# TD n° 4 : Régression linéaire multiple 2

**Exercice 1.** Soient $Y_{-n}, \ldots, Y_n$, $2n+1$ *var* indépendantes telles que, pour tout $i \in \{-n, \ldots, n\}$,

$$Y_i = ax_i + b\cos(x_i) + \epsilon_i,$$

où $x_i = \dfrac{i\pi}{n}$ et $\epsilon_{-n}, \ldots, \epsilon_n$ sont $2n+1$ *var iid* suivant chacune la loi normale $\mathcal{N}(0, \sigma^2)$. Les paramètres $a$, $b$ et $\sigma$ sont des réels inconnus.

1. *Question préliminaire.* Montrer que $\displaystyle\sum_{i=-n}^{n} x_i^2 = \frac{\pi^2(n+1)(2n+1)}{3n}$ et $\displaystyle\sum_{i=-n}^{n} x_i \cos(x_i) = 0$. En admettant que $\displaystyle\sum_{i=1}^{n} \cos(2x_i) = 0$, montrer que $\displaystyle\sum_{i=-n}^{n} \cos^2(x_i) = n+1$.
2. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, en indiquant ce que sont ici $Y$, $X$, $\beta$ et $\epsilon$.
3. Calculer l'*emco* $\widehat{a}$ de $a$ et l'*emco* $\widehat{b}$ de $b$.
4. Déterminer la loi de $\widehat{a}$ et la loi de $\widehat{b}$.
5. Donner un estimateur $\widehat{\sigma}^2$ sans biais de $\sigma^2$.
6. On considère les hypothèses :

$$H_0 : a \leq 0 \qquad \text{contre} \qquad H_1 : a > 0.$$

   Quelle démarche doit-on suivre pour décider du rejet de $H_0$ (ou non) au risque 1% ?

**Exercice 2.** Soient $Z_1, \ldots, Z_n$, $n$ *var* indépendantes telles que, pour tout $i \in \{1, \ldots, n\}$,

$$Z_i = ai^{\frac{1}{2}} + bi^{\frac{3}{2}} + \xi_i,$$

où $\xi_1, \ldots, \xi_n$ sont $n$ *var iid* suivant chacune la loi normale $\mathcal{N}(0, \theta^2)$. Les paramètres $a$, $b$ et $\theta$ sont des réels inconnus.

1. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, en indiquant ce que sont ici $Y$, $X$, $\beta$ et $\epsilon$.
2. Calculer les *emco* de $a$ et $b$. Déterminer la loi de ces estimateurs.
3. Donner un estimateur $\widehat{\theta}^2$ sans biais de $\theta^2$. Déterminer un réel $c$ inconnu tel que la *var* $c\widehat{\theta}^2$ suive une loi usuelle.
4. Déterminer un intervalle de confiance (aléatoire) pour $a$ au niveau 90%.
5. Déterminer un intervalle de confiance (aléatoire) pour $\theta^2$ au niveau 90% de la forme $]0, U]$, où $U$ est une *var* que l'on explicitera.

**Exercice 3.** Soit $Z$ le *vectar* de $\mathbb{R}^n$ défini par

$$Z = T\beta + \delta,$$

où $T$ est une matrice à $n$ lignes et $p$ colonnes connue et $\delta$ est un *vectar* suivant la loi normale multivariée $\mathcal{N}_n(0_n, \sigma^2\Omega)$, où $\Omega$ est une matrice symétrique définie positive connue. Ici, $\beta$ est un vecteur inconnu de $\mathbb{R}^p$ et $\sigma$ est un réel inconnu.

1. Montrer qu'il existe une matrice carrée $U$ telle que si $Y = UZ$, alors

   $$Y = X\beta + \epsilon,$$

   où $X$ est une matrice à $n$ lignes et $p$ colonnes connue que l'on précisera et $\epsilon$ suit la loi normale multivariée $\mathcal{N}_n(0_n, \sigma^2\mathbb{I}_n)$ (*on utilisera le resultat : si $A$ est une matrice carrée symétrique définie positive, alors il existe une matrice orthogonale $Q$ et une matrice diagonale $S$ à coefficients diagonaux strictement positifs telles que $A = QS^2Q^t$).*

   Exprimer $U^tU$ en fonction de $\Omega$.

2. Montrer que l'*emco* $\widehat{\beta}$ de $\beta$ peut s'écrire en fonction de $Z$, $T$ et $\Omega$. Déterminer la loi de $\widehat{\beta}$ en fonction de ces éléments.

3. Donner un estimateur sans biais de $\sigma^2$ en fonction de $Z$, $T$ et $\Omega$.

**Exercice 4.** On observe $n$ valeurs d'un *car* $(Z, X)$ notées $(z_1, x_1), \ldots, (z_n, x_n)$ telles que, pour tout $i \in \{1, \ldots, n\}$, $z_i$ est une réalisation de

$$Z_i = a + bx_i + cx_i^2 + \epsilon_i,$$

où $\epsilon_1, \ldots, \epsilon_n$ sont $n$ *var* indépendantes suivant la loi normale centrée avec, pour tout $i \in \{1, \ldots, n\}$, $\mathbb{V}(\epsilon_i) = \sigma^2\delta_i^2$, $\delta_1, \ldots, \delta_n$ sont des réels fixés non nuls. Les paramètres $a$, $b$, $c$ et $\sigma$ sont des réels inconnus. On suppose que $x_1, \ldots, x_n$ et $\delta_1, \ldots, \delta_n$ vérifient

$$\sum_{i=1}^{n} \frac{x_i}{\delta_i^2} = 0, \qquad \sum_{i=1}^{n} \frac{x_i^3}{\delta_i^2} = 0.$$

On pose

$$u_n = \sum_{i=1}^{n} \frac{1}{\delta_i^2}, \qquad v_n = \sum_{i=1}^{n} \frac{x_i^2}{\delta_i^2}, \qquad w_n = \sum_{i=1}^{n} \frac{x_i^4}{\delta_i^2}.$$

1. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, où $\epsilon$ suit la loi normale multivariée $\mathcal{N}_n(0_n, \sigma^2\mathbb{I}_n)$, en indiquant ce que sont ici $Y$, $X$, $\beta$ et $\epsilon$.

2. Montrer que la matrice $X^tX$ est inversible si, et seulement si, $v_n \neq 0$ et $u_nw_n \neq v_n^2$. Sous ces conditions, calculer $(X^tX)^{-1}$ en fonction de $u_n$, $v_n$ et $w_n$.

On suppose désormais que $v_n \neq 0$ et $u_nw_n \neq v_n^2$.

3. Calculer l'*emco* $\widehat{a}$ de $a$, l'*emco* $\widehat{b}$ de $b$ et l'*emco* $\widehat{c}$ de $c$, en fonction de $u_n$, $v_n$, $w_n$, des composantes du vecteur $Y$, de $\delta_1, \ldots, \delta_n$ et de $x_1, \ldots, x_n$.

4. On suppose que $n = 13$. Déterminer un intervalle de confiance (aléatoire) pour $\sigma^2$ au niveau $90\%$ de la forme $]0, U]$, où $U$ est une *var* que l'on explicitera.