

TD et TP: Régression linéaire

Exercice 1 : Modèle de croissance humaine

Un père a deux garçons, et s'inquiète de la croissance de son cadet qu'il trouve petit. Il décide de faire un modèle familial à partir des mesures de tailles en fonction de l'âge de l'aîné :

âge	3	4	5	6	7	8	9	10	11	12
taille	96	104.8	110.3	115.3	121.9	127.4	130.8	136	139.7	144.5

1°) Représenter les données sur un graphique et justifier l'utilisation d'un modèle de régression linéaire simple. Discuter les conditions si nécessaires.

2°) Estimer les coefficients de la régression et tracez sur le graphique la droite estimée .

3°) Calculer le R^2 et représenter les résidus. La régression semble-t elle valable ?

Exercice 2 : Hauteur des arbres

Nous souhaitons exprimer la hauteur Y d'un arbre en fonction de son diamètre X à 1m30 du sol. Pour cela, nous avons mesuré 20 couples de diamètre-hauteur et les résultats ci-dessous sont disponibles :

$$\bar{x} = 34.9; \frac{1}{20} \sum_{i=1}^{20} (x_i - \bar{x})^2 = 28.29; \bar{y} = 18.34$$

$$\frac{1}{20} \sum_{i=1}^{20} (y_i - \bar{y})^2 = 2.85; \frac{1}{20} \sum_{i=1}^{20} (y_i - \bar{y})(x_i - \bar{x}) = 6.26$$

1°) On note $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$, l'estimation de la droite de régression. Donner l'expression de $\hat{\beta}_0$ et $\hat{\beta}_1$ en fonctions des statistiques élémentaires ci-dessus. Calculer $\hat{\beta}_0$ et $\hat{\beta}_1$.

2°) Donner une mesure de qualité d'ajustement des données au modèle. Exprimer cette mesure à l'aide de statistique élémentaires. Calculer et commenter.

3°) Testez $H_0 : \beta_j = 0$ contre $H_1 : \beta_j \neq 0$ pour $j = 0, 1$. Commenter.

Exercice 3

On étudie l'influence des heures de travail et du capital utilisé sur la production industrielle. Pour cela, on dispose des observations de 9 entreprises résumées dans le tableau ci-dessus:

Obs	Travail (heures)	Capital (machines/heures)	Production (100 tonnes)
1	1100	300	60
2	1200	400	120
3	1430	420	190
4	1500	400	250
5	1520	510	300
6	1620	590	360
7	1800	600	380
8	1820	630	430
9	1800	610	440

On suppose que la production est expliquée par un modèle de régression linéaire multiple avec deux variables explicatives, le capital et le travail.

1°) Écrire le modèle sous forme matricielle .

2°) Estimer le vecteur β puis donner l'équation de l'hyperplan des moindres carrés. Pour cela on donne

$$(X'X)^{-1} = \begin{pmatrix} 6.30477 & -0.007800 & 0.011620 \\ -0.007800 & 0.000015 & -0.000031 \\ 0.011620 & -0.000031 & 0.000072 \end{pmatrix}$$

3°) Calculer les estimations de σ^2 et $V(\hat{\beta})$.

4°) Calculer les intervalles de confiance à 95% pour β_j , $j = 0, 1, 2$.

5°) Calculer les intervalles de confiance simultanées pour β_j , $j = 0, 1, 2$ au niveau de confiance 95% par la méthode de Bonferroni et la méthode de Scheffé.

6°) Donner l'expression de la région de confiance de β et calculer les régions de confiance des couples (β_i, β_j) , $i \neq j$.

7°) Tester l'hypothèse nulle $H_0 : "\beta_j = 0"$ contre l'alternative $H_1 : "\beta_j \neq 0"$ pour $j = 0, 1, 2$.

8°) Construire le tableau d'analyse de variance et réaliser le test de Fisher global d'hypothèse nulle $H_0 : "\beta_1 = \beta_2 = 0"$ au risque $\alpha = 5\%$. Conclure.