# CORRIGES DES EXERCICES : Tests de Wilcoxon "signes et rangs" pour la médiane

## *Exercice 3.1*

$\mathcal{P}$ ={personnes atteintes de schizophrénie} de taille N inconnue, X = volume de l'hippocampe gauche (en cm$^3$), variable quantitative de médiane $M$ **inconnue** dans $\mathcal{P}$ ;
$\mathcal{P}_0$ ={personnes (non atteintes de schizophrénie)}, X de médiane $M_0$ **connue** dans $\mathcal{P}_0$ : $M_0$=1,75 cm$^3$

Une diminution du volume de l'hippocampe gauche chez les personnes atteintes de schizophrénie se traduit par le fait que **le volume médian** de l'hippocampe gauche des personnes atteintes de schizophrénie est inférieur à celui des personnes non atteintes de schizophrénie, c'est à dire **inférieur à** 1,75 cm$^3$ : ce sera l'hypothèse à accepter avec un minimum de risque, donc l'hypothèse alternative, que l'on acceptera à tort au risque d'erreur maximum de $\alpha$=1%

**Comparaison de la médiane *M inconnue* à la valeur théorique $M_0$=1,75**

hypothèse nulle $H_0$ : "le volume médian $M$ dans $\mathcal{P}$ est égal à celui dans $\mathcal{P}_0$ " s'écrit
$H_0 : M = M_0 = 1{,}75$
hypothèse alternative $H_1$ : " le volume médian $M$ dans $\mathcal{P}$ est inférieur à celui dans $\mathcal{P}_0$ "
$H_1 : M < M_0 = 1{,}75$ unilatérale gauche, au risque $\alpha$=1%

1. *Peut-on à l'aide d'un test de rangs et au risque $\alpha$=1% , accepter l'hypothèse d'une diminution du volume de l'hippocampe gauche chez les personnes atteintes de schizophrénie ?*

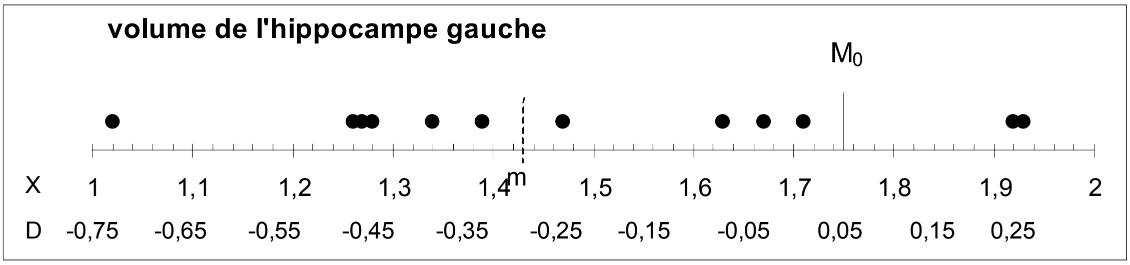
**Comparaison d'une médiane à une valeur théorique : test de Wilcoxon (signes et rangs)**

On suppose que le volume X est symétrique dans $\mathcal{P}$ (par rapport à sa médiane $M$)

| personne | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| volume X | 1,27 | 1,63 | 1,47 | 1,39 | 1,93 | 1,26 | 1,71 | 1,67 | 1,28 | 1,92 | 1,02 | 1,34 |
| *rang*(X) | 3 | 8 | 7 | 6 | 12 | 2 | 10 | 9 | 4 | 11 | 1 | 5 |
| D=X−1,75 | −0,48 | −0,12 | −0,28 | −0,36 | 0,18 | −0,49 | −0,04 | −0,08 | −0,47 | 0,17 | −0,73 | −0,41 |
| *signe*(D) | − | − | − | − | + | − | − | − | − | + | − | − |
| *rang*(\|D\|) | 10 | 3 | 6 | 7 | 5 | 11 | 1 | 2 | 9 | 4 | 12 | 8 |

pour $n$ pair la médiane observée $m$ de X est égale à toute valeur comprise entre la $\left(\frac{n}{2}\right)^{\text{ème}}$ et la $\left(\frac{n}{2}+1\right)^{\text{ème}}$ valeur ordonnée de X, c'est à dire de rang $\left(\frac{n}{2}\right)$ et $\left(\frac{n}{2}+1\right)$ ; on choisit en général le milieu de ces deux valeurs

pour $n$=12 la médiane $m$ se trouve entre la 6$^{\text{ème}}$ et la 7$^{\text{ème}}$ valeur ordonnée de X donc entre les valeurs de rang 6 et 7 : $m$ se situe entre 1,39 et 1,47 cm$^3$ (milieu $m$= 1,43 cm$^3$)




Echantillon de D où D=X−$M_0$=X−1,75 de taille $n$=12−0=12 (nombre de valeurs de D non nulles)

