Université Claude Bernard Lyon 1
L2 MASS UE 42
Année 2011-2012

Proba-Stat : fiche n° 5

Exercice 1. On cherche à étudier un phénomène Y que l'on pense lié à deux paramètres $X1$ et $X2$. Les huit observations sont regroupées dans le tableau ci-dessous.

Obs.	Y	$X1$	$X2$
1	0.2	5	4
2	2.4	8	8
3	0.3	3	3
4	2.6	5	7
5	2.7	5	7
6	0.7	8	6
7	1.8	4	5
8	1	1	6

1. Tracer les nuages de points de Y en fonction de $X1$ puis de Y en fonction de $X2$.
2. Lequel de ces deux nuages vous semble le plus proche d'un modèle linéaire ?
3. Effectuer la régression linéaire de Y en $X2$. Déterminer notamment le coefficient de détermination de cette régression.
4. Effectuer maintenant la régression linéaire multiple de Y en $X1$ et $X2$: écrire la matrice X usuelle, le modèle sous la forme matricielle (en précisant la nature de chacun des objets apparaissant dans ce modèle !).
   On donne les matrices $X^t.X$ et $X.X^t$: utiliser celle qui convient pour déterminer les estimateurs de la régression !
$$X^t.X = \begin{pmatrix} 8 & 42 & 46 \\ 42 & 244 & 255 \\ 46 & 255 & 284 \end{pmatrix} \qquad X.X^t = \begin{pmatrix} 42 & 73 & 28 & 54 & 54 & 65 & 41 & 45 \\ 73 & 129 & 49 & 97 & 97 & 113 & 73 & 81 \\ 28 & 49 & 19 & 37 & 37 & 43 & 28 & 31 \\ 54 & 97 & 37 & 75 & 75 & 83 & 56 & 63 \\ 54 & 97 & 37 & 75 & 75 & 83 & 56 & 63 \\ 65 & 113 & 43 & 83 & 83 & 101 & 63 & 69 \\ 41 & 73 & 28 & 56 & 56 & 63 & 42 & 47 \\ 45 & 81 & 31 & 63 & 63 & 69 & 47 & 53 \end{pmatrix}$$
5. Déterminer le coefficient de détermination de la régression multiple et comparer celui de la régression simple. Ce résultat est-il attendu ?

Exercice 2. *Régression polynomiale.* On étudie un phénomène Y dont on pense qu'il est lié polynomialement à une variable X. On suppose que le degré du polynôme en question est inférieur à d (un entier fixé supérieur ou égal à 2) : le modèle est donc
$$Y = \sum_{j=1}^{d} a_j X^j + \epsilon_j$$
où les a_j sont des constantes inconnues et où les ϵ_j sont des variables aléatoires indépendantes de loi Normale centrée et de variance σ^2 inconnue.

On dispose de n observations $(x_i, y_i)_{i \leq d}$.

1. Écrire ce modèle sous forme vectorielle.
2. Donner une estimation de la variance σ^2.
3. Donner plusieurs méthodes permettant de choisir le degré du polynôme, sachant que l'on privilégie pour ce modèle un polynôme de petit degré.