

## Devoir Maison

### Exercice 1

**1.a.** On est donc avec :  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ ,  $\sigma$  connu.

Alors, le quotient  $\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma}$  suit une loi normale centrée réduite. avec  $q_{1-\alpha/2}$  le quantile de niveau  $\alpha/2$  de la loi normale on a (pour un niveau  $\alpha$  car la loi est symétrique) :

$$\begin{aligned} \mu - q_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} &< \hat{\mu} < \mu + q_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \\ \hat{\mu} - q_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} &< \mu < \hat{\mu} + q_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Soit les intervalles de confiance avec  $q_{97.5\%} = 1.96$  et  $q_{99.5\%} = 2.58$  les quantiles de niveau 5% et 1% :

$$\begin{aligned} IC_{95\%} &= \left[ 20.31 - 1.96 \frac{5}{\sqrt{8}} ; 20.31 + 1.96 \frac{5}{\sqrt{8}} \right] \\ IC_{99\%} &= \left[ 20.31 - 2.58 \frac{5}{\sqrt{8}} ; 20.31 + 2.58 \frac{5}{\sqrt{8}} \right] \end{aligned}$$

On a donc le test statistique suivant :

$$H_0 : \mu = 16 \quad \text{vs} \quad \mu \neq 16$$

qu'on peut reformuler de la manière suivante en posant :

$$T = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma} = \frac{\sqrt{8}(20.31 - 16)}{5} = 2.44$$

Si  $T \in ]q_{1-\alpha/2}^{N(0,1)}; q_{1-\alpha/2}^{N(0,1)}[$  on accepte  $H_0$ , sinon on rejette.

On a donc  $T \notin ]-1.96; 1.96[$  donc on rejette  $H_0$  avec un niveau  $\alpha = 5\%$  et  $T \in ]-2.58; 2.58[$  donc on accepte  $H_0$  avec un niveau  $\alpha = 1\%$ .

**1.b.** On a maintenant  $\sigma$  inconnu. On se place d'abord dans le cas du test :

$$H_0 : \mu = 16 \quad \text{vs} \quad H_1 : \mu \neq 16$$

On passe donc par la statistique conventionnelle :  $T = \sqrt{n} \frac{(\hat{\mu} - 16)}{\hat{\sigma}}$  avec :  $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ .

$$T = \sqrt{n} \frac{\hat{\mu} - 16}{\hat{\sigma}} = \frac{\sqrt{n} \frac{\hat{\mu} - 16}{\sigma}}{\frac{\hat{\sigma}}{\sigma}} = \frac{\sqrt{n} \frac{\hat{\mu} - 16}{\sigma}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}$$

Alors on a :

$$\begin{aligned} \sqrt{n} \frac{\hat{\mu} - 16}{\sigma} &\sim N(0, 1) \\ \frac{\hat{\sigma}^2}{\sigma^2} &= \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2} \sim \frac{\chi^2(n-1)}{n-1} \frac{n-1}{n} \end{aligned}$$

On pose donc une statistique  $\tilde{T}$  qui vérifie :

$$T = \tilde{T} \sqrt{\frac{n}{n-1}}$$

et dans ce cas,  $\tilde{T}$  est le quotient d'une loi normale centrée réduite et d'une loi de chi deux de degré  $(n-1)$  divisée par  $(n-1)$ , indépendantes entre elles. Donc  $\tilde{T}$  suit une loi de Student à  $(n-1)$  degrés de liberté. Donc :

$$\tilde{T} = T \sqrt{\frac{n-1}{n}} \sim \mathcal{T}(n-1)$$

On a donc le test :  $\Phi_\alpha^1 = \mathbb{1}_{\{|\tilde{T}| > q_{1-\alpha}^{\mathcal{T}(n-1)}\}}$  car la fonction puissance pour  $H_1 : \mu \neq 16$  est  $\mathbb{P}_{16}(|\tilde{T}| > q_{1-\alpha}^{\mathcal{T}(n-1)})$