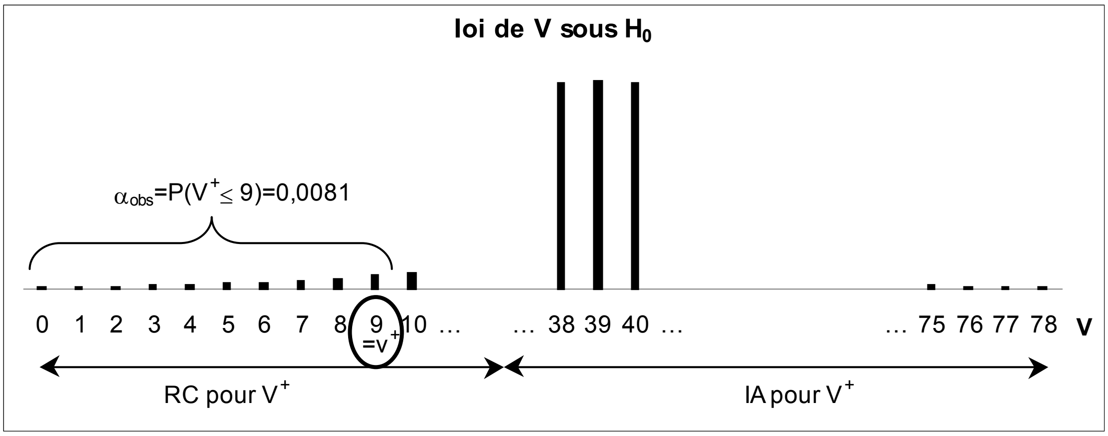
Statistiques de test $V^+$ somme des rangs de |D| associés aux valeurs positives de D, et $V^-$ :

somme $V^+ + V^- = 12\times 13/2 = 78$
valeur observée de $V^+$ : $v^+ = 5+4=9$
valeur observée de $V^-$ : $v^- = 78-9 = 69$

① ***loi exacte de la statistique de test*** $\mathbf{V}^+$ ***sous l'hypothèse nulle*** (pas d'ex æquo et $n$=12 < 20)

sous $H_0$ : $M = M_0$

la statistique de test $V^+$ est quantitative discrète sur {0, …, 78} symétrique, de moyenne 78/2= 39




***- degré de signification du test unilatéral gauche (région de rejet à gauche du domaine de variation de*** $\mathbf{V}^+$***)***

le degré de signification est la probabilité sous $H_0$ d'observer une valeur de la statistique de test $V^+$ au plus égale à la valeur $v^+$=9 observée car sous l'hypothèse $H_1$ on attend "peu" de valeurs positives de D avec des rangs faibles, donc de "petites" valeurs de $V^+$

$\alpha_{obs} = P_{H_0}(V^+ \le v^+) = P_{H_0}(V^+ \le 9) = 0{,}0081$

car $v=v^+=9<n=12$ table 2 colonne $n$=12 et ligne v=9

***- décision***

$\alpha_{obs} < 0{,}0081 \le \alpha=1\%$ donc on rejette $H_0$ en faveur de $H_1$ au risque $\alpha$=1%

② ***loi de la statistique de test sous l'hypothèse nulle :*** approximation normale ($n$=12 ≥10)

sous $H_0$ : $M = M_0$ les statistiques $V^+$ et $V^-$ suivent approximativement une loi normale

de moyenne $\frac{n(n+1)}{4}$=78/2=39 et de variance $\frac{n(n+1)(2n+1)}{24}$=12×13×25/24=13×25/2=162,5

donc la statistique de test $Z = \dfrac{V^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ suit approximativement une loi normale $\mathcal{N}(0,1)$

***- valeur observée de la statistique de test*** : $z = \dfrac{9-39}{\sqrt{162{,}5}} = -2{,}35$

**a.** ***région de rejet du test unilatéral gauche au risque*** $\alpha$***=1%***

$IR_{1\%} = ]-\infty\,;\,-z_{0,99}\,] = ]-\infty\,;\,-2{,}32]$ où $z_{0,99}$=2,32 est le quantile d'ordre 1−α=99% de la loi $\mathcal{N}(0,1)$

décision : on observe z= −2,35 donc $z \in IR_{1\%}$ donc on rejette $H_0$ en faveur de $H_1$ au risque $\alpha$=1%

**b.** ***degré de signification du test unilatéral gauche (région de rejet à gauche du domaine de variation de Z)***

le degré de signification est la probabilité sous $H_0$ d'observer une valeur de la statistique de test Z au plus égale à la valeur z=−2,35 observée car sous $H_1$ on attend de "petites" valeurs de Z

$\alpha_{obs} = P_{H_0}(Z \le -2{,}35) = F_Z(-2{,}35) = 1 - F_Z(2{,}35) = 1 - 0{,}9906 = 0{,}0094$

où $F_Z$ est la fonction de répartition de la loi $\mathcal{N}(0,1)$

décision : $\alpha_{obs} = 0{,}0094 \le \alpha=1\%$ donc on rejette $H_0$ en faveur de $H_1$ au risque $\alpha$=1%

- ***conclusion***

le volume médian de l'hippocampe gauche des personnes atteintes de schizophrénie est significativement inférieur à celui des personnes non atteintes de schizophrénie, au risque $\alpha$=1% et au degré de signification $\alpha_{obs}\approx$1%

