MAP 433 - EN 2 :

1- Déterminons l'EMV de $\theta = (\mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$

On a $P(x_1, \dots, x_n, \mu_0, \sigma_0^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x_i-\mu_0)^2}{2\sigma_0^2}}$

$$= \frac{1}{(2\pi)^{n/2}} \frac{1}{(\sigma_0^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma_0^2}}$$

$$\ln P(x_1, \dots, x_n, \mu_0, \sigma_0^2) = c - \frac{n}{2}\ln \sigma_0^2 - \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma_0^2}$$

$$\frac{\partial \ln P(X, \mu_0, \sigma_0^2)}{\partial \mu_0} = 0 \iff \frac{\sum x_i}{n} = \mu_0 \iff \mu_0 = \overline{X}_n$$

$$\frac{\partial \ln P(X, \mu_0, \sigma_0^2)}{\partial \sigma_0^2} = 0 \iff \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma_0^2} = n \iff \sigma_0^2 = \frac{\sum(x_i-\mu_0)^2}{n}$$
$$= \frac{\sum_{i=1}^{n}(x_i-\overline{X}_n)^2}{n}$$

Ainsi $\hat{\theta} = \left(\overline{X}_n, \overline{Y}_p, \frac{\sum_{i=1}^{n}(x_i-\overline{X}_n)^2}{n}, \frac{\sum_{i=1}^{p}(y_i-\overline{Y}_p)^2}{p}\right)$

2- En utilisant la méthode du RV généralisé, déterminons un test de niveau α

On prend comme statistique :

$$\Lambda(z) = \frac{\sup_{(\mu_0, \sigma_0^2) \in \mathbb{R}^+ \times \mathbb{R}^*_+} P(x_1, \dots, x_n, \mu_0, \sigma_0^2)}{\sup_{(\mu_1, \sigma_1^2) \in \mathbb{R}^+ \times \mathbb{R}^*_+} P(y_1, \dots, y_p, \mu_1, \sigma_1^2)}$$

Maximiser les numérateur et denominateur suivant μ_0, μ_1 respectivement revient à ~~xxxxxxxx~~ utiliser la statistique :

$$\Lambda'(z) = \frac{(n-1)^{-1} R_{n,0}(z)}{(p-1)^{-1} R_{p,1}(z)} \quad \text{où} \quad \begin{cases} R_{n,0}(z) = \sigma_0^{-2} \sum_{i=1}^{n}(x_i-\hat{\mu}_0)^2 \sim \chi^2_{(n-1)} \\ R_{p,1}(z) = \sigma_1^{-2} \sum_{i=1}^{p}(y_i-\hat{\mu}_1)^2 \sim \chi^2_{(p-1)} \end{cases}$$

Ainsi, $\Lambda'(z;\sigma_0^2,\sigma_1^2)$ suit une loi de Fisher de paramètre $(n-1, p-1)$.

On prend le test : $\Theta_0 = \{\sigma_1^2\}$ $\Theta_1 :]\sigma_1^2 ; +\infty[$

$$\phi_\alpha(z) = \begin{cases} 1 \text{ si } \Lambda'(z,\sigma_0^2,\sigma_1^2) > k_\alpha \\ 0 \text{ si } \Lambda'(z,\sigma_0^2,\sigma_1^2) < k_\alpha. \end{cases}$$

Ainsi, k_α est tel que $P_{\theta_0}(\phi_\alpha(z)=1) = \alpha$.

$\Rightarrow P_{\theta_0}(\Lambda'(z) > k_\alpha) = \alpha$

$\Rightarrow k_\alpha = q_{1-\alpha}^{F(n-1,p-1)}$

3) Code python pour calculer la p-valeur du test

$$\hat{\alpha}(\cancel{z}) = \inf\{\alpha \mid \phi_\alpha(z) = 1\}$$

$$= \inf\{\alpha \mid \Lambda'(z,\sigma_0^2,\sigma_1^2) > q_{1-\alpha}\}$$

$F_{n-1,p-1}^{-1}$ est décroissante car et bijective

ainsi $\hat{\alpha}(x)$ est tq $1-\hat{\alpha}(x) = F(\Lambda'(z,\sigma_0^2,\sigma_1^2))$, avec F la fonction de répartition de Fisher$(n-1,p-1)$

$$\Rightarrow \hat{\alpha}(\cancel{z}) = 1 - F(\Lambda'(z,\sigma_0^2,\sigma_1^2)) . \forall \theta \text{ tq } \sigma_0^2 = \sigma_1^2$$

$$= P_{\theta_0}(\Lambda'(z,\sigma_0^2,\sigma_1^2) < F_{n-1,p-1})$$

$$= P\left(\frac{\sum_{i=1}^{n}(x_i-\hat{\mu}_0)^2}{\sum_{i=1}^{p}(y_i-\hat{\mu}_1)^2} < F_{n-1,p-1}\right)$$

$$\hat{\alpha}(z) = 1 - F\left(\frac{p-1}{n-1}\cdot\frac{\sum(x_i-\hat{\mu}_0)^2}{\sum_{i=1}^{n}(y_i-\hat{\mu}_1)^2}\right)$$

