n par la loi normale de moyenne $\mu(S_n) = \frac{n}{2} = 4$ et de variance $var(S_n) = \frac{n}{4} = \frac{8}{4} = 2$; la statistique de test $Z = \frac{S_n - \mu(S_n)}{\sqrt{var(S_n)}} = \frac{S_n - 4}{\sqrt{2}} \stackrel{\text{approx}}{\sim} \mathcal{N}(0, 1)$ (approximation peu justifiée car $n = 8 < 30$).

La valeur observée de Z : $z_{obs} = \frac{s_{obs} - \mu(S_n)}{\sqrt{\text{var}(S_n)}} = \frac{6-4}{\sqrt{2}} \simeq 1,4142$

Région critique et niveau de signification du test

$\alpha_{obs} \simeq P[Z \geq 1,4142] = 1 - P[Z \leq 1,4142] = 1 - 0,92135 = 0,07865$ (cf Table de la loi $\mathcal{N}(0,1)$)

3.3.3 Décision et conclusion

- Règle de décision basée sur α_{obs}
- Conclusion

Exemple 4

Décision et conclusion

loi exacte : $\alpha_{obs} = 14,45\% > \alpha = 5\%$ donc on conserve H_0 au seuil $\alpha = 5\%$ avec un risque β .

approximation normale : $\alpha_{obs} = 7,87\% > \alpha = 5\%$, décision identique à la précédente.

NB : dans ce cas l'approximation normale étant peu justifiée puisque $n < 30$, on préférera donc conclure à partir de la décision découlant de la loi exacte.

On ne peut pas conclure au seuil 5% que les aînés ont tendance à être plus indépendants que leurs cadets, au risque de seconde espèce β inconnu.

Résultats obtenus avec STATISTICA

STATISTICA calcule :

- la valeur observée $\frac{n-s_{obs}}{n}$ exprimé en % c'est à dire le pourcentage de signes négatifs ou de valeurs $x_i < y_i$ parmi les n valeurs $x_i \neq y_i$
- la valeur observée de l'approximation normale de S_n avec correction de continuité

$$z_{obs}^c = \frac{|s_{obs} - \mu(S_n)| - 0,5}{\sqrt{\text{var}(S_n)}} \quad \text{et la } p\text{-valeur bilatérale correspondante.}$$

C'est une correction de continuité classique permettant d'éviter de trop sous estimer la vraie p -valeur en utilisant l'approximation normale au lieu de test binomial exact : elle diminue la valeur de la statistique de test afin de rendre le test plus conservateur (diminution de la p -valeur par rapport à celle obtenue avec z_{obs}).

STATISTICA ne calcule que la p -valeur bilatérale ; pour obtenir le niveau de signification d'un test unilatéral il faut diviser la valeur obtenue par 2.

STATISTICA ne propose pas le test binomial exact mais fait systématiquement l'approximation normale avec correction de continuité quelle que soit la taille de l'échantillon.

Exemple 5

Huit adolescentes de 13 à 18 ans présentant une carence en fer sans anémie ont reçu un traitement pour améliorer leur fonction cognitive. On mesure cette fonction par le score au test d'apprentissage verbal de Hopkin avant et après traitement. On obtient les valeurs du tableau suivant

score avant traitement	5	3,1	2,8	1,5	1,8	0,5	3,7	6
score avant après traitement	3,1	6	6,7	1,5	3,5	5,5	4,3	5,7

Peut-on conclure, au risque $\alpha = 0,10$ que le traitement est efficace ?

Contexte

$\mathcal{P} = \{\text{adolescentes ayant une carence en fer sans anémie, sous traitement}\}$

$X = \text{score de Hopkin avant traitement}$ $Y = \text{score de Hopkin après traitement}$

Les 2 variables mesurent le même caractère (fonction cognitive) et sont quantitatives continues.

Hypothèses et risque $\alpha = 10\%$

On souhaite tester si le traitement est efficace, donc si les scores Y après traitement sont globalement supérieurs aux scores X avant traitement, noté $Y \succ X$ ou $X \prec Y$.

On teste $\begin{cases} H_0 : \text{le traitement n'est pas efficace} \\ H_1 : \text{le traitement est efficace} \end{cases}$ ce qui s'écrit $\begin{cases} H_0 : X \equiv Y \\ H_1 : X \prec Y \end{cases}$

On définit la variable auxiliaire "différence" : $D = X - Y$.