*2. Même question avec le test du signe et au risque $\alpha$=1% ?*

**Comparaison d'une médiane à une valeur théorique : test du signe pour la médiane**

| personne | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| volume X | 1,27 | 1,63 | 1,47 | 1,39 | 1,93 | 1,26 | 1,71 | 1,67 | 1,28 | 1,92 | 1,02 | 1,34 |
| *rang*(X) | 3 | 8 | 7 | 6 | 12 | 2 | 10 | 9 | 4 | 11 | 1 | 5 |
| D=X−1,75 | −0,48 | −0,12 | −0,28 | −0,36 | 0,18 | −0,49 | −0,04 | −0,08 | −0,47 | 0,17 | −0,73 | −0,41 |
| *signe*(D) | − | − | − | − | + | − | − | − | − | + | − | − |

La variable *signe*(D) où D=X−$M_0$=X−1,75 est qualitative dichotomique définie sur E={positif, négatif}, $p$ proportion de signes strictement positifs **inconnue** dans $\mathcal{P}$

Comparaison de la proportion $p$ *inconnue* à la valeur théorique $p_0$=50%=0,5
hypothèse nulle $H_0 : M = M_0$=1,75 s'écrit $H_0 : p = p_0$= 0,5
hypothèse alternative $H_1 : M < M_0$=1,75 s'écrit $H_1 : p < p_0$ =0,5 unilatérale gauche, au risque $\alpha$=1%

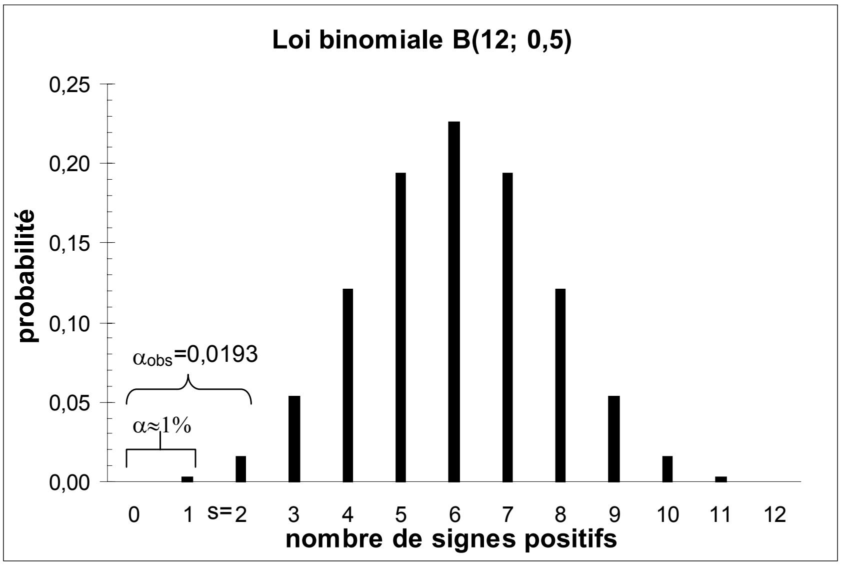
Echantillon de X issu de $\mathcal{P}$ de taille $n$=12−0=12 (nombre de valeurs de D non nulles) : estimation ponctuelle de $p$ par $f$ fréquence observée de signes strictement positifs dans l'échantillon où $s$=2 donc $f$ = 2/12 = 0,167

***- loi de la statistique de test sous l'hypothèse nulle***

sous $H_0 : p = p_0 = 0,5$

la statistique de test $S_n$ effectif empirique de signes strictement positifs, quantitative discrète définie sur $\{0, \ldots, n\} = \{0, \ldots, 12\}$ suit une loi binomiale $\mathcal{B}(n, p_0) = \mathcal{B}(12; 0,5)$ symétrique, de moyenne $np_0 = 12 \times 0,5 = 6$

| | $S_n \sim \mathcal{B}$ (12; 0,5) | |
|---|---|---|
| $s$ | $P(S_n = s)$ | $P(S_n \le s)$ |
| 0 | 0,0002 | 0,0002 |
| 1 | 0,0029 | 0,0032 |
| 2 | 0,0161 | 0,0193 |
| 3 | 0,0537 | 0,0730 |
| 4 | 0,1208 | 0,1938 |
| 5 | 0,1934 | 0,3872 |
| 6 | 0,2256 | 0,6128 |
| 7 | 0,1934 | 0,8062 |
| 8 | 0,1208 | 0,9270 |
| 9 | 0,0537 | 0,9807 |
| 10 | 0,0161 | 0,9968 |
| 11 | 0,0029 | 0,9998 |
| 12 | 0,0002 | 1,0000 |




***- degré de signification du test unilatéral gauche (région de rejet à gauche du domaine de variation de*** **$S_n$*)***

le degré de signification est la probabilité sous $H_0$ d'observer une valeur de la statistique de test $S_n$ au plus égale à la valeur $s$=2 observée car sous $H_1$ on attend "peu" de valeurs de D positives, donc on prendra le risque de conclure $H_1$ (rejeter $H_0$ en faveur de $H_1$) à tort pour les "petites" valeurs de $S_n$

