Université de Lyon 2
Licence 3 MIASHS-IDS

Modèle linéaire
Année 2021-22

Contrôle continu 1

Il sera tenu compte de la présentation et de la qualité des explications pour justifier les résultats.

Exercice 1 Les données suivantes représentent le score en anglais sur 100 d'un échantillon de 12 personnes d'âge différent :

âge (années)	29	28	27	26	25	24	23	22	21	20	19	18
score	76.1	77	78.1	78.2	78.8	79.7	79.9	81.1	81.2	81.8	82.8	83.5

1. Tracer les points, la droite de régression et en déduire graphiquement l'équation de la droite de régression (on cherchera deux points visiblement proches de la vraie droite de régression ce qui permettra de calculer son équation, à une légère erreur près).

2. Calculer une estimation des coefficients de la droite de régression pour ces données.

3. *On se focalise maintenant sur le point de vue de la statistique inférentielle pour le modèle linéaire. Pour ce faire, on suppose*

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

avec $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$.

   (a) Quelle est la loi de Y_i, $i = 1, \ldots, n$?

   Il a été vu au TD 1 que

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right)$$

   De plus, la variance σ^2 *des résidus peut être estimée par*

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2.$$

   On en a déduit que

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim \tau_{n-2}. \tag{1}$$

(b) Proposez un test afin de savoir si le coefficient est nul. Le résultat vous surprend-t-il ?

Exercice 2 (Droite de régression et points atypiques) 7 personnes sont inscrites à une formation à la conduite automobile. Au début de la formation, ces stagiaires subissent un test A notée sur 20. A la fin de la formation, elles subissent un test B de niveau identique. Les résultats sont donnés dans le tableau suivant :

Test A	3	4	6	7	9	10	9
Test B	8	9	10	13	15	14	13

1. Représenter le nuage de points.

2. La fonction "somme des carrés des erreurs" est donnée par

$$f(b) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2, \tag{2}$$

comme fonction des coefficients b_0 et b_1.

(a) Déterminer la dérivée de cette fonction, lorsqu'elle est considérée comme fonction de b_1 uniquement.

(b) Déterminer la dérivée de cette fonction, lorsqu'elle est considérée comme fonction de b_0 uniquement.

(c) Déterminer les coefficients de la droite de régression. Commenter.

3. Deux stagiaires semblent se distinguer des autres. Lesquels ?

4. *La "somme des carrés des erreurs" peut aussi se regarder comme fonction du vecteur y dont les composantes sont les y_i, $i = 1, \ldots, n$. On l'écrit dans le cas spécial où on prend $b_0 = \hat{\beta}_0$ et $b_1 = \hat{\beta}_1$:*

$$g(y) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2, \tag{3}$$

   (a) Calculer la dérivée de g lorsqu'elle est considérée comme fonction de y_{i_0} pour un indice i_0 particulier.

   (b) Calculer les valeurs numériques de ces dérivées pour chaque $i_0 = 1, \ldots, n$. Commenter.