C'est une variable quantitative continue de médiane M inconnue.

Elle représente la mesure de l'effet du traitement sur le score de Hopkin (la fonction cognitive).

Observations

On dispose de deux échantillons appariés de taille $n' = 8$. On calcule les différences $d_i = x_i - y_i$. Une différence est nulle, donc on enlève cette valeur de l'échantillon et on continue à travailler avec $n = 7$.

On dispose d'un échantillon $\mathcal{E}$ de la variable différence D de taille $n = 7$.

On construit la variable $\text{signe}(D)$ qualitative à 2 modalités '+' et '-' pour les $n = 7$ valeurs non nulles de D .

x_i	5	3,1	2,8	1,5	1,8	0,5	3,7	6
y_i	3,1	6	6,7	1,5	3,5	5,5	4,3	5,7
d_i	1,9	-2,9	-3,9	0	-1,7	-5	-0,6	0,3
$\text{signe}(d_i)$	+	-	-	X	-	-	-	+

Analyse descriptive des données

La médiane observée de D $m = -1,15$ (milieu des valeurs observées de rang 4 et 5 car $n' = 8$, donc entre $-1,7$ et $-0,6$) est légèrement négative : est-elle significativement inférieure à 0 ?

Par ailleurs la boîte à moustache des valeurs observées de D indique une distribution plutôt symétrique : $\min(D) = -5$, $\max(D) = 1,9$, 1^{er} quartile = $-3,4$ et 3^{ème} quartile = $0,15$.

Test du signe

Hypothèses de test et risque

Sur la médiane M de la différence des scores de Hopkin, le test se traduit alors par

$$\begin{cases} H_0 : M = 0 \\ H_1 : M < 0 \end{cases} \text{ test unilatéral gauche, au risque } \alpha = 10\%.$$

En effet, H_1 signifie que les valeurs de X sont globalement inférieures aux valeurs de Y , donc les différences $D = X - Y$ sont globalement négatives et ainsi la médiane de la variable D est négative.

On considère la variable $\text{signe}(D)$ variable qualitative à 2 modalités '+' et '-' de paramètre p = proportion de signes '+'. Les hypothèses se traduisent pour le test du signe par :

$$\begin{cases} H_0 : p = \frac{1}{2} & \text{les proportions de signes '+' et '-' sont égales} \\ H_1 : p < \frac{1}{2} & \text{la proportion de signes '+' est plus faible que celle des signes '-'} \end{cases}$$

test unilatéral gauche, au risque $\alpha = 10\%$.

Statistique de test

S_n = nombre de signes '+' sur les échantillons de taille $n = 7$

variable quantitative discrète définie sur $\{0, 1, \dots, n = 7\}$.

Sur l'échantillon $\mathcal{E}$ de taille $n = 7$ on observe $s_{obs} = 2$ signes '+' pour les différences d_i .

Lois sous H_0 , région critique et niveau de signification du test

- loi exacte

Sous H_0 , S_n quantitative discrète sur $\{0, 1, \dots, 7\}$ suit une loi binomiale $\mathcal{B}(7, \frac{1}{2})$ de moyenne 3,5.

Sous $H_1 : M < 0$ ou $p < \frac{1}{2}$ les valeurs de D sont en majorité négatives, on s'attend à observer moins de la moitié de signes '+' (moins de 3,5), donc S_n est faible ; RC se situe à gauche du domaine de S_n

$$\alpha_{obs} = P_{H_0}[S_n \leq s_{obs}] = P_{H_0}[S_n \leq 2] = 0,2266 \text{ (Table de la loi binomiale } \mathcal{B}(7, \frac{1}{2}) \text{)}.$$

- approximation normale

Sous H_0 , S_n suit approximativement une loi normale de moyenne $\frac{7}{2} = 3,5$ et de variance $\frac{7}{4} = 1,75$ donc la variable centrée réduite $Z = \frac{S_n - 3,5}{\sqrt{1,75}}$ suit approximativement une loi $\mathcal{N}(0, 1)$ (approximation peu valable car $n = 7 < 30$).

$$\text{La valeur observée de } Z : z_{obs} = \frac{s_{obs} - 3,5}{\sqrt{1,75}} = \frac{2 - 3,5}{\sqrt{1,75}} = -1,1339$$