$\alpha_{obs} = P_{H_0}(S_n \le s) = P_{H_0}(S_n \le 2) = F_{\mathcal{B}}(2) = 0,0193$

où $F_{\mathcal{B}}$ fonction de répartition de la loi $\mathcal{B}(12; 0,5)$ donnée dans la table 1 (ligne $n$=12 colonne $k$=2)

***- décision***

$\alpha_{obs} = 0,0193 > \alpha$=1%

$\Rightarrow$ on ne rejette pas $H_0$ au seuil de signification $\alpha$=1% et au risque d'erreur de seconde espèce $\beta$ inconnu ; le test n'est pas significatif au seuil $\alpha$=1% et au risque $\beta$

***- conclusion***

la proportion $p$ de signes positifs de la différence de X à la valeur théorique ($M_0$ =) 1,75 $cm^3$ n'est pas significativement inférieure à 50%, au risque $\alpha$=1% ; donc le volume médian des personnes atteintes de schizophrénie n'est pas significativement inférieur à celui des personnes non atteintes de schizophrénie, au seuil $\alpha$=1% et au risque $\beta$ inconnu

***remarques***

- les degrés de signification du test de Wilcoxon obtenus avec la table exacte de la loi ($\alpha_{obs} = 0{,}0081$) et avec l'approximation normale ($\alpha_{obs} = 0{,}0094$) sont très proches et aboutissent à la même conclusion
- les résultats des deux tests, signe et Wilcoxon, sont différents car si la variable étudiée est symétrique, le test de Wilcoxon est plus puissant que le test du signe puisqu'il tient compte non seulement du signe des différences à la médiane, mais aussi de leurs grandeurs (tailles) relatives

## ***Exercice 3.2***

$\mathcal{P}$ ={personnes ayant absorbé la drogue} de taille N inconnue, X = force de préhension, variable quantitative de médiane $M$ **inconnue** dans $\mathcal{P}$ ;
$\mathcal{P}_0$ ={personnes ayant absorbé la drogue }, X de médiane $M_0$ **connue** dans $\mathcal{P}_0$ : $M_0$=18,2

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

Le produit est efficace s'il augmente la force musculaire mesurée par la force de préhension se traduit par le fait que **la force de préhension médiane** des personnes ayant absorbé la drogue est supérieure à celle des personnes n'ayant pas absorbé la drogue, c'est à dire **supérieure à** 18,2 : ce sera l'hypothèse à accepter avec un minimum de risque, donc l'hypothèse alternative, que l'on acceptera à tort au risque d'erreur maximum de $\alpha$=5%

**Comparaison de la médiane *M inconnue* à la valeur théorique $M_0$=18,2**

hypothèse nulle $H_0$ : "la force de préhension médiane $M$ dans $\mathcal{P}$ est égale à celle dans $\mathcal{P}_0$ " s'écrit
$H_0 : M = M_0 = 18{,}2$

hypothèse alternative $H_1$ : "la force de préhension médiane $M$ dans $\mathcal{P}$ est supérieur à celle dans $\mathcal{P}_0$ "
$H_1 : M > M_0 = 18{,}2$ unilatérale droite, au risque $\alpha$=5%

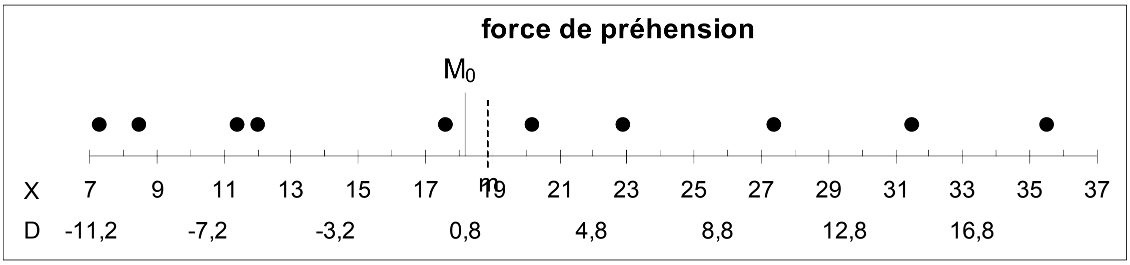
**❶ Comparaison d'une médiane à une valeur théorique : test de Wilcoxon (signes et rangs)**