On se place maintenant dans le cas du test :

$$H_0 : \mu \leq 16 \quad \text{vs} \quad H_1 : \mu > 16$$

Ici  $\Theta_0$  n'est pas un singleton donc  $\bar{x}_n = \hat{\mu}$  n'a pas de loi fixe sous  $H_0$ . On cherche alors le plus petit  $C \in \mathbb{R}$  qui vérifie :

$$\sup_{\mu \leq 16} \mathbb{P}_\mu(\hat{\mu} > C) \leq \alpha$$

or ;

$$\begin{aligned} \mathbb{P}_\mu(\hat{\mu} > C) &= \mathbb{P}_\mu\left(\frac{\hat{\mu} - \mu}{\sigma} > \frac{C - \mu}{\sigma}\right) \\ &= 1 - \mathbb{P}_\mu\left(\frac{\hat{\mu} - \mu}{\sigma} \leq \frac{C - \mu}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{C - \mu}{\sigma}\right), \quad \Phi \text{ CDF de } N(0, 1) \end{aligned}$$

Donc la fonction qui à  $\mu$  associe  $\mathbb{P}_\mu(\hat{\mu} > C)$  est croissante. On a donc :

$$\sup_{\mu \leq 16} \mathbb{P}_\mu(\hat{\mu} > C) = \mathbb{P}_{16}(\hat{\mu} > C)$$

Or sous  $\mu = 16$ ,  $\hat{\mu} \sim N(\mu, \frac{\sigma^2}{n})$ , soit tout comme plus haut,  $\frac{\sqrt{n}(\hat{\mu}-16)}{\sigma} \sim N(0, 1)$ . On pose alors la statistique :

$$\tilde{T} = T \sqrt{\frac{n-1}{n}} \quad \text{et} \quad T = \sqrt{n} \frac{(\hat{\mu} - 16)}{\hat{\sigma}}$$

Alors on a  $\tilde{T} = \sqrt{n-1} \frac{(\hat{\mu}-16)}{\hat{\sigma}} \sim \mathcal{T}(n-1)$ . On a donc bien  $\mathbb{P}_{\mu=16}(\tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)}) = \alpha$

On fait donc ici de même que plus haut, mais il s'agit de montrer que :

$$\sup_{\mu \leq 16} \mathbb{P}_\mu(\tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)}) \leq \alpha$$

On a alors, avec  $K = q_{1-\alpha}^{\mathcal{T}(n-1)} \frac{\hat{\sigma}}{\sqrt{n-1}} + 16$  :

$$\begin{aligned} \mathbb{P}_\mu(\tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)}) &= \mathbb{P}_\mu\left(\hat{\mu} > q_{1-\alpha}^{\mathcal{T}(n-1)} \frac{\hat{\sigma}}{\sqrt{n-1}} + 16\right) \\ &= \mathbb{P}_\mu(\hat{\mu} > K) \\ &= 1 - \mathbb{P}_\mu(\hat{\mu} \leq K) \\ &= 1 - \mathbb{E}_\mu[\mathbb{1}_{\hat{\mu} \leq K}] \\ &= 1 - \mathbb{E}_\mu[\mathbb{E}_\mu[\mathbb{1}_{\hat{\mu} \leq K} | K]] \\ &= 1 - \mathbb{E}_\mu[g(\mu, K)] \end{aligned}$$

On a donc posé  $g(\mu, K) = \mathbb{P}_\mu(\hat{\mu} \leq K)$  et donc  $1 - g(\mu, K) = \mathbb{P}_\mu(\hat{\mu} > K)$  est croissante pour  $\mu$  comme vu ci-dessus. On a donc bien :

$$\sup_{\mu \leq 16} \mathbb{P}_\mu \left( \tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)} \right) = \mathbb{P}_{\mu=16} \left( \tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)} \right) = \alpha$$

On peut donc définir le test :  $\Phi_\alpha^2 = \mathbb{1}_{\{\tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)}\}}$ .