Sous $H_1 : M < 0$ les valeurs de D sont en majorité négatives, donc S_n est faible ; RC est à gauche du domaine de S_n et de Z , d'où

$$\alpha_{obs} \simeq P_{H_0}[Z \leq z_{obs}] = P_{H_0}[Z \leq -1,1339] = 1 - P_{H_0}[Z \leq 1,1339] = 1 - 0,87158 = 0,12842 \text{ (Table de la loi } \mathcal{N}(0, 1) \text{)}.$$

Décision et conclusion

loi exacte : $\alpha = 10\% < \alpha_{obs} = 22,66\%$ donc on conserve H_0 au risque β inconnu.

approximation normale : $\alpha = 10\% < \alpha_{obs} = 12,84\%$ décision identique à la précédente, bien que peu justifiée.

On ne peut pas conclure à l'efficacité du traitement sur l'amélioration de la fonction cognitive des adolescentes présentant une carence en fer, au seuil $\alpha = 10\%$ et au risque β inconnu.

3.4 Test de Wilcoxon (signes et rangs) sur deux échantillons appariés

Ce test de comparaison de deux distributions s'applique à une variable quantitative continue et symétrique. Nous utilisons ici un cas particulier de ce test pour deux échantillons appariés, c'est à dire pour une variable représentant la **différence** D entre deux autres variables, notées précédemment X et Y , qui sera supposée **symétrique**.

Le test du signe ne tient compte que d'une petite partie de l'information contenue dans les données puisqu'il ne tient compte que du signe ('+' ou '-') de la différence. Pour obtenir un test plus puissant (moindre erreur β de seconde espèce) on veut prendre en compte non seulement les signes des différences, mais également les amplitudes de ces différences. Le test des **"signes et rangs"** attribue un poids plus important à une différence de plus grande amplitude.

Ce test nécessite seulement de connaître les signes (sens) des différences et de savoir les ordonner les unes par rapport aux autres (ne nécessite pas d'avoir des valeurs quantitatives précises).

3.4.1 Observations

On dispose d'un échantillon $\mathcal{E}$ de la variable "différence" $D = X - Y$ de taille n .

Comme pour le test du signe, on supprime les individus dont les différences sont nulles, c'est à dire pour lesquels $x_i = y_i$.

On calcule successivement pour chaque individu :

- la différence $d_i = x_i - y_i$
- le signe de la différence $+$ ou $-$ (pas de 0)
- la valeur absolue de la différence $|d_i|$
- le rang de la valeur absolue $\text{rang}(|d_i|)$

3.4.2 Statistiques de test et loi sous H_0

Statistiques de Wilcoxon V^+ et V^- pour 2 échantillons appariés

$V^+ =$ somme des $\text{rang}(|d_i|)$ associés aux différences positives (de signe '+')

$V^- =$ somme des $\text{rangs}(|d_i|)$ associés aux différences négatives (de signe '-')

Remarque : en cas d'ex aequo, on leur attribue leur rang moyen.

Exemple 4

Contexte

Il est identique à celui du test du signe (cf section 3.1) en supposant la symétrie de la variable "différence" D .

Observations

On dispose d'un échantillon $\mathcal{E}$ de $D = X - Y$ de taille $n = 8$ pour lequel on a calculé les signes $\text{signe}(d_i)$. Il faut calculer également les valeurs absolues des différences $|d_i|$ et les rangs de ces valeurs absolues $\text{rang}(|d_i|)$ de 1 à $n = 8$.

score de l'aîné	x_i	8	12	14	15	5	13	18	20	17
score du puîné	y_i	9	10	14	12	9	8	12	10	10
différence	$d_i = x_i - y_i$	-1	2	0	3	-4	5	6	10	7
	$\text{signe}(d_i)$	-	+	X	+	-	+	+	+	+
	$ d_i $	1	2	0	3	4	5	6	10	7
	$\text{rang}(d_i)$	1	2	X	3	4	5	6	8	7

On vérifie que les rangs vont de 1 à 8.

Statistiques de Wilcoxon

Les valeurs observées des statistiques de Wilcoxon

pour V^+ : $v^+ = 2 + 3 + 5 + 6 + 7 + 8 = 31$

pour V^- : $v^- = 1 + 4 = 5$

Domaine de variation de V^+ et de V^-

Exemple 4

Les rangs vont de 1 à 8.