On suppose que le résultat X est symétrique dans $\mathcal{P}$ (par rapport à sa médiane $M$)

| Sujet | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Résultat X | 20,2 | 11,4 | 8,5 | 27,4 | 31,5 | 17,6 | 7,3 | 35,5 | 12 | 22,9 |
| *Rang*(X) | 6 | 3 | 2 | 8 | 9 | 5 | 1 | 10 | 4 | 7 |
| D=X−1,75 | 2 | −6,8 | −9,7 | 9,2 | 3,3 | −0,6 | −10,9 | 17,3 | −6,2 | 4,7 |
| *Signe*(D) | + | − | − | + | + | − | − | + | − | + |
| *Rang*(\|D\|) | 2 | 5 | 7 | 6 | 9 | 1 | 8 | 10 | 4 | 3 |

pour $n$ pair la médiane observée $m$ de X est égale à toute valeur comprise entre la $\left(\frac{n}{2}\right)^{\text{ème}}$ et la $\left(\frac{n}{2}+1\right)^{\text{ème}}$ valeur ordonnée de X, c'est à dire de rang $\left(\frac{n}{2}\right)$ et $\left(\frac{n}{2}+1\right)$ ; on choisit en général le milieu de ces deux valeurs

pour $n$=10 la médiane $m$ se trouve entre la 5$^{\text{ème}}$ et la 6$^{\text{ème}}$ valeur ordonnée de X donc entre les valeurs de rang 5 et 6 : $m$ se situe entre 17,6 et 20,2 (milieu $m$= 18,9)




Echantillon de D où D=X−$M_0$=X−18,2 de taille $n$=10−0=10 (nombre de valeurs de D non nulles)

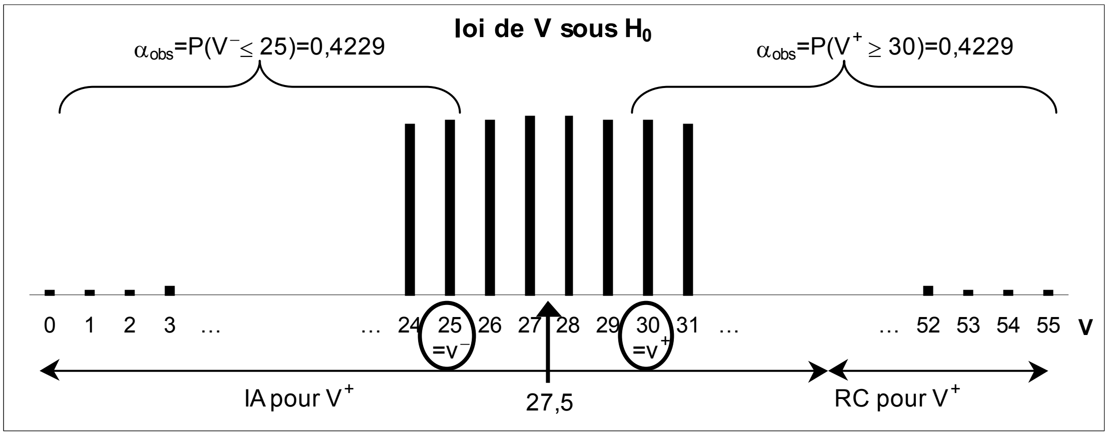
Statistiques de test $V^+$ somme des rangs de |D| associés aux valeurs positives de D, et $V^-$ :

somme $V^+ + V^- = 10\times 11/2 = 55$
valeur observée de $V^+$ : $v^+ = 2+6+9+10+3=30$
valeur observée de $V^-$ : $v^- = 55-30 = 25$

① ***loi exacte de la statistique de test*** $\mathbf{V^+}$ ***sous l'hypothèse nulle*** (pas d'ex æquo et $n$=10 < 20)

sous $H_0$ : $M = M_0$

la statistique de test $V^+$ est quantitative discrète sur {0, …, 55} symétrique, de moyenne 55/2= 27,5




***- degré de signification du test unilatéral droit (région de rejet à droite du domaine de variation de*** $\mathbf{V^+}$***)***

le degré de signification est la probabilité sous $H_0$ d'observer une valeur de la statistique de test $V^+$ au moins égale à la valeur $v^+$=30 observée car sous l'hypothèse $H_1$ on attend "beaucoup" de valeurs positives de D avec des rangs élevés, donc de "grandes" valeurs de $V^+$

$\alpha_{obs} = P_{H_0}(V^+ \geq v^+) = P_{H_0}(V^+ \geq 30) = 1 - P_{H_0}(V^+ \leq 29) = 1 - 0{,}5771 = 0{,}4229$

car v=29>$n$=10 table 3 colonne $n$=10 et ligne v−$n$=29−10=19 donc $P_{H_0}(V^+ \leq 29) = 0{,}5771$

ou $\alpha_{obs} = P_{H_0}(V^+ \geq v^+) = P_{H_0}(V^- \leq v^-) = P_{H_0}(V^- \leq 25) = 0{,}4229$

car v=$v^-$=25>$n$=10 table 3 colonne $n$=10 et ligne v−$n$=25−10=15

***- décision***

$\alpha_{obs}$ < 0,4229 > α=5% donc on ne rejette pas $H_0$ (on conserve $H_0$) au seuil α=5% et au risque d'erreur de seconde espèce β inconnu ; le test n'est pas significatif au seuil α=5% et au risque β

② ***loi de la statistique de test sous l'hypothèse nulle*** *:* approximation normale ($n$=10 ≥10)

sous $H_0$ : $M = M_0$ les statistiques $V^+$ et $V^-$ suivent approximativement une loi normale

de moyenne $\frac{n(n+1)}{4}$=55/2=27,5 et de variance $\frac{n(n+1)(2n+1)}{24}$=10×11×21/24=2310/24=96,25

donc la statistique de test $Z = \dfrac{V^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ suit approximativement une loi normale $\mathcal{N}(0,1)$