Nos lois sont continues, on a donc :

$$\Phi_\alpha^1 = \mathbb{1}_{\{|\tilde{T}| > q_{1-\alpha}^{\mathcal{T}(n-1)}\}} = \mathbb{1}_{\{|\tilde{T}| \geq q_{1-\alpha}^{\mathcal{T}(n-1)}\}} \quad \text{et} \quad \Phi_\alpha^2 = \mathbb{1}_{\{\tilde{T} > q_{1-\alpha}^{\mathcal{T}(n-1)}\}} = \mathbb{1}_{\{\tilde{T} \geq q_{1-\alpha}^{\mathcal{T}(n-1)}\}}$$

D'après le théorème de Wassermann on a la  $p$ -value de la manière suivante :

$$p_1(x) = \sup_{\mu \leq 16} \mathbb{P}_\mu \left( |\tilde{T}| \geq q_{1-\alpha}^{\mathcal{T}(n-1)} \right) = \mathbb{P}_{\mu=16} \left( |\tilde{T}| \geq q_{1-\alpha}^{\mathcal{T}(n-1)} \right) = 1 - \left[ F_{\mathcal{T}(n-1)}(\tilde{T}(x)) - F_{\mathcal{T}(n-1)}(-\tilde{T}(x)) \right]$$

$$p_2(x) = \sup_{\mu \leq 16} \mathbb{P}_\mu \left( \tilde{T} \geq q_{1-\alpha}^{\mathcal{T}(n-1)} \right) = \mathbb{P}_{\mu=16} \left( \tilde{T} \geq q_{1-\alpha}^{\mathcal{T}(n-1)} \right) = 1 - F_{\mathcal{T}(n-1)}(\tilde{T}(x))$$

On a donc avec  $n = 8$  :

$$\begin{cases} \tilde{T}(X) = \sqrt{7} \frac{20.31-16}{3.96} = 2.88 \\ p_1(x) = 1 - (F_{\mathcal{T}(7)}(2.88) - F_{\mathcal{T}(7)}(-2.88)) = 0.024 \\ p_2(x) = 1 - F_{\mathcal{T}(7)}(2.88) = 0.012 \end{cases}$$

On a calculé les  $p$ -values avec le code ci-dessous avec *Python* :

```
1 %pylab inline
2 from scipy import stats
3
4 T_tilde_x = sqrt(7)*(20.31-16)/(3.96)
5 pv_1 = 1 - (stats.t.cdf(T_tilde_x, df = 7) - stats.t.cdf(-T_tilde_x, df = 7))
6 pv_2 = 1 - stats.t.cdf(T_tilde_x, df = 7)
7
8 print("P-value test 1 = ", pv_1)
9 print("P-value test 2 = ", pv_2)
```

Donnant la sortie :

```
1 [Out] P-value test 1 = 0.023665341010521757
2 [Out] P-value test 2 = 0.011832670505260823
```

les  $p$ -value des deux test sont donc inférieures à 5%, on rejette donc  $H_0$  dans les deux cas.

Cela semble logique, à un niveau de 5% on a rejeté  $H_0$  connaissant  $\sigma = 5$ , ici on a un  $\hat{\sigma}$  empirique plus petit, les valeurs sont donc plus recentrées autour de la moyenne empirique et donc il est encore moins probable qu'on valide l'hypothèse déjà rejetée avant.