Au minimum : aucune différence positive donnerait $v^+ = 0$

Au maximum : toutes les différences sont positives, alors $v^+ = 1+2+3+4+5+6+7+8 = S = \frac{8 \times 9}{2} = 36$

Donc V^+ varie dans l'ensemble $\{0, 1, 2, \dots, 36\}$.

Dans le cas général, pour un échantillon de taille n , les rangs vont de 1 à n .

Domaine de variation de $V^+ : \left\{ 0, 1, 2, \dots, \frac{n(n+1)}{2} \right\}$

V^+ variable quantitative discrète.

De même V^- est une variable quantitative discrète sur $\left\{ 0, 1, 2, \dots, \frac{n(n+1)}{2} \right\}$

Propriété : $V^+ + V^- = \frac{n(n+1)}{2}$

en effet, cette somme est celle de tous les rangs de 1 à n .

V^+ et V^- jouent un rôle symétrique :

quand V^+ augmente, V^- diminue et vice-versa, leur somme restant constante, égale à $\frac{n(n+1)}{2}$.

Exemple 4

Ici : $v^+ = 31$, $v^- = 5$ donc $v^+ + v^- = 36$. Or $n = 6$ donc $\frac{n(n+1)}{2} = \frac{6 \times 7}{2} = 21$.

V^+ et V^- sont quantitatives discrètes et varient sur le domaine $\{0, 1, \dots, 36\}$.

3.4.3 Lois des statistiques de Wilcoxon sous H_0 et niveau de signification du test

Sous H_0 les statistiques V^+ et V^- suivent la même loi, notée V , définie sur $\left\{ 0, 1, 2, \dots, \frac{n(n+1)}{2} \right\}$ symétrique autour de la moyenne (et milieu) $\mu(V) = \frac{n(n+1)}{4}$.

• **Loi exacte des statistiques de Wilcoxon sous H_0**

Pour des échantillons de petites tailles ($n \leq 20$) les probabilités cumulées $P_{H_0} [V \leq v]$ sont données dans les tables de Wilcoxon.

Elles permettent de trouver le niveau de signification α_{obs} ou p -valeur. En effet,

$v^+ + v^- = \frac{n(n+1)}{2}$ donc v^+ et v^- sont symétriques par rapport au milieu sur la distribution de V ; on en déduit que : $P_{H_0} [V^+ \geq v^+] \stackrel{\text{sym}}{=} P_{H_0} [V^- \leq v^-] = P_{H_0} [V \leq v^-]$.

Exemple 4

Loi exacte des statistiques de Wilcoxon sous H_0

Sous $H_0 : X \equiv Y$

- V^+ et V^- ont la même distribution, définie sur $\{0, 1, \dots, \frac{8 \times 9}{2} = 36\}$ symétrique autour de la moyenne (et milieu) $\frac{n(n+1)}{4} = \frac{36}{2} = 18$ représentée par son diagramme en bâtons (Figure 4).
- les différences d_i sont équilibrées donc les rangs faibles (forts) apparaissent associés indifféremment à des signes '+' ou '-'.

Alors les valeurs de V^+ et V^- sont proches, situées près du centre de la distribution.