4) Construisons un intervalle de confiance pour $\mu_0 - \mu_1$ de niveau de confiance $1-\alpha$

On a : $\hat{\mu}_0 = \frac{\sum_{i=1}^{n} x_i}{n}$ et $\hat{\mu}_1 = \frac{\sum_{i=1}^{p} y_i}{p}$

$\sim \mathcal{N}(\mu_0 ; \frac{\sigma^2}{n})$ $\sim \mathcal{N}(\mu_1, \frac{\sigma^2}{p})$

Comme $\hat{\mu}_0 \perp\!\!\!\perp \hat{\mu}_1$; on a : $\hat{\mu}_0 - \hat{\mu}_1 \sim \mathcal{N}(\mu_0 - \mu_1, \sigma^2(\frac{1}{n} + \frac{1}{p}))$

$$\Rightarrow \sqrt{\frac{np}{n+p}} \left\{ \frac{(\hat{\mu}_0 - \hat{\mu}_1) - (\mu_0 - \mu_1)}{\sigma} \right\} \sim \mathcal{N}(0,1)$$

On remplace σ^2 par la variance empirique car σ^2 inconnue :

$$S^2_{n,p} = \frac{1}{n+p-2} \left(\sum_{i=1}^{n} (x_i - \hat{\mu}_0)^2 + \sum_{j=1}^{p} (y_j - \hat{\mu}_1)^2 \right) = \frac{n+p}{n+p-2} \hat{\sigma}^2_{n,p}$$

Comme fait à la PC3 - Exercice 5 ; on trouve que :

$$\left\{ (\mu_0 - \mu_1) : t^{\alpha/2}_{n+p-2} \leq \sqrt{\frac{np}{n+p}} \left\{ \frac{(\hat{\mu}_0 - \hat{\mu}_1) - (\mu_0 - \mu_1)}{S_{n,p}} \right\} \leq t^{1-\alpha/2}_{n+p-2} \right\}$$

est un intervalle de confiance de niveau $1-\alpha$.

$$\Rightarrow \boxed{I(Z) = \left[\hat{\mu}_0 - \hat{\mu}_1 - t^{1-\alpha/2}_{n+p-2} \sqrt{\frac{n+p}{np}} S_{n,p}(Z) ; \hat{\mu}_0 - \hat{\mu}_1 + t^{1-\alpha/2}_{n+p-2} \sqrt{\frac{n+p}{np}} S_{n,p}(Z) \right]}$$

Car $t^{\alpha/2}_{n+p-2} = - t^{1-\alpha/2}_{n+p-2}$.

Avec $t^{\alpha/2}_{n+p-2}$: quantile d'ordre $\frac{\alpha}{2}$ de la loi de Student de paramètre $n+p-2$.

5- Test de niveau $\alpha = 0{,}01$ de l'hypothèse

$\mu_0 = \mu_1$ contre $\mu_0 \neq \mu_1$

On ~~a~~ considère : $\Theta_0 = \{\mu_1\}$ et $\Theta_1 = \mathbb{R} \setminus \{\mu_1\}$

On considère la statistique :

$$r(Z) = \sqrt{\frac{np}{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1 - (\mu_0 - \mu_1)|}{S_{n,p}(Z)}$$

On prend le test : $\phi_\alpha(z) = \mathbb{1}\{r(z) > k_\alpha\}$

avec k_α tel que : $P_{\Theta_0}(\phi_\alpha(Z) = 1) = \alpha$.

$$\Rightarrow P\left(\sqrt{\frac{np}{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1|}{S_{n,p}(Z)} > k_\alpha \right) = \alpha \Rightarrow p(|r(z)| > k_\alpha) = \alpha$$

$\cancel{\Rightarrow 1 - p(|r(z)| \leq k_\alpha) = 1 - \alpha}$

~~On prend k_α =~~

$$\Rightarrow 1 - p(|r(z)| \leq k_\alpha) = \alpha$$

$$\Rightarrow 1 - p(-k_\alpha \leq r(z) \leq k_\alpha) = \alpha$$

$$\Rightarrow 1 - F(k_\alpha) + F(-k_\alpha) = \alpha$$ avec F : fct de répartition de ~~S_{n+p}~~ St_{n+p-2}

$$\Rightarrow 1 - F(k_\alpha) + 1 - F(k_\alpha) = \alpha$$ car F paire

$$\rightarrow F(k_\alpha) = 1 - \frac{\alpha}{2}$$

$$k_\alpha = t_{n+p-2}^{1-\alpha/2}$$

Ainsi, le test est défini par la zone de rejet : $R_\alpha = \left\{ Z \in \mathbb{R}^{n+p} ; \sqrt{\frac{np}{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1|}{S_{n,p}(Z)} \geq t_{n+p-2}^{1-\alpha/2} \right\}$

