TD n^o 5 : Régression linéaire multiple 3

Exercice 1. Soient $(k, T) \in (\mathbb{N}^*)^2$, $n = kT$ et $Z_1, \ldots, Z_n$ n *var* indépendantes telles que, pour tout $i \in \{1, \ldots, n\}$,

$$Z_i = \beta_i + \xi_i,$$

où $\xi_1, \ldots, \xi_n$ sont n *var iid* suivant chacune la loi normale $\mathcal{N}(0, \sigma^2)$. On se place dans le cas T-périodique : pour tout $(u, v) \in \{0, \ldots k-1\} \times \{1, \ldots, T\}$,

$$\beta_{uT+v} = \beta_v.$$

Les paramètres $\beta_1, \ldots, \beta_T$ sont des réels inconnus. On pose $\beta = (\beta_1, \ldots, \beta_T)^t$.

1. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, en indiquant ce que sont ici Y, X et ϵ.
2. Calculer l'*emco* $\widehat{\beta}$ de β, puis donner sa loi.
3. Donner un estimateur sans biais $\widehat{\sigma}^2$ de σ^2. Exprimer $\widehat{\sigma}^2$ sous la forme d'une somme double.

Exercice 2. Soient $\ell \in \mathbb{N}^*$, $n \in \mathbb{N}^*$ tel que $n > 2\ell + 1$, et $Z_1, \ldots, Z_n$ n *var* indépendantes telles que, pour tout $i \in \{1, \ldots, n\}$,

$$Z_i = \alpha_0 + \sum_{v=1}^{\ell} (\alpha_v \cos(x_{v,i}) + \theta_v \sin(x_{v,i})) + \xi_i,$$

où $x_{v,i} = \dfrac{2vi\pi}{n}$ et $\xi_1, \ldots, \xi_n$ sont n *var iid* suivant chacune la loi normale $\mathcal{N}(0, \sigma^2)$. Les paramètres $\alpha_0, \alpha_1, \theta_1, \ldots, \alpha_\ell, \theta_\ell$ et σ sont des réels inconnus. On pose $\beta = (\alpha_0, \alpha_1, \theta_1, \ldots, \alpha_\ell, \theta_\ell)^t$.

1. Écrire le modèle linéaire associé sous la forme matricielle usuelle : $Y = X\beta + \epsilon$, en indiquant ce que sont ici Y, X et ϵ.
2. On admet que, pour tout $(u, v) \in (\mathbb{N}^*)^2$,

$$\sum_{i=1}^{n} \cos(x_{v,i}) = 0, \qquad \sum_{i=1}^{n} \sin(x_{v,i}) = 0, \qquad \sum_{i=1}^{n} \cos(x_{v,i}) \sin(x_{u,i}) = 0,$$

$$\sum_{i=1}^{n} \cos(x_{v,i}) \cos(x_{u,i}) = \begin{cases} \dfrac{n}{2} & \text{si } u = v, \\ 0 & \text{sinon,} \end{cases}, \qquad \sum_{i=1}^{n} \sin(x_{v,i}) \sin(x_{u,i}) = \begin{cases} \dfrac{n}{2} & \text{si } u = v, \\ 0 & \text{sinon.} \end{cases}$$

   Calculer l'*emco* $\widehat{\alpha}_0$ de α_0 et, pour tout $v \in \{1, \ldots, \ell\}$, l'*emco* $\widehat{\alpha}_v$ de α_v et l'*emco* $\widehat{\theta}_v$ de θ_v.
3. Donner un estimateur sans biais $\widehat{\sigma}^2$ de σ^2. Exprimer $\widehat{\sigma}^2$ sous la forme d'une somme faisant intervenir $(Z_i)_{i \in \{1,\ldots,n\}}$, $(x_{v,i})_{(v,i) \in \{1,\ldots,\ell\} \times \{1,\ldots,n\}}$, $\widehat{\alpha}_0$, $(\widehat{\alpha}_v)_{v \in \{1,\ldots,\ell\}}$ et $(\widehat{\theta}_v)_{v \in \{1,\ldots,\ell\}}$.

Exercice 3. On considère le modèle :

$$Y = U\alpha + V\gamma + \epsilon,$$

où U est une matrice à n lignes et p_1 colonnes connue, V est une matrice à n lignes et p_2 colonnes connue, et ϵ suit la loi normale multivariée $\mathcal{N}_n(0_n, \mathbb{I}_n)$. Ici, $\alpha = (\alpha_1, \ldots, \alpha_{p_1})^t$ est un vecteur inconnu de $\mathbb{R}^{p_1}$ et $\gamma = (\gamma_1, \ldots, \gamma_{p_2})^t$ est un vecteur inconnu de $\mathbb{R}^{p_2}$. On pose

$$P = U(U^tU)^{-1}U^t, \qquad M = \mathbb{I}_n - P.$$

Dans ce que suit, $\|.\|$ et $< ., . >$ désignent respectivement la norme et le produit scalaire euclidiens dans $\mathbb{R}^n$.

1. *Questions préliminaires.*
   (a) Montrer que $PU = U$, $MU = 0$, $P^t = P$, $P^2 = P$, $M^t = M$, $M^2 = M$ et $M^tP = 0$.
   (b) Montrer que, pour tout vecteur A à n composantes, on a
   $$< MA, PA >= 0, \qquad \|A\|^2 = \|MA\|^2 + \|PA\|^2.$$
2. En utilisant les résultats de la question 1, montrer que, pour tout $(u, v) \in \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$,
   $$\|Y - Uu - Vv\|^2 = \|MY - MVv\|^2 + \|PY - Uu - PVv\|^2.$$
3. On définit
   $$\begin{cases} \widehat{\gamma} = \underset{v \in \mathbb{R}^{p_2}}{\operatorname{argmin}} \|MY - MVv\|^2, \\ \\ \widehat{\alpha} \text{ tel que } PY - U\widehat{\alpha} - PV\widehat{\gamma} = 0. \end{cases}$$
   On suppose que $\widehat{\gamma}$ et $\widehat{\alpha}$ existent et sont uniques.
   Montrer que, pour tout $(u, v) \in \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$, on a
   $$\|MY - MVv\|^2 + \|PY - Uu - PVv\|^2 \geq \|MY - MV\widehat{\gamma}\|^2 + \|PY - U\widehat{\alpha} - PV\widehat{\gamma}\|^2.$$
   En déduire que
   $$(\widehat{\alpha}, \widehat{\gamma}) = \underset{(u,v) \in \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}}{\operatorname{argmin}} \|Y - Uu - Vv\|^2.$$
4. Montrer que
   $$MY = MV\gamma + \epsilon^*,$$
   où ϵ^* suit la loi normale multivariée $\mathcal{N}_n(0_n, M)$.
5. On suppose désormais que M est définie positive. En utilisant le résultat de l'exercice 3 de la feuille de TD 4, montrer que
   $$\widehat{\gamma} = (V^tMV)^{-1}V^tMY.$$