Région critique et niveau de signification du test

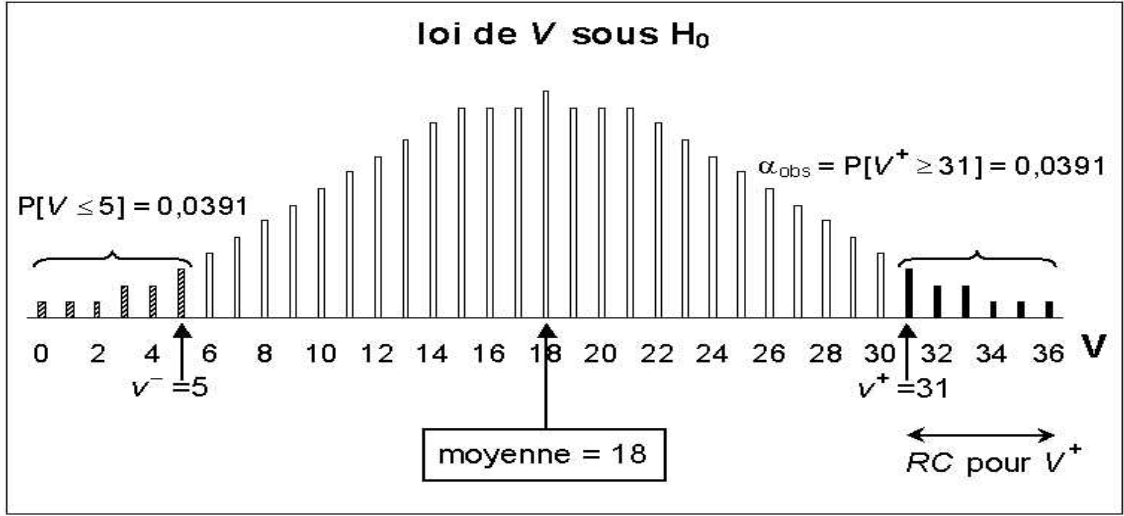
Sous $H_1 : X \succ Y$ les valeurs de $D = X - Y$ sont en majorité positives associées à des rangs élevés ce qui implique de grandes valeurs de V^+ , et à l'opposé de petites valeurs de V^- . La région critique du test RC est située à droite de la distribution de V^+ (à gauche pour V^-).

Si on choisit de travailler avec V^+ :

$$\alpha_{obs} = P_{H_0} [V^+ \geq v^+] = P_{H_0} [V^+ \geq 31] \stackrel{\text{sym}}{=} P_{H_0} [V \leq v^-] = P_{H_0} [V \leq 5] = 0,0391$$

(cf table de Wilcoxon)

FIG. 4 – Diagramme en bâtons de la loi de V sous H_0 pour $n = 8$



- Région critique et niveau de signification du test

Test unilatéral

$$\begin{aligned} H_1 : X \succ Y & \quad RC \text{ à droite pour } V^+ \text{ (à gauche pour } V^-) & \alpha_{obs} = P_{H_0}[V^+ \geq v^+] \stackrel{\text{sym}}{=} P_{H_0}[V \leq v^-] \\ H_1 : X \prec Y & \quad RC \text{ à gauche pour } V^+ & \alpha_{obs} = P_{H_0}[V^+ \leq v^+] \end{aligned}$$

Test bilatéral

$$\begin{aligned} H_1 : X \neq Y & \quad RC \text{ aux 2 extrémités de } V^+ \text{ (ou de } V^-) & \alpha_{obs} = 2 \times P_{H_0}[V \leq v_{min}] \\ & & \text{où } v_{min} = \min(v^+, v^-) \end{aligned}$$

- Approximation normale des statistiques de Wilcoxon sous H_0

Pour des tailles d'échantillons suffisamment grandes ($n > 20$) la distribution de V sous H_0 suit approximativement une loi normale de moyenne $\mu(V) = \frac{n(n+1)}{2}$

et de variance $var(V) = \frac{n(n+1)(2n+1)}{24}$.

On en déduit que la variable centrée et réduite : $Z = \frac{V - \mu(V)}{\sqrt{var(V)}} = \frac{V - \frac{n(n+1)}{2}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \approx \mathcal{N}(0, 1)$.

Si on choisit de travailler avec V^+ la valeur observée de Z : $z_{obs} = \frac{v^+ - \mu(V)}{\sqrt{var(V)}} \left(= -\frac{v^- - \mu(V)}{\sqrt{var(V)}} \right)$

- Région critique et niveau de signification du test

Si on choisit de travailler avec V^+ (dans le cas contraire, pour V^- le signe de Z et les inégalités sont inversés)

Test unilatéral

$$\begin{aligned} H_1 : X \succ Y & \quad RC \text{ à droite pour } V^+ \text{ et pour } Z & \alpha_{obs} = P_{H_0}[V^+ \geq v^+] \simeq P_{H_0}[Z \geq z_{obs}] \\ H_1 : X \prec Y & \quad RC \text{ à gauche pour } V^+ \text{ et pour } Z & \alpha_{obs} = P_{H_0}[V^+ \leq v^+] \simeq P_{H_0}[Z \leq z_{obs}] \end{aligned}$$

