Feuille 6 : Régression linéaire

Allan Merino

DUT Mesures Physiques, Metz - Mars 2016

Exercice 1

L'étude statistique ci-dessous porte sur les poids respectifs des pères et de leur fils aîné.

Père	65	63	67	64	68	62	70	66	68	67	69	71
Fils	68	66	68	65	69	66	68	65	71	67	68	70

Voici les résultats numériques que l'on a obtenus :

$$\sum_{i=1}^{12} p_i = 800 \qquad \sum_{i=1}^{12} \mathrm{p}_\mathrm{i}^2 = 53418 \qquad \sum_{i=1}^{12} p_i f_i = 54107 \qquad \sum_{i=1}^{12} f_i = 811 \qquad \sum_{i=1}^{12} \mathrm{f}_\mathrm{i}^2 = 54849$$

1. Calculez la droite des moindres carrés du poids des fils en fonction du poids des pères.
2. Calculez la droite des moindres carrés du poids des pères en fonction du poids des fils.

Exercice 2

Douze personnes sont inscrites à une formation. Au début de la formation, ces stagiaires subissent une épreuve A notée sur 20. A la fin de la formation, elles subissent une épreuve B de niveau identique. Les résultats sont donnés dans le tableau suivant :

Epreuve A	3	4	6	7	9	10	9	11	12	13	15	4
Epreuve B	8	9	10	13	15	14	13	16	13	19	6	19

1. Représenter le nuage de points. Déterminer la droite de régression. Calculer le coefficient de détermination. Commenter.
2. Deux stagiaires semblent se distinguer des autres. Les supprimer et déterminer la droite de régression sur les dix points restants. Calculer le coefficient de détermination. Commenter.

Exercice 3

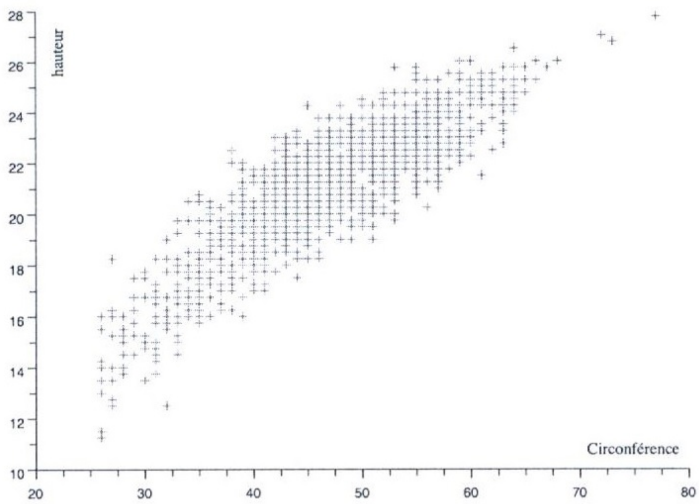
Nous souhaitons exprimer la hauteur y (en pieds) d'un arbre d'une essence donnée en fonction de son diamètre x (en pouces) à $1m30$ du sol. Pour ce faire, nous avons mesuré 20 couples (diamètre, hauteur) et effectué les calculs suivants : $\bar{x} = 4.53$, $\bar{y} = 8.65$ et

$$\frac{1}{20}\sum_{i=1}^{20}(x_i-\bar{x})^2 = 10,97 \qquad \frac{1}{20}\sum_{i=1}^{20}(y_i-\bar{y})^2 = 2,24 \qquad \frac{1}{20}\sum_{i=1}^{20}(x_i-\bar{x})(y_i-\bar{y}) = 3,77$$

1. On note $y = \hat{\beta}_0 + \hat{\beta}_1 x$ la droite de régression. Calculer $\hat{\beta}_0$ et $\hat{\beta}_1$.
2. Donner et commenter une mesure de la qualité de l'ajustement des données au modèle. Exprimer cette mesure en fonction des statistiques élémentaires. Commenter le résultat.
3. On donne les estimations de l'écart-type de $\hat{\beta}_0$, $\hat{\sigma}_0 = 1,62$ et de $\hat{\beta}_1$, $\hat{\sigma}_1 = 0,05$. On suppose les perturbations ε_i gaussiennes, centrées, de même variance et indépendantes. Tester $H_0 : \beta_j = 0$ contre $H_1 : \beta_j \neq 0$ pour $j = 0, 1$. Pourquoi ce test est-il intéressant dans notre contexte ? Que pensez-vous du résultat ?

Exercice 4

On souhaite expliquer la hauteur y (en mètres) d'un arbre en fonction de sa circonférence x (en centimètres) à $1m30$ du sol. On a relevé $n = 1429$ couples (x_i, y_i), le nuage de points étant représenté ci-dessous.



On a obtenu $(\bar{x}, \bar{y}) = (47,3; 21,2)$ et

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 102924 \qquad \sum_{i=1}^{n}(y_i-\bar{y})^2 = 8857 \qquad \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = 26466$$

1. Calculer la droite des moindres carrés pour le modèle $y = \beta_1 + \beta_2 x + \varepsilon$. et la représenter sur la figure précédente.
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 alors $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 2052$. Si on suppose les perturbations ε_i gaussiennes, centrées, indépendantes et de même variance σ^2, en déduire un estimateur $\hat{\sigma^2}$ de σ^2.
4. Donner un estimateur $\hat{\sigma_1^2}$ de la variance de $\hat{\beta}_1$.
5. Tester l'hypothèse $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 = 0$.

Exercice 5

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-fréquence-mètre. On cherche à savoir si l'âge d'un sportif a une influence sur sa fréquence seuil. On dipose 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 :

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 1991 \qquad \sum_{i=1}^{n}(y_i - \bar{y})^2 = 189,2 \qquad \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = -195,4$$

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 $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 170$. Si on suppose les perturbations ε_i gaussiennes, centrées, indépendantes et de même variance σ^2, en déduire un estimateur $\hat{\sigma^2}$ de σ^2.
4. Donner un estimateur $\hat{\sigma_2^2}$ de la variance de $\hat{\beta}_2$.
5. Tester l'hypothèse $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$ pour un risque de 5 pourcents. Conclure sur la question de l'influence de l'âge sur la fréquence seuil.