**2.a.** On a :

$$\begin{aligned} \Delta_n &= \sup_{t \in \mathbb{R}} \left| \hat{F}_n(t) - N_{\hat{\mu}, \hat{\sigma}^2}(t) \right| \\ &= \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq t} - N_{\hat{\mu}, \hat{\sigma}^2}(t) \right| \end{aligned}$$

On pose alors :  $Y_i = \frac{X_i - \mu}{\sigma}$  ; On a donc :  $Y_i \sim \mathcal{N}(0, 1)$ .

$$\Delta_n = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq \frac{t-\mu}{\sigma}} - N_{\hat{\mu}, \hat{\sigma}^2}(t) \right|$$

On pose alors le changement de variable :  $y = \frac{t-\mu}{\sigma}$ , soit  $t = \sigma y + \mu$ . Alors :

$$\Delta_n = \sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq y} - N_{\hat{\mu}, \hat{\sigma}^2}(\sigma y + \mu) \right|$$

Or, on a que :

$$N_{\mu, \sigma^2}(t) = \mathbb{P}(X_1 \leq t) = \mathbb{P}\left(\frac{X_1 - \mu}{\sigma} \leq \frac{t - \mu}{\sigma}\right) = N_{0,1}\left(\frac{t - \mu}{\sigma}\right) = \Phi\left(\frac{t - \mu}{\sigma}\right)$$

On a donc :

$$N_{\hat{\mu}, \hat{\sigma}^2}(\sigma y + \mu) = \Phi\left(\frac{\sigma y + \mu - \hat{\mu}}{\hat{\sigma}}\right)$$

On décide alors de poser :  $\hat{\mu}_2$  l'estimateur de la moyenne de  $(Y_i)$ , et  $\hat{\sigma}_2^2$  l'estimateur biaisé de la variance de  $Y_i$ .

$$\begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n \sigma Y_i + \mu = \sigma \hat{\mu}_2 + \mu \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n (\sigma Y_i + \mu - (\sigma \hat{\mu}_2 + \mu))^2 = \sigma^2 \hat{\sigma}_2^2 \end{cases}$$

On peut donc trouver :

$$\Phi\left(\frac{\sigma y + \mu - \hat{\mu}}{\hat{\sigma}}\right) = \Phi\left(\frac{\sigma y + \mu - \sigma \hat{\mu}_2 - \mu}{\sigma \hat{\sigma}_2}\right) = \Phi\left(\frac{y - \hat{\mu}_2}{\hat{\sigma}_2}\right)$$

On revient donc à  $\Delta_n$  :

$$\Delta_n = \sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq y} - N_{\hat{\mu}, \hat{\sigma}^2}(\sigma y + \mu) \right| = \sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq y} - \Phi\left(\frac{y - \hat{\mu}_2}{\hat{\sigma}_2}\right) \right| \equiv \tilde{h}_n$$

On a donc le test :  $\Phi = \mathbb{1}_{\{\Delta_n > q_{1-\alpha}^{\tilde{h}_n}\}}$ .

**2.b.** l'écart entre la courbe rouge et les paliers est maximal pour  $t \simeq 24$ , pour ce  $t$ ,  $\Delta_n \simeq 0.8 - 0.62 = 0.18$ .

**2.c.** On a le test :  $\Phi = \mathbb{1}_{\{\Delta_n > q_{1-\alpha}^{\tilde{h}_n}\}}$ . Ici,  $\Delta_n \simeq 0.18$ , pour  $\alpha = 5\%$  on a donc  $q_{95\%} = 0.29$ . On a donc  $\Delta_n < q_{95\%}$ , on accepte donc  $H_0$ .

**2.d.**

## Exercice 2

**1.a** On a ici des données appariées (même personne avant et après la prise de sang) mais on ne possède pas les valeurs, uniquement les signes. On se propose alors d'utiliser le test du signe. On pose donc :

- $m$  médiane des  $X_i$
- $X_i = U_i - V_i$
- $U_i$  rythme cardiaque avant le don du sang
- $V_i$  rythme cardiaque après le don du sang

Nous ne savons que le signe des  $X_i$  : il est à 6 reprises positif et à 2 négatif. Il n'est apparemment jamais nul, on a donc bien  $\mathbb{P}_{H_0}(X_i = 0) = 0$ . De plus les  $X_i$  sont indépendants entre eux et ne sont pas forcément de même loi mais on suppose la médiane  $m$  commune.