Test bilatéral

$$H_1 : X \neq Y \quad RC \text{ aux 2 extrémités de } V^+ \text{ et de } Z \quad \alpha_{obs} \simeq 2 \times P_{H_0}[Z \geq |z_{obs}|]$$

Exemple 4

Approximation normale sous H_0

Si on approxime la loi de V par la loi normale de moyenne $\mu(V) = \frac{n(n+1)}{4} = \frac{8 \times 9}{4} = 18$ et de variance $var(V) = \frac{n(n+1)(2n+1)}{24} = \frac{8 \times 9 \times 17}{24} = 51$
la statistique de test $Z = \frac{V - \mu(V)}{\sqrt{var(V)}} = \frac{V - 18}{\sqrt{51}} \approx \mathcal{N}(0, 1)$ (mais cette approximation est peu justifiée ici puisque $n = 8 < 20$).

On choisit de travailler avec V^+ donc la valeur observée de Z :

$$z_{obs} = \frac{V^+ - \mu(V)}{\sqrt{\text{var}(V)}} = \frac{31-18}{\sqrt{51}} \simeq 1,82036$$

Région critique et niveau de signification du test

Sous $H_1 : X \succ Y$ RC se situe à droite de la distribution de V^+ et de Z .

$$\alpha_{obs} \simeq P[Z \geq 1,82036] = 1 - P[Z \leq 1,82036] = 1 - 0,96565 = 0,03435 \quad (\text{cf Table de la loi } \mathcal{N}(0,1)).$$

3.5 Décision et conclusion

- Règle de décision basée sur α_{obs}

- si $\alpha_{obs} > \alpha$ on conserve H_0 (on ne valide pas H_1) au seuil α et au risque de 2^{de} espèce β inconnu
- si $\alpha_{obs} \leq \alpha$ on rejette H_0 en faveur de H_1 (on valide H_1) au risque α et au niveau de signification (p -valeur) α_{obs}

- Conclusion

Exemple 4

Décision et conclusion

loi exacte : $\alpha_{obs} = 3,91\% < \alpha = 5\%$ donc on rejette H_0 en faveur de H_1 au risque $\alpha = 5\%$

approximation normale : $\alpha_{obs} \simeq 3,44\% < \alpha = 5\%$ donc on rejette H_0 en faveur de H_1 au risque $\alpha = 5\%$.

NB : dans ce cas l'approximation normale est peu justifiée puisque $n < 20$, on préférera donc conclure à partir de la décision découlant de la loi exacte. Cependant, on remarque qu'ici le niveau de signification obtenu avec l'approximation normale est proche de celui du test exact.

On peut conclure que les aînés ont tendance à être plus indépendants que leurs cadets, au risque 5% et au niveau de signification $\alpha_{obs} = 3,91\%$.

Remarques :

- le test du signe avec $\alpha_{obs} = 14,45\%$ aboutissait à une conclusion inverse.
C'est une constatation fréquente puisque le test des signes et rangs de Wilcoxon est plus puissant (moins conservateur) que le test du signe.
- le test de Student de comparaison de deux moyennes pour deux échantillons appariés sous les conditions de normalité de $D = X - Y$ (tests de normalité non significatifs) donnerait $t_{obs} = 2,13498$ et $\alpha_{obs} = P_{H_0}[T \geq t_{obs}] = P_{H_0}[T \geq 2,13498] = 0,03264 \simeq 3,26\%$ d'après la loi de Student T_8 et induirait une conclusion identique.
Cependant, étant donné les petites tailles d'échantillons, les tests vérifiant la normalité sont peu puissants pour détecter un écart à ces conditions, leurs résultats sont peu fiables ; il est donc préférable de se référer au test non paramétrique présenté ici.

Résultats obtenus avec STATISTICA

ATTENTION : *STATISTICA* donne la taille des échantillons appariés n' mais ne précise ni le nombre de différences nulles (ou de valeurs $x_i = y_i$) ni la taille de l'échantillon n à partir de laquelle les calculs sont réalisés.

STATISTICA calcule la valeur observée de la statistique $V_{min} = \min(V^+, V^-)$ mais il ne calcule pas le niveau de signification exact correspondant.

STATISTICA calcule systématiquement, quelle que soit la taille de l'échantillon, la valeur observée de Z approximation normale de $V_{max} = \max(V^+, V^-)$ (z_{obs} est toujours positive) sans correction de continuité ni correction pour les ex aequo.

