UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS
L3 MASS - Statistiques - 2008-2009

TD5 : Régression linéaire

Ex. 1 On reprend les données de l'exercice 4 de la feuille 4.

X	Y
12,5	0,15
17,5	0,16
22,5	0,21
27,5	0,26
32,5	0,35
37,5	0,57
42,5	0,57
47,5	1,21
52,5	1,69
57,5	2,33
62,5	3,51
67,5	5,72
72,5	9,36
77,5	17,68
82,5	30,85
87,5	40,09

Calculer le coefficient de détermination R^2 associé à ces valeurs. Que peut-on en déduire ?

Ex. 2 Montrer que les estimateurs $\widehat{\beta}_1$ et $\widehat{\beta}_2$ sont sans biais.

Ex. 3 Calculer la variance de $\widehat{\beta}_2$.

Ex. 4 Calculer la convariance de $\widehat{\beta}_1$ et $\widehat{\beta}_2$.

Ex. 5 Les arbres. Nous souhaitons exprimer la hauteur d'un arbre d'une essence donnée en fonction de son diamètre x à 1m30 du sol. Pour ce faire, nous avons mesuré 20 couples « diamètre-hauteur ». Nous avons effectué les calculs suivants

$$\bar{x} = 39,49 \qquad \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})^2 = 28,29 \qquad \bar{y} = 18,34$$
$$\frac{1}{20}\sum_{i=1}^{20}(y_i - \bar{y})^2 = 2,85 \quad \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})(y_i - \bar{y}) = 6,26\ .$$

(a) On note $\widehat{y} = \widehat{\beta}_1 + \widehat{\beta}_2 x$ la droite de régression. Donner l'expression de $\widehat{\beta}_2$ en fonction des statistiques élémentaires ci-dessus. Calculer $\widehat{\beta}_1$ et $\widehat{\beta}_2$.

(b) Donner et commenter une mesure de la qualité de l'ajustement des données au modèle. Exprimer cette mesure en fonction des statistiques élémentaires .

Ex. 6 Démonstration partielle du théorème de Gauss-Markov. On a n variables explicatives $x_1, \ldots, x_n$ et n variables à expliquer $y_1, \ldots, y_n$. Soit $\widetilde{\beta}_2$ un estimateur linéaire de β_2. On peut donc écrire $\widetilde{\beta}_2 = \sum_{i=1}^{n} \widetilde{\lambda}_i y_i$.

(a) Trouver deux conditions sur la somme des $\widetilde{\lambda}_i$ pour que $\widetilde{\beta}_2$ ne soit pas biaisé.

(b) Montrer que $\mathbb{E}((\widetilde{\beta}_2 - \widehat{\beta}_2)(\widehat{\beta}_2 - \beta_2)) = 0$. En déduire que $\mathrm{Var}(\widehat{\beta}_2) \leq \mathrm{Var}(\widetilde{\beta}_2)$.