RÉGRESSION LINÉAIRE MULTIPLE

1 Hémoglobine.

On souhaite étudier la variation du taux d'hémoglobine dans le sang au cours d'une opération chirurgicale en fonction de la durée de l'opération et du volume de sang perdu pendant l'opération. On dispose des résultats suivants où y_i représente la valeur observée en pourcentage de la variation du taux d'hémoglobine, x_{i1} est la durée de l'opération en heures décimales et x_{i2} est le volume en litres de sang perdu.

y_i	-1.70	-4.61	-5.82	-1.17	-4.23	-3.31	+0.42	-2.98
x_{i1}	1.75	1.33	1.43	1.86	1.81	1.66	1.60	2.00
x_{i2}	0.52	0.59	0.61	0.50	0.54	0.49	0.27	0.47

On suppose que y_i est une réalisation d'une variable aléatoire Y_i de loi $\mathcal{N}(a+bx_{i1}+cx_{i2}, \sigma^2)$.

1) Estimer les paramètres inconnus a, b, c et σ^2.

2) Tester l'hypothèse suivant laquelle la variation du taux d'hémoglobine ne dépend ni de la durée de l'opération ni du volume de sang perdu.

3) Tester l'hypothèse suivant laquelle la variation du taux d'hémoglobine ne dépend pas de la durée de l'opération.

2 Orthogonalité.

On considère la régression linéaire multiple

$$Y_i = a + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \varepsilon_i, \qquad i = 1, 2, \cdots, n$$

où (ε_i) est une suite de variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$. On lui associe les résultats matriciels suivants: $\| Y \|^2 = 640$,

$$X^t X = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 20 & 0 & -4 \\ 0 & 0 & 5 & 2 \\ 0 & -4 & 2 & 1 \end{pmatrix} \qquad \text{et} \qquad X^t Y = \begin{pmatrix} 25 \\ -20 \\ 40 \\ 20 \end{pmatrix}.$$

1) Quel est le nombre d'observations n? Interpréter les 0 de la matrice X^tX.

2) Trouver les lois des estimateurs du maximum de vraisemblance $\widehat{a}$, $\widehat{b}$ et $\widehat{\sigma^2}$ de a, b et σ^2 puis calculer les résultats numériques associés à ces estimateurs.

3) Tester l'hypothèse $b_3 = 0$ contre $b_3 > 0$ puis l'hypothèse $4a+b_1 = 2$ et $2b_2-5b_3 = 14$ contre $4a + b_1 \neq 2$ ou $2b_2 - 5b_3 \neq 14$.

3 Régression polynômiale.

On considère le modèle de régression polynômiale défini, pour $n \geq 3$, par

$$Y_i = a + bx_i + cx_i^2 + \varepsilon_i, \qquad i = 1, 2, \cdots, n$$

où (ε_i) est une suite de variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$.

1) Trouver une condition simple pour que cette régression soit identifiable.

2) Tester l'hypothèse $c = 0$ contre $c \neq 0$ pour les observations suivantes

x_i	-5	-4	-3	-2	-1	0	1	2	3	4	5
y_i	-3.37	-2.11	-2.24	1.59	3.28	3.96	6.42	8.57	10.71	14.32	15.91

4 Régression périodique.

On considère le modèle de régression périodique défini, pour $n \geq 1$, par

$$Y_i = ax_i + b\cos(x_i) + \varepsilon_i, \qquad i = -n, \cdots, n$$

où (ε_i) est une suite de variables aléatoires centrées, non corrélées avec $Var(\varepsilon_i) = \sigma^2$ et (x_i) est la suite de nombres réels donnée par $x_i = i\pi/n$.

1) Vérifier que

$$\sum_{i=-n}^{n} \cos^2(x_i) = n + 1 \qquad \text{et} \qquad \sum_{i=-n}^{n} x_i^2 = \frac{\pi^2(n+1)(2n+1)}{3n}.$$

2) Déterminer les estimateurs des moindres carrés $\widehat{a}$ et $\widehat{b}$ de a et b.

2) Montrer qu'ils sont sans biais et calculer leur variance.

3) Proposer un estimateur sans biais $\widehat{\sigma^2}$ de σ^2.

4) Dans le cadre gaussien, trouver les lois de $\widehat{a}$, $\widehat{b}$ et $\widehat{\sigma^2}$.

5) Construire un test de l'hypothèse $a = 0$ contre $a > 0$ puis de $a = 4b$ contre $a \neq 4b$.