STATISTICA ne calcule que la p -valeur bilatérale ; pour obtenir le niveau de signification d'un test unilatéral il faut diviser la valeur donnée par 2.

Exemple 5

Huit adolescentes de 13 à 18 ans présentant une carence en fer sans anémie ont reçu un traitement pour améliorer leur fonction cognitive. On mesure cette fonction par le score au test d'apprentissage verbal de Hopkin avant et après traitement. On obtient les valeurs du tableau suivant

score avant traitement	5	3,1	2,8	1,5	1,8	0,5	3,7	6
score avant après traitement	3,1	6	6,7	1,5	3,5	5,5	4,3	5,7

Peut-on conclure, au risque $\alpha = 0,10$ que le traitement est efficace ?

Contexte

$\mathcal{P} = \{\text{adolescentes ayant une carence en fer sans anémie, sous traitement}\}$

$X = \text{score de Hopkin avant traitement}$ $Y = \text{score de Hopkin après traitement}$

Les 2 variables mesurent le même caractère (fonction cognitive) et sont quantitatives continues.

Hypothèses et risque

On souhaite tester si le traitement est efficace, donc si les scores Y après traitement sont globalement supérieurs aux scores X avant traitement, noté $Y \succ X$ ou $X \prec Y$.

On teste $\begin{cases} H_0 : \text{le traitement n'est pas efficace} \\ H_1 : \text{le traitement est efficace} \end{cases}$ ce qui s'écrit $\begin{cases} H_0 : X \equiv Y \\ H_1 : X \prec Y \end{cases}$

On définit la variable auxiliaire "différence" : $D = X - Y$.

C'est une variable quantitative continue supposée symétrique.

Test des signes et rangs de Wilcoxon

Observations

On dispose de deux échantillons appariés de taille $n' = 8$ puis d'un échantillon $\mathcal{E}$ de $D = X - Y$ de taille $n = 7$ pour lequel on a calculé les signes $\text{signe}(d_i)$.

Il faut compléter en calculant les valeurs absolues des différences $|d_i|$ et les rangs de ces valeurs absolues $\text{rang}(|d_i|)$ de 1 à $n = 7$.

x_i	5	3,1	2,8	1,5	1,8	0,5	3,7	6
y_i	3,1	6	6,7	1,5	3,5	5,5	4,3	5,7
d_i	1,9	-2,9	-3,9	0	-1,7	-5	-0,6	0,3
$\text{signe}(d_i)$	+	-	-	X	-	-	-	+
$ d_i $	1,9	2,9	3,9	X	1,7	5	0,6	0,3
$\text{rang}(d_i)$	4	5	6	X	3	7	2	1

On vérifie que les rangs vont de 1 à 7.

Statistiques de Wilcoxon

V^+ = somme des $\text{rang}(|d_i|)$ pour $d_i > 0$ et V^- = somme des $\text{rang}(|d_i|)$ pour $d_i < 0$

Les valeurs observées des statistiques de Wilcoxon

pour V^+ : $v^+ = 4 + 1 = 5$

pour V^- : $v^- = 2 + 3 + 5 + 6 + 7 = 23$

Vérification : $v^+ = 5$, $v^- = 23$ donc $v^+ + v^- = 28$. Or $n = 7$ donc $\frac{n(n+1)}{2} = \frac{7 \times 8}{2} = 28$.

V^+ et V^- sont quantitatives discrètes et varient sur le domaine $\{0, 1, \dots, 28\}$.

Lois sous H_0 , région critique et niveau de signification du test

- loi exacte

Sous $H_0 : X \equiv Y$ V^+ et V^- ont la même distribution, définie sur $\{0, 1, \dots, 28\}$ symétrique autour de la moyenne (et milieu) $\frac{n(n+1)}{4} = \frac{28}{2} = 14$.

Sous $H_1 : X \prec Y$ les valeurs de $D = X - Y$ sont en majorité négatives associées à des rangs élevés ce qui implique de grandes valeurs de V^- , et à l'opposé de petites valeurs de V^+ ; RC est à gauche de la distribution de V^+ (à droite pour V^-).