On pose alors :

$$H_0 : m = 0 \quad \text{vs.} \quad H_1 : m > 0$$

La statistique de test est alors :

$$S_8 = \sum_{i=1}^8 \mathbb{1}_{X_i > 0}$$

et le test est de la forme  $\Phi_\alpha(X_i) = \mathbb{1}_{\{S_8 > q_{1-\alpha}^{B(8,1/2)}\}}$

**1.b** On a dans notre cas :  $S_8 = 6$  et  $q_{1-\alpha}^{B(8,1/2)}$  le quantile d'ordre  $1 - \alpha$  de la loi binomiale  $B(8, 1/2)$ .  
On code alors dans *Python* :

```
1 %pylab inline
2 from scipy import stats
3
4 X = array([1,1,1,1,1,1,-1,-1])
5 masque = array([X[i] > 0 for i in range(len(X))])
6 S_8 = sum(X[masque])
7 q_1_alpha = stats.binom.ppf(q = 0.95, n = 8, p = 1/2)
8 if (S_8 > q_1_alpha):
9     print("H0 rejetee")
10 else : print("H0 acceptee")
```

Nous donnant la sortie :

```
1 [Out] H0 acceptee
```

En effet,  $q_{95\%}^{B(8,1/2)} = 6$  on est donc à la limite la zone de rejet avec  $S_8 = 6$ , mais l'inégalité est stricte et le reste étant donné que la fonction de répartition est discrète.

**2.a.** On a ici des données appariées, et on a accès à leur valeurs et non uniquement à leur signe. On se propose alors d'utiliser un test de Wilcoxon des rangs signés. On pose donc à nouveau :

- $m$  médiane des  $X_i$
- $X_i = U_i - V_i$
- $U_i$  rythme cardiaque avant le don du sang
- $V_i$  rythme cardiaque après le don du sang

Mais cette fois ci, on dispose des valeurs.

On a donc ici les deux hypothèses :

$$H_0 : m = 0 \quad \text{vs.} \quad H_1 : m > 0$$

La statistique de test  $W_8^+$  est alors :

$$W_8^+ = \sum_{i=1}^8 R_{|X|}(i) \mathbb{1}_{X_i > 0}$$

Son espérance sous  $H_0$  est :

$$\mathbb{E}_{H_0}(W_8^+) = \frac{n(n+1)}{4} = \frac{72}{4} = 18$$

i	1	2	3	4	5	6	7	8
$X_i$	-6	21	10	3	13	-1	2	7
$ X_i $	6	21	10	3	13	1	2	7
$R_{ X_i }(i)$	4	8	6	3	7	1	2	5

**2.b.** On a la table suivante :

On observe alors :

$$W_8^+ = 8 + 6 + 3 + 7 + 2 + 5 = 31$$

Le test est de la forme :  $\Phi_\alpha = \mathbb{1}_{\{W_8^+ > q_{1-\alpha}\}}$  où  $q_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha$  de la loi de  $W_8^+$  sous  $H_0$ . la  $p$ -valeur observée se trouve donc par :

$$\mathbb{P}_{H_0}(W_8^+ > 31) = 1 - \mathbb{P}_{H_0}(W_8^+ \leq 30) = 1 - 0.9610 = 0.0390$$

On a alors que :  $0.0390 < 0.05$ , on ne conserve donc pas  $H_0$  au niveau 5%.

Dans un premier cas, on acceptait  $H_0$  (le rythme cardiaque n'ont pas baissés après le don de sang) sans avoir accès aux valeurs, dans le second cas on rejette  $H_0$  (le rythme cardiaque est plus faible après le don de sang). L'accès aux valeurs nous a permis de modifier notre résultat.