Séance 3
M1 Actuariat

Exercices

Exercice 1 : Coût d'un véhicule

Afin d'étudier comment varie le coût de maintenance d'un véhicule utilitaire en fonction de l'âge de celui-ci, une entreprise a collecté les données suivantes :

Age (x_1) (en mois)	Coût annuel (y) (en centaine d'euros)
15	48
8	43
36	77
41	89
16	50
8	40
21	56
21	62
53	100
10	47
32	71
17	58
58	102
6	35
20	60

Les valeurs suivantes ont été calculées :

$$\sum_{i=1}^{15} x_{1i} = 362 \qquad \sum_{i=1}^{15} x_{1i}^2 = 12490 \qquad \sum_{i=1}^{15} x_{1i} y_i = 27437$$
$$\sum_{i=1}^{15} y_i = 938 \qquad \sum_{i=1}^{15} y_i^2 = 64926$$

Nombre d'observations : $N = 15$.

Partie 1

On cherche à estimer les coefficients d'une régression linéaire entre les variables x_1 et y, de la forme :

$$y_i = a_0 + a_1 x_{1i} + u_i \tag{1}$$

On suppose que $u \sim \mathcal{N}\left(0, \sigma^2 I_N\right)$.

1. Démontrer la formule permettant de calculer $\widehat{a}_0$ et $\widehat{a}_1$, les estimateurs des MCO des paramètres a_0 et a_1.
   Montrer que ces estimateurs sont sans biais.
   Calculer la valeur de ces estimateurs à partir des données et interpréter les résultats.
2. Rappeler l'équation de décomposition de la variance.
   Sachant que la valeur de la somme des carrés expliqués (SCE) par le modèle est : $\sum_{i=1}^{15}\left(\widehat{y}_i - \overline{\overline{y}}\right)^2 = 6137.71889$, calculer $\sum_{i=1}^{15} e_i^2$, la somme des carrés des résidus (SCR).
3. Calculer le coefficient de détermination R^2.
   Interpréter littérairement la valeur obtenue.
4. Donner la formule d'un estimateur sans biais de σ^2 (démonstration non demandée).
   Calculer sa valeur.
   En déduire une estimation des variances de $\widehat{a}_0$ et $\widehat{a}_1$.
5. Expliquer la construction d'un intervalle de confiance au seuil α pour $\widehat{a}_1$.
   Le calculer pour $\alpha = 5\%$.
   Interpréter littérairement.
6. Tester si les coefficients a_0 et a_1 sont significativement différents de 0 au seuil $\alpha = 5\%$.
   Interpréter littérairement les résultats des tests.
7. Déterminer une prévision du coût de maintenance pour un véhicule de 4 ans.
   Calculer son intervalle de confiance au seuil de 5 %.
   Interpréter littérairement.

Partie 2

On souhaite maintenant comparer les résultats de l'estimation du modèle (1) avec ceux de l'estimation du modèle suivant :

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + u_i \tag{2}$$

où :
- y_i et x_{1i} sont les mêmes variables que celles utilisées dans le modèle (1).
- x_{2i} est une variable dichotomique prenant la valeur 1 si le véhicule est de couleur claire, 0 si la couleur est foncée.
- x_{3i} est une variable dichotomique prenant la valeur 1 si le véhicule a un moteur diesel et 0 s'il a un moteur essence.

Les résultats de l'estimation de ce modèle sur le même échantillon sont les suivants :

$$\begin{array}{ccccccccccc} \widehat{y}_i & = & 31.748 & + & 1.152 & x_{1i} & - & 0.025 & x_{2i} & + & 5.600 & x_{3i} \\ & & (1.253) & & (0.061) & & & (1.549) & & & (2.022) & \end{array}$$

(.) écart-types estimés

$SCE' = \sum_{i=1}^{15}\left(\widehat{y}_i - \overline{\overline{y}}\right)^2 = 6194.34598$

1. Tester la significativité des variables x_{1i}, x_{2i} et x_{3i}.
   Interpréter leurs coefficients estimés ($\widehat{b}_1$,$\widehat{b}_2$ et $\widehat{b}_3$) et commenter.
2. Tester la significativité globale du modèle (2) au seuil $\alpha = 5\%$.
3. Comparer les résultats de l'estimation des modèles (1) et (2).
   Quel modèle privilégier ? Argumenter la réponse.