Si on choisit de travailler avec V^+ :

$\alpha_{obs} = P_{H_0}[V^+ \leq v^+] = P_{H_0}[V^+ \leq 5] = 0,0781$ (cf table de Wilcoxon)

- approximation normale

Si on approxime la loi de V par la loi normale de moyenne $\mu(V) = \frac{n(n+1)}{4} = \frac{7 \times 8}{4} = 14$ et de variance

$\text{var}(V) = \frac{n(n+1)(2n+1)}{24} = \frac{7 \times 8 \times 15}{24} = 35$

la statistique de test $Z = \frac{V - \mu(V)}{\sqrt{\text{var}(V)}} = \frac{V - 14}{\sqrt{35}} \stackrel{\text{approx}}{\sim} \mathcal{N}(0, 1)$

(mais cette approximation est peu justifiée ici puisque $n = 7 < 20$).

On choisit de travailler avec V^+ donc la valeur observée de Z :

$$z_{obs} = \frac{V^+ - \mu(V)}{\sqrt{\text{var}(V)}} = \frac{5 - 14}{\sqrt{35}} \simeq -1,521278$$

Sous $H_1 : X < Y$ RC est à gauche pour V^+ et pour Z .

$\alpha_{obs} \simeq P[Z \leq -1,521278] = 1 - P[Z \leq 1,521278] = 1 - 0,9359 = 0,0641$ (cf Table de la loi $\mathcal{N}(0, 1)$).

Décision et conclusion

loi exacte : $\alpha_{obs} = 7,81\% < \alpha = 10\%$ donc on rejette H_0 en faveur de H_1 au risque $\alpha = 10\%$.

approximation normale : $\alpha_{obs} = 6,41\% < \alpha = 10\%$ décision identique à la précédente.

NB : dans ce cas l'approximation normale est peu justifiée puisque $n < 20$, on préférera donc conclure à partir de la décision découlant de la loi exacte. Cependant, on remarque qu'ici encore le niveau de signification obtenu avec l'approximation normale est proche de celui du test exact.

On peut conclure à l'efficacité du traitement sur l'amélioration de la fonction cognitive des adolescentes présentant une carence en fer, au risque $\alpha = 10\%$ et au niveau de signification $\alpha_{obs} = 7,81\%$.

Remarques :

– le test du signe avec $\alpha_{obs} = 22,66\%$ aboutissait à une conclusion opposée.

C'est une constatation fréquente puisque le test des signes et rangs de Wilcoxon est plus puissant (moins conservateur) que le test du signe.

– le test de Student de comparaison de deux moyennes pour deux échantillons appariés sous les conditions de normalité de $D = X - Y$ (tests de normalité non significatifs) donnerait $t_{obs} = -1,80888$ et $\alpha_{obs} = P_{H_0}[T \geq t_{obs}] = P_{H_0}[T \leq -1,80888] = 0,0567 \simeq 5,67\%$ d'après la loi de Student T_7 et induirait une conclusion identique.

Cependant, étant donné les petites tailles d'échantillons, les tests vérifiant la normalité sont peu puissants pour détecter un écart aux conditions, leurs résultats sont peu fiables ; il est donc préférable de se référer au test non paramétrique présenté ici.

3.6 Remarques

- Pour des variables X et Y continues, la probabilité de trouver un individu ayant la même valeur de X et de Y est nulle.

De même la probabilité de trouver deux individus (ou plus) ayant la même valeur de D est nulle, en théorie les ex aequo n'existent pas.

En pratique, ces cas de figure sont produits par les arrondis ou les erreurs de mesure :

- on ignore les individus donnant une valeur de D nulle (ils n'apportent pas d'information sur la différence entre X et Y),

- les formules (donc les tables) changent en présence d'ex aequo (voir logiciels). Il est important de les utiliser uniquement s'il n'y a que "peu" d'ex aequo.

- Le test du signe est plus conservateur (conserve plus souvent H_0) ou moins puissant (détecte moins souvent H_1) que le test de Wilcoxon.

Démarche à suivre :

- faire le **test du signe** dont le champ d'application est très étendu puisqu'il ne nécessite aucune condition ;
- si la variable $D = X - Y$ peut être supposée symétrique et si les données le permettent (on ne dispose pas que des signes), faire le **test de Wilcoxon** ;

- le plus souvent les décisions concordent, mais si le test du signe conserve H_0 alors que le test de Wilcoxon rejette H_0 , on garde cette dernière décision puisque le test de Wilcoxon est plus puissant.