TD DE MODÈLES LINÉAIRES I - SÉRIE 7

Exercice 1. Nous estimons le paramètre vectoriel $\theta_1 \in \mathbb{R}^{p_1}$ du modèle de régression linéaire approché :

$$Y = W\theta_1 + U$$

où $\mathbb{E}_\theta(U) = 0_n$ et $Var_\theta(U) = \sigma^2 \mathrm{I}_n$, alors que le vrai modèle est

$$Y = W\theta_1 + Z\theta_2 + V$$

où $\mathbb{E}_\theta(V) = 0_n$, $Var_\theta(V) = \sigma^2 \mathrm{I}_n$, $\theta_2 \in \mathbb{R}^{p_2}$, et $X = (W, Z) \in \mathcal{M}(n, p)$ avec $p = p_1 + p_2$. Nous supposons que la matrice X est de rang p.

1) Calculer $\mathbb{E}_\theta(\tilde{\theta}_1)$ où $\tilde{\theta}_1$ est l'estimateur MCO du paramètre θ_1 du modèle approché.

Nous pouvons considérer que l'estimation dans le modèle approché revient à estimer le vrai modèle sous contraintes.

2) Ecrire le modèle sous contraintes.

3) Nous étudions ici le cas d'un modèle $Y = X\theta + \xi$ sous la contrainte $G\theta = b$ avec $\mathbb{E}_\theta(\xi) = 0_n$ et $Var_\theta(\xi) = \sigma^2 \mathrm{I}_n$, $G \in \mathcal{M}(r, p)$ et G de rang r.

a) Notons $\hat{\theta}$ l'estimateur MCO de θ sans considérer la contrainte et $\hat{\theta}^0$ l'estimateur MCO de θ en considérant la contrainte. Montrer que :

$$\hat{\theta}^0 = \hat{\theta} + (X'X)^{-1}G'[G(X'X)^{-1}G']^{-1}(b - G\hat{\theta}).$$

b) Calculer $\mathbb{E}_\theta(\hat{\theta}^0)$.

c) Montrer que

$$Var_\theta(\hat{\theta}^0) = \sigma^2 \left[(X'X)^{-1} - (X'X)^{-1}G'[G(X'X)^{-1}G']^{-1}G(X'X)^{-1}\right].$$

d) Comparer $Var_\theta(\hat{\theta}^0)$ et $Var_\theta(\hat{\theta})$.

4) Interpréter le résultat précédent dans le cas d'omission de variables..

Exercice 2 Considérons le modèle

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, 1 \le i \le I, 1 \le j \le n_i,$$

où les ε_{ij} sont indépendantes de loi $\mathcal{N}(0, \sigma^2)$ et $\sum_{i=1}^{I} n_i\alpha_i = 0$.

1) Montrer que $X = A\theta + \varepsilon$, où $A = [1A_1...A_I]$ avec $1, A_1, ..., A_I$, éléments de $\mathbb{R}^{IJ}$ que l'on précisera.

2) Soit $F_0 = \{\mu 1, \mu \in \mathbb{R}\}, F_1 = \Big\{ \sum_{i=1}^{I} \alpha_i A_i, \alpha_i \in \mathbb{R}, \sum_{i=1}^{I} n_i\alpha_i = 0\Big\}$.

Montrer que F_0 et F_1 sont des sous-espaces vectoriels deux-à-deux orthogonaux. En déduire qu'il existe un espace G tel que

$$\mathbb{R}^{IJ} = F_0 \oplus F_1 \oplus G.$$

3) Calculer $P_0, P_0 + P_1$ (P_0 étant la projection orthogonale sur F_0). En déduire P_1. Calculer P_G.

4) Construire la table d'analyse de la variance.

5) Tester l'hypothèse H_0: "tous les α_i sont nuls" contre H_1: "tous les α_i ne sont pas nuls".

Application. Les mesures de teneur en octane sur des échantillons de carburant prélevés dans quatre régions du nord-est des États Unis durant l'été 1953, sont reproduites dans le tableau ci-après.

Notant X_{ij} la jème mesure effectuée dans la région i, on donne les quantités suivantes :

Région	A	B	C	D
n_i	16	13	18	22
$\bar{X}_{i.}$	83.875	82.846	83.22	83.009
$\sum_j (X_{ij} - \bar{X}_{i.})^2$	17.81	2.67	28.97	33.04

Peut-on conclure que la teneur en octane est différente suivant les régions?

Région A	Région B	Région C	Région D
84.0	82.4	83.2	80.2
83.5	82.4	82.8	82.9
84.0	83.4	83.4	84.6
85.0	83.3	80.2	84.2
83.1	83.1	82.7	82.8
83.5	83.3	83.0	83.0
81.7	82.4	85.0	82.9
85.4	83.3	83.0	83.4
84.1	82.6	85.0	83.1
83.0	82.0	83.7	83.5
85.8	83.2	83.6	83.6
84.0	83.1	83.3	86.7
84.2	82.5	83.8	82.6
82.2		85.1	82.4
83.6		83.1	83.4
84.9		84.2	82.7
		80.6	82.9
		82.3	83.7
			81.5
			81.9
			81.7
			82.5

Data from O.C. Blade "National motor-gasoline survey" Bureau of Mines Report of Investigation 5041.

Exercice 3. On considère une expérience dont le résultat est un tableau $\{X_{ij}, 1 \leq i, j \leq 2\}$ composé de 4 variables indépendantes de même variance. On adopte le modèle $X_{ij} \sim \mathcal{N}(\mu + \alpha_i - \alpha_j, \sigma^2), 1 \leq i, j \leq 2$, où $\alpha_1 + \alpha_2 = 0$.

1) Déterminer les ESBVM de μ, α_1 et α_2.

2) Tester l'hypothèse $\mu = 0$ au seuil α.

3) A.N. $\alpha = 0.05, X_{11} = 4, X_{12} = 1, X_{21} = -3, X_{22} = 2$.