***- valeur observée de la statistique de test*** : $z = \dfrac{30 - 27{,}5}{\sqrt{96{,}25}} = 0{,}25$

**a.** ***région de rejet du test unilatéral droit au risque*** $\boldsymbol{\alpha}$***=5%***

$IR_{1\%}$ = [$z_{0,95}$ ; +∞ [ = [1,645 ; +∞ [ où $z_{0,95}$=1,645 est le quantile d'ordre 1−α=95% de la loi $\mathcal{N}(0,1)$

décision : on observe z= 0,25 donc z ∉ $IR_{5\%}$ donc on ne rejette pas $H_0$ (on conserve $H_0$) au seuil α=5% et au risque d'erreur de seconde espèce β inconnu ; le test n'est pas significatif au seuil α=5% et au risque β

**b.** ***degré de signification du test unilatéral droit (région de rejet à droite du domaine de variation de Z)***

le degré de signification est la probabilité sous $H_0$ d'observer une valeur de la statistique de test Z au moins égale à la valeur z=0,255 observée car sous $H_1$ on attend de "grandes" valeurs de Z

$\alpha_{obs} = P_{H_0}(Z \geq 0{,}25) = 1 - F_Z(0{,}25) = 1 - 0{,}5987 = 0{,}4013$

où $F_Z$ est la fonction de répartition de la loi $\mathcal{N}(0,1)$

décision : $\alpha_{obs}$ = 0,4013 ≥ α=5% donc on ne rejette pas $H_0$ (on conserve $H_0$) au seuil α=5% et au risque d'erreur de seconde espèce β inconnu ; le test n'est pas significatif au seuil α=5% et au risque β

- ***conclusion***

la force de préhension médiane des personnes ayant absorbé la drogue n'est pas significativement supérieure à celle des personnes n'ayant pas absorbé la drogue, au seuil α=5% et au risque β inconnu

**❷ Comparaison d'une médiane à une valeur théorique : test du signe pour la médiane**

| sujet | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| résultat X | 20,2 | 11,4 | 8,5 | 27,4 | 31,5 | 17,6 | 7,3 | 35,5 | 12 | 22,9 |
| *rang*(X) | 6 | 3 | 2 | 8 | 9 | 5 | 1 | 10 | 4 | 7 |
| D=X−1,75 | 2 | −6,8 | −9,7 | 9,2 | 3,3 | −0,6 | −10,9 | 17,3 | −6,2 | 4,7 |
| *signe*(D) | + | − | − | + | + | − | − | + | − | + |

La variable *signe*(D) où $D=X-M_0=X-18{,}2$ est qualitative dichotomique définie sur E={positif, négatif}, $p$ proportion de signes strictement positifs **inconnue** dans $\mathcal{P}$

Comparaison de la proportion $p$ *inconnue* à la valeur théorique $p_0$=50%=0,5

hypothèse nulle $H_0 : M = M_0=18{,}2$ s'écrit $H_0 : p = p_0= 0{,}5$

hypothèse alternative $H_1 : M > M_0=18{,}2$ s'écrit $H_1 : p > p_0 =0{,}5$ unilatérale droite, au risque $\alpha$=5%

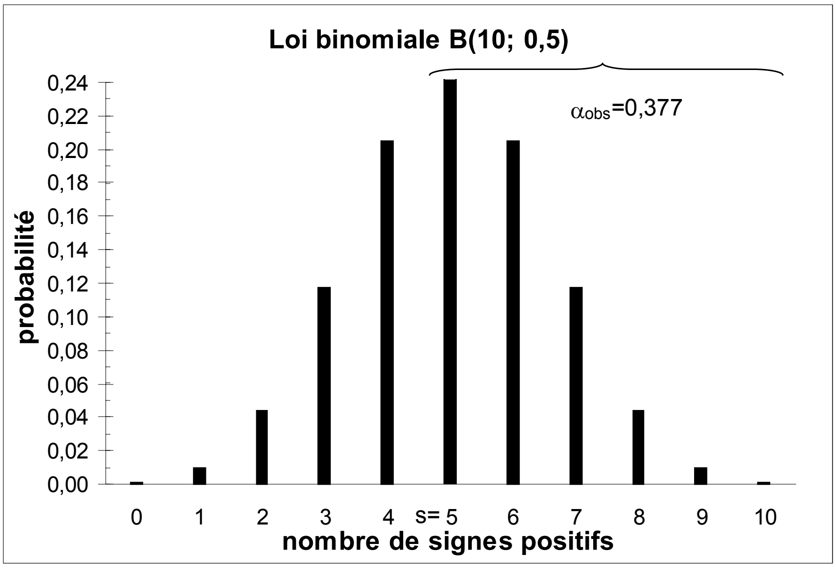
Echantillon de X issu de $\mathcal{P}$ de taille $n$=10−0=10 (nombre de valeurs de D non nulles) : estimation ponctuelle de $p$ par $f$ fréquence observée de signes strictement positifs dans l'échantillon où $s$=5 donc $f$ = 5/10 = 0,5

