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

Proba-Stat : CC n° 2

Exercice 1. Questions de cours

1. On considère des données $(x_i, y_i)_{i\leq n}$. Montrer l'égalité

$$\sum_{i=1}^{n} [(x_i - \overline{x})(y_i - \overline{y})] = \left(\sum_{i=1}^{n} x_i y_i\right) - n\overline{x}\,\overline{y}$$

   Qu'obtient-on dans le cas où, pour tout i, $y_i = x_i$?

2. On souhaite effectuer une régression linéaire multiple pour expliquer un paramètre observé Y par deux facteurs $X1$ et $X2$: décrire le modèle statistique (en précisant toutes les hypothèses). Préciser la quantité que l'on cherche à minimiser. Une matrice X intervient dans la recherche : l'expliciter. Indiquer l'estimateur de $\hat{y}_i$ des y_i qui découle de cette minimisation. Quelle est la différence entre erreur et résidu ?
3. On cherche à expliquer linéairement une variable Y par une variable $X1$. Montrer que la régression linéaire multiple (appliquée avec une seule variable explicative) donne le même estimateur de Y que la régression linéaire simple.

Exercice 2. On étudie le loyer mensuel (en euros, hors charge) de studios à Villeurbanne, montant que l'on cherche à expliquer linéairement par la surface des appartements (en m^2). Les données observées début mai 2012 sur le site seloger.com sont les suivantes :

obs.	1	2	3	4	5	6	7	8	9	10
surface	17	18	19	22	27	31	32	33	36	47
loyer	390	305	310	320	396	427	370	430	480	620

On donne également

$$\overline{x} = 28,2 \qquad \overline{y} = 404,8 \qquad \sum x_i^2 = 8766 \qquad \sum y_i^2 = 1719370 \qquad \sum x_i y_i = 121429$$

1. Donner l'équation de la droite de régression linéaire obtenue par la méthode de moindres carrés.
2. Recopier et compléter le tableau d'analyse de la variance suivant :

Source	Degrés de liberté	Somme des carrés
Modèle		SM=
Résidus		SR=
Total		ST=

3. Calculer le coefficient de détermination de la régression.
4. Un propriétaire possède un studio dont la surface est $x_{11} = 30\text{m}^2$. Quel est le montant mensuel de location donné par le modèle de régression linéaire ? Donner un intervalle de confiance de y_{11}.
5. Donner un estimateur (sans biais) de la variance de la loi des erreurs.

Tournez S.V.P

Exercice 3. On souhaite expliquer un phénomène Y en fonction de deux paramètres $X1$ et $X2$. Les observations sont regroupées dans le tableau ci-dessous.

Obs.	Y	$X1$	$X2$
1	-4	3	1
2	-2	2	-1
3	5	-4	0
4	7	0	-8
5	-6	1	5
6	-4	-2	3

1. Écrire le modèle de régression linéaire multiple sous forme matricielle, en précisant les différentes hypothèses du modèle.
2. Quelles lignes de commande du logiciel R permettent de créer les vecteurs y, $x1$ et $x2$ contenant les données et d'effectuer cette régression linéaire ?
3. Déterminer le vecteur $\widehat{\beta} = {}^t(\widehat{b_0}, \widehat{b_1}, \widehat{b_2})$ des estimations des coefficients de la régression par la méthode des moindres carrés.
4. Calculer les estimations $\hat{y}_i$ et les résidus e_i, pour i de 1 à 6. En déduire $\hat{\sigma}^2$ une estimation sans biais de la variance des termes d'erreurs.

   On donne ${}^t\big(Y - \bar{Y}\big) . \big(Y - \bar{Y}\big) \approx 143,33$.
5. Recopier et compléter le tableau de décomposition ci-dessous :

Source de variation	ddl	Somme des carrés des écarts	Carré moyen	Fisher
Régression	$p = \ldots$	$SM = \ldots$	$SM/\ldots = \ldots$	$\frac{SM/\ldots}{SR/\ldots} = \ldots$
Résiduelle	$\ldots = \ldots$	$SR = \ldots$	$SR/\ldots = \ldots$	
Total	$n-1 = \ldots$	$ST = \ldots$		

6. Tester la significativité du modèle de régression entier.

   On donne :
   – la somme des carrés des écarts résiduels de la régression linéaire simple de Y en fonction de $X1$: $SR_{X_1} = 109,33$
   – la somme des carrés des écarts résiduels de la régression linéaire simple de Y en fonction de $X2$: $SR_{X_2} = 43,33$
7. Tester la significativité de la variable $X1$ sachant que $X2$ est dans le modèle pour un risque de 5%. Tester la significativité de la variable $X2$ sachant que $X1$ est dans le modèle pour un risque de 5%.