6 - pValeur du test :

$$\hat{\alpha}(\mathcal{Z}) = \inf \{ \alpha \mid \Phi_\alpha(z) = 1 \}$$

$$= \inf \left\{ \alpha \mid \sqrt{\frac{np}{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1 - (\mu_0 - \mu_1)|}{S_{n,p}(\mathcal{Z})} > t_{n+p-2}^{1-\alpha/2} \right\}$$

$\forall \theta \in \Theta_0$; $\hat{\alpha}(\mathcal{Z}) = \inf \left\{ \alpha / \sqrt{\frac{np}{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1|}{S_{n,p}} > t_{n+p-2}^{1-\alpha/2} \right\}$

$$\Rightarrow 1 - \frac{\hat{\alpha}(\mathcal{Z})}{2} = F\left(\frac{\sqrt{np}}{\sqrt{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1|}{S_{n,p}} \right)$$ avec F fct de répartition de St_{n+p-2}

$$\Rightarrow \boxed{\hat{\alpha}(\mathcal{Z}) = 2\left(1 - F\left(\sqrt{\frac{np}{n+p}} \frac{|\hat{\mu}_0 - \hat{\mu}_1|}{S_{n,p}} \right)\right)}$$

7 - Test de niveau $\alpha = 0{,}01$ avec

$H_0 : \mu_0 \leq \mu_1$ contre $\mu_0 > \mu_1$

On considère : - $\Theta_0 :]-\infty ; \mu_1]$ et $\Theta_1 :]\mu_1 ; +\infty[$

- La statistique $\sqrt{\frac{np}{n+p}} \frac{[\hat{\mu}_0 - \hat{\mu}_1 - (\mu_0 - \mu_1)]}{S_{n,p}(z)} = r(z)$

- Le test $\Phi_\alpha(\mathcal{Z}) = \mathbb{1}\{ r(z) > k_\alpha \}$ avec k_α tel que $P_{\Theta_0}(\Phi_\alpha(\mathcal{Z}) = 1) = \alpha$

$$\Rightarrow P_{\Theta_0}\left(\sqrt{\frac{np}{n+p}} \frac{[\hat{\mu}_0 - \hat{\mu}_1 - (\mu_0 + \mu_1)]}{S_{n,p}(\mathcal{Z})} > k_\alpha \right) = \alpha$$

Or $P_{\Theta_0}(r(z) > k_\alpha) = P_{\Theta_0}\left(\sqrt{\frac{np}{n+p}} \frac{\hat{\mu}_0 - \hat{\mu}_1}{S_{n,p}(\mathcal{Z})} > k_\alpha + (\mu_0 - \mu_1) \right)$

$\geq P\left(\sqrt{\frac{np}{n+p}} \frac{\hat{\mu}_0 - \hat{\mu}_1}{S_{n,p}(\mathcal{Z})} > k_\alpha \right)$ car $\mu_0 - \mu_1 \leq 0$ sous Θ_0

$\Rightarrow k_\alpha = t_{n+p-2}^{1-\alpha}$ avec t_{n+p-2}^{α} : quantile d'ordre α de la loi de Student de paramètre $n+p-2$

ainsi
$$\boxed{\begin{array}{l} \sup_{\theta;\, \mu_0-\mu_1 \leq 0} P_\theta(z \in R_\alpha) = \alpha \\ \text{avec } R_\alpha = \left\{ z \in \mathbb{R}^{n+p} ;\ \sqrt{\frac{np}{n+p}} \left(\frac{\hat{\mu}_0 - \hat{\mu}_1}{S_{n,p}} \right) \geq t_{n+p-2}^{1-\alpha} \right\} \end{array}}$$

8 - Code python pour la p-valeur du test

$$\hat{\alpha}(Z) = \inf \{ \alpha \mid \phi_\alpha(z) = 1 \}$$

$$= \inf \{ \alpha \mid r(z) \geq t_{n+p-2}^{1-\alpha} \}$$

$$\forall \theta \in \Theta_0 ;\ \hat{\alpha}(Z) = \inf \left\{ \sqrt{\frac{np}{n+p}} \left(\frac{\hat{\mu}_0 - \hat{\mu}_1}{S_{n,p}} \right) \geq t_{n+p-2}^{1-\alpha} \right\}$$

$$\Rightarrow 1 - \hat{\alpha}(Z) = F\left(\sqrt{\frac{np}{n+p}} \left(\frac{\hat{\mu}_0 - \hat{\mu}_1}{S_{n,p}} \right) \right)$$ avec F : fct° de repartition d'une student St_{n+p-2}

$$\Rightarrow \hat{\alpha}(Z) = 1 - F\left(\sqrt{\frac{np}{n+p}} \left(\frac{\hat{\mu}_0 - \hat{\mu}_1}{S_{n,p}} \right) \right)$$