***- loi de la statistique de test sous l'hypothèse nulle***

sous $H_0 : p = p_0 = 0{,}5$

la statistique de test $S_n$ effectif empirique de signes strictement positifs, quantitative discrète définie sur $\{0, \ldots, n\} = \{0, \ldots, 10\}$ suit une loi binomiale $\mathcal{B}(n, p_0) = \mathcal{B}(10; 0{,}5)$ symétrique, de moyenne $np_0 = 10\times0{,}5 = 5$

| | $S_n \sim\mathcal{B}$ (10; 0,5) | |
|---|---|---|
| $s$ | $P(S_n = s)$ | $P(S_n \le s)$ |
| 0 | 0,0010 | 0,0010 |
| 1 | 0,0098 | 0,0107 |
| 2 | 0,0439 | 0,0547 |
| 3 | 0,1172 | 0,1719 |
| 4 | 0,2051 | 0,3770 |
| 5 | 0,2461 | 0,6230 |
| 6 | 0,2051 | 0,8281 |
| 7 | 0,1172 | 0,9453 |
| 8 | 0,0439 | 0,9893 |
| 9 | 0,0098 | 0,9990 |
| 10 | 0,0010 | 1,0000 |




***- degré de signification du test unilatéral droit (région de rejet à droite du domaine de variation de*** $S_n$***)***

le degré de signification est la probabilité sous $H_0$ d'observer une valeur de la statistique de test $S_n$ au moins égale à la valeur $s$=5 observée car sous $H_1$ on attend "beaucoup" de valeurs de D positives, donc on prendra le risque de conclure $H_1$ (rejeter $H_0$ en faveur de $H_1$) à tort pour les "grandes" valeurs de $S_n$

$\alpha_{obs} = P_{H_0} (S_n \ge s) = P_{H_0} (S_n \ge 5) = 1-P_{H_0} (S_n < 5) = 1-F_{\mathcal{B}}(4) = 1-0{,}377 = 0{,}623$

où $F_{\mathcal{B}}$ fonction de répartition de la loi $\mathcal{B}(10; 0{,}5)$ donnée dans la table 1 (ligne $n$=10 colonne $k$=4)

***- décision***

$\alpha_{obs} = 0{,}623 > \alpha=5\%$

⇒ on ne rejette pas $H_0$ au seuil de signification $\alpha$=5% et au risque d'erreur de seconde espèce $\beta$ inconnu ; le test n'est pas significatif au seuil $\alpha$=5% et au risque $\beta$

***- conclusion***

la proportion $p$ de signes positifs de la différence de X à la valeur théorique ($M_0$ =) 18,2 n'est pas significativement supérieure à 50%, au risque $\alpha$=5% ; donc la force de préhension médiane des personnes ayant absorbé la drogue n'est pas significativement supérieure à celle des personnes n'ayant pas absorbé la drogue, au seuil $\alpha$=5% et au risque $\beta$ inconnu

***remarques***

- les degrés de signification du test de Wilcoxon obtenus avec la table exacte de la loi ($\alpha_{obs}$ = 0,4229) et avec l'approximation normale ($\alpha_{obs}$ = 0,4013) sont très proches et aboutissent à la même conclusion

- les résultats des deux tests, signe et Wilcoxon, aboutissent à la même conclusion : si le test de Wilcoxon, plus puissant que le test du signe, n'est pas significatif, celui du signe ne sera pas significatif non plus

## *Exercice 3.3*

$\mathcal{P}$ ={agneaux en régime normal avec hormone} de taille N inconnue, X = croissance sur une période de 3 semaines, variable quantitative de médiane $M$ **inconnue** dans $\mathcal{P}$ ;
$\mathcal{P}_0$ ={ agneaux en régime normal}, X de médiane $M_0$ **connue** dans $\mathcal{P}_0$

*Peut-on au risque $\alpha$=1% , accepter l'hypothèse que l'hormone agit sur le taux de croissance ?*

L'hormone agit sur le taux de croissance si elle modifie la croissance ce qui se traduit par le fait que **la croissance médiane** des agneaux en régime normal avec hormone est différente de celle des agneaux en régime normal, c'est à dire **différente de** $M_0$: ce sera l'hypothèse à accepter avec un minimum de risque, donc l'hypothèse alternative, que l'on acceptera à tort au risque d'erreur maximum de $\alpha$=1%

**Comparaison de la médiane *M inconnue* à la valeur théorique $M_0$**

hypothèse nulle H$_0$ : "la croissance médiane $M$ dans $\mathcal{P}$ est égale à celle dans $\mathcal{P}_0$ " s'écrit
H$_0$ : $M = M_0$
hypothèse alternative H$_1$ : " la croissance médiane $M$ dans $\mathcal{P}$ est différente de celle dans $\mathcal{P}_0$ "
H$_1$ : $M \neq M_0$ bilatérale, au risque $\alpha$=1%

