Université de Jijel. **2024/ 2025** Z. Djeridi
Département de Mathématiques.
Master 1 proba-stat.

Série de TD n 2

Exercice 1:

On appelle “fréquence seuil” d’un sportif amateur sa fréquence cardiaque obtenue après trois quarts d’heure d’un effort soutenu de course à pied. Celle-ci est mesurée à l’aide d’un cardio-fréquencemètre.
On cherche à savoir si l’âge d’un sportif a une influence sur sa fréquence seuil. On dispose pour cela de 20 valeurs du couple (x_i, y_i), où x_i est l’âge et y_i la fréquence seuil du sportif. On a obtenu $(\bar{x}, \bar{y}) = (35,6;\ 170,2)$ et :
$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 1991$; $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 189,2$; $\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = -195.$□□$4$

1. Calculer la droite des moindres carrés pour le modèle $y = \beta_1 + \beta_2 x + \varepsilon$.
2. Calculer le coefficient de détermination R^2. Commenter la qualité de l’ajustement des données au modèle.
3. Avec ces estimateurs, la somme des carrés des résidus vaut $\sum_{i=1}^{n}(y_i - \hat{y})^2 = 170$. Si on suppose les perturbations ε_i gaussiennes, indépendantes et de même variance σ^2, en déduire un estimateur non biaisé $\hat{\sigma}^2$ de σ^2.
4. Donner un estimateur $\hat{\sigma}_2^2$ de la variance de $\widehat{\beta}_2$.
5. Tester l’hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$ pour un risque de 5%. Conclure sur la question de l’influence de l’âge sur la fréquence seuil.

Exercice 2:

Le taux d’hémoglobine dans le sang au cours d’une opération chirurgicale varie en fonction de la durée de l’opération et de volume de sang perdu pendant l’opération. On dispose d’un échantillon de taille 8 donnant pour un individu i : y_i la variation du taux d’hémoglobine (en pourcentage), x_i^1 la durée de l’opération en heures décimales et x_i^2 le volume en litres de sang perdu.

y	-1.71	-4.51	-5.81	-1.16	-4.33	-3.33	0.41	-2.95
x^1	1.74	1.35	1.41	1.85	1.81	1.68	1.50	1.98
x^2	0.51	0.56	0.62	0.49	0.56	0.52	0.25	0.46

I- On suppose que y_i dépend de façon linéaire de x_i^1 et x_i^2. On suppose de plus que ce modèle linéaire conduit à des erreurs ϵ_i indépendantes de même loi $\mathcal{N}(0, \sigma^2)$.

1. Ecrire le modèle décrit ci-dessus pour tout $i = 1, \ldots, 8$ puis sous forme matricielle.
2. Calculer $\hat{\beta}$. On donne :
$$(X'X)^{-1} = \begin{bmatrix} 11.63 & -4.97 & -6.50 \\ -4.97 & 2.86 & 0.43 \\ -6.50 & 0.43 & 11.67 \end{bmatrix} \; et \; X'Y = \begin{bmatrix} -23.39 \\ -38.06 \\ -12.98 \end{bmatrix}$$
3. Donner la table d’analyse de la variance. On donne *RSS*=*SCR*= 5.534 et $\sum_i y_i^2 = 97.07$

	d.d.l	SC	SC moyen	Stat F
Régression Résiduelle				
Totale				

4. Calculer le coefficient de détermination R^2.

5. Ecrire en fonction de vos paramètres puis tester l'hypothèse selon laquelle le taux d'hémoglobine ne dépend ni de la durée de l'opération ni du volume de sang perdu. On donne qf(0.95,2,5) = 5.786.
6. Ecrire en fonction de vos paramètres puis tester l'hypothèse selon laquelle le taux d'hémoglobine ne dépend que du volume de sang perdu. qf (0.95,1,5) = 6.608

Exercice 3:

Nous voulons expliquer la concentration e l'ozone sur Jijel en fonction des variables T9, T12, Ne9, Ne12 et Vx. Les sorties données par GNU-R sont :

Coefficients :

	Estimate	Std.Error	t value	Pr(> $\|t\|$
(Intercept)	62	10	1	0
T9	-4	2	-5	0
T12	5	0.75	3	0
Ne9	-1.5	1	4	0.08
Ne12	-0.5	0.5	5	0.53
Vx	0.8	0.15	5.5	0

Multiple R-squared : 0.6666, Adjusted R-squared :0.6081

Residual Standard error: 16 on 124 degrees of freedom

F-statistic: 6 on 7 and 8 DF, p-value: 0

I)

1. Ecrire le modèle linéaire décrit par ces données pour tout $i = 1, \dots, n$, puis sous forme matricielle en donnons les hypothèses associées.
2. Donner l'expression de l'estimateur de MCO des paramètres inconnus, avec sa valeur.

II)

1. Compléter approximativement la sortie ci-dessus.
2. Rappeler la statistique de test et tester la nullité des paramètres séparément au seuil de 5%.
3. Rappeler la statistique de test et tester la nullité des paramètres simultanée autres que la constante au seuil de 5%.
4. Voici une nouvelle valeur (T9=10, T12=20, NE9=, Ne12=0, Vx=1), peut-on effectuer la prévision ?
5. Interpréter le nombre « Multiple R-squared : 0.6666 » du tableau.