UPS TOULOUSE III **MODÉLISATION STATISTIQUE**
MAGISTÈRE 2002

RÉGRESSION LINÉAIRE SIMPLE

1 Capitole.

Un particulier cherche à acquérir un appartement aux alentours immédiats de la place du Capitole. Il a sélectionné les 24 offres de vente suivantes où X représente la surface en mètres carrés et Y correspond au prix en milliers d'Euros

X	28	50	196	55	190	110	60	48	90	35	86	65
Y	130	280	800	268	790	500	320	250	378	250	350	300

X	32	52	40	70	28	30	105	52	80	60	20	100
Y	155	245	200	325	85	78	375	200	270	295	85	495

1) Déterminer l'équation de la droite de régression linéaire de Y en X.

2) Proposer un intervalle de confiance à 95% pour la variance résiduelle σ^2.

3) Tester la significativité de X.

4) Pour une nouvelle valeur $x^* = 100$, calculer la prédiction naturelle $\widehat{Y}^*$ de Y^* et trouver un intervalle de prévision à 95% pour Y^*.

2 Combinaison Linéaire.

On considère la régression linéaire simple identifiable

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

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

1) Pour $a = 0$, donner l'estimateur des moindres carrés $\widehat{b}$ de b et déterminer sa loi.

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

3) Soient $\alpha, \beta \in \mathbb{R}$ et $c \in \mathbb{R}$ donnés. Construire un test de l'hypothèse $\alpha a + \beta b = c$ contre $\alpha a + \beta b \neq c$.

3 Réduction de variance.

On considère la régression linéaire simple identifiable

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

On va vérifier qu'ajouter une observation (Y_{n+1}, x_{n+1}) à la régression permet d'améliorer l'estimation des paramètres a et b. On pose

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{et} \qquad \overline{x_n^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

1) Si $v_n = \overline{x_n^2} - (\overline{x}_n)^2$, montrer la relation récursive

$$v_{n+1} = \frac{n}{(n+1)} v_n + \frac{n}{(n+1)^2}(x_{n+1} - \overline{x}_n)^2.$$

2) En déduire que $Var(\widehat{a}_{n+1}) \leq Var(\widehat{a}_n)$ et $Var(\widehat{b}_{n+1}) \leq Var(\widehat{b}_n)$ puis conclure à l'amélioration de l'estimation des paramètres a et b.

4 Moindres Carrés Pondérés.

On considère la régression linéaire simple identifiable

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

où (ε_i) est une suite de variables aléatoires centrées et non-corrélées avec $Var(\varepsilon_i) = \sigma^2 z_i^2$ où (z_i) est strictement positive connue.

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

2) Montrer qu'ils sont des estimateurs sans biais et de variance minimale.

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

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

5 Durbin-Watson.

On considère la régression linéaire gaussienne identifiable définie, pour $i = 1, 2, \cdots, n$, par

$$\begin{cases} Y_i = a + bx_i + \varepsilon_i \\ \varepsilon_i = \rho\varepsilon_{i-1} + e_i \end{cases}$$

où $|\rho| < 1$ et (e_i) est une suite de variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$. Proposer une statistique permettant de tester la non-corrélation des résidus.