**❶ Comparaison d'une médiane à une valeur théorique : test de Wilcoxon (signes et rangs)**

On suppose que la croissance X est symétrique dans $\mathcal{P}$ (par rapport à sa médiane $M$)

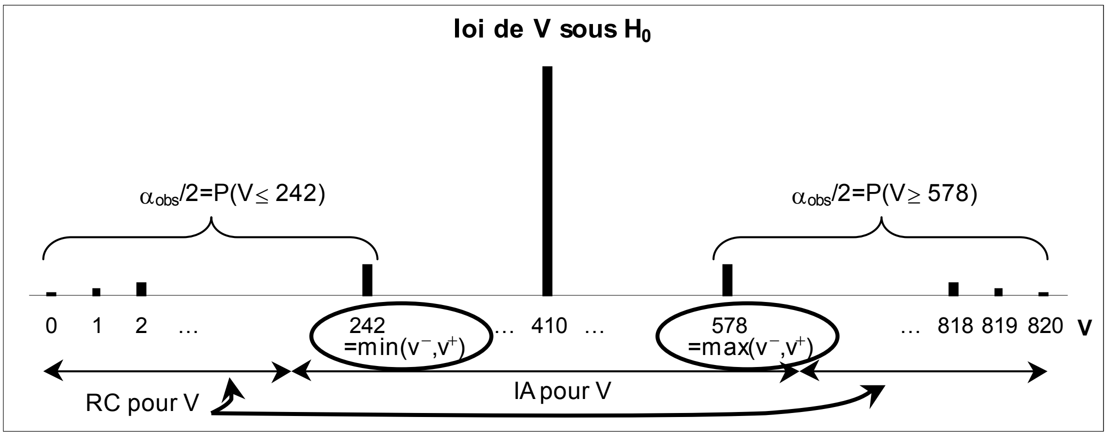
Echantillon de D où D=X–$M_0$ de taille $n$=40 (nombre de valeurs de D non nulles ?)

Statistiques de test V$^+$ somme des rangs de |D| pour les différences positives, et V$^-$ :

somme V$^+$ + V$^-$= 40×41/2=820
valeur observée de V = min(V$^-$ , V$^+$) : v=242
valeur observée de max(V$^-$ , V$^+$) : 820–242=578

**① loi exacte de la statistique de test V (V$^+$ ou V$^-$) sous l'hypothèse nulle**

sous H$_0$ : $M = M_0$ la statistique V est quantitative discrète sur {0, …, 820} symétrique, de moyenne 820/2=410




***- degré de signification du test bilatéral (région de rejet bilatérale, petites ou grandes valeurs de* V*)***

pour v = min(v$^+$, v$^-$) = 242 $\alpha_{obs} = 2\times P_{H_0}$ (V ≤ v) = $2\times P_{H_0}$ (V ≤ 242) = ? (pas de table pour $n$=40)

**② approximation normale** (justifiée car $n$ =40 ≥10)

***- loi de la statistique de test sous l'hypothèse nulle***

sous H$_0$ : " X ≡ Y " les statistiques V$^+$ et V$^-$ suivent approximativement une loi normale

de moyenne $\frac{n(n+1)}{4}$=820/2=410 et de variance $\frac{n(n+1)(2n+1)}{24}$=40×41×81/24=132 840/24=5 535

donc la statistique de test $Z = \frac{V - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ suit approximativement une loi normale $\mathcal{N}(0,1)$

**- *valeur observée de la statistique de test*** : $z = \dfrac{v - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} = \dfrac{242 - 410}{\sqrt{5535}} = -2{,}258$

**- *degré de signification du test bilatéral (région de rejet bilatérale)***

$\alpha_{obs} = 2 \times P_{H_0}(Z \le z) = 2 \times P_{H_0}(Z \le -2{,}258) \approx 2 \times F_Z(-2{,}26) = 2 \times (1 - F_Z(2{,}26)) = 2 \times (1 - 0{,}9881)$

$= 2 \times 0{,}0119 = 0{,}0238$ où $F_Z$ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$

- ***décision*** : $\alpha_{obs} = 0{,}0238 > \alpha = 1\%$ donc on ne rejette pas $H_0$ au seuil $\alpha$=1% et au risque $\beta$ inconnu

**conclusion**

la croissance médiane avec hormone n'est pas significativement différente de celle sans hormone au seuil $\alpha$=1% et au risque $\beta$ inconnu, donc l'hormone n'agit pas significativement sur la croissance au seuil $\alpha$=1% et au risque $\beta$ inconnu

**❷ Comparaison de deux distributions sur échantillons appariés : test du signe pour la médiane**

La variable *signe*(D) est qualitative dichotomique définie sur E={positif, négatif}, *p* proportion de signes strictement positifs **inconnue** dans $\mathcal{P}$,

échantillon de D issu de $\mathcal{P}$ de taille *n*=40 : estimation ponctuelle de *p* par *f* fréquence observée de signes strictement positifs dans l'échantillon que l'on ne connaît pas ici : *s*=?

Il est impossible de faire le test du signe avec ces données