Exercice n°4

1) moyenne empirique

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{ici} \quad \bar{x}_{10} = \frac{1}{10}\sum_{i=1}^{10} x_i = 20.3$$

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{ici} \quad \bar{y}_{10} = \frac{1}{10}\sum_{i=1}^{10} y_i = 113.2$$

variance empirique :

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \quad \text{ici} \quad \sigma_x^2 = \frac{1}{9}\sum_{i=1}^{10} (x_i - \bar{x}_n)^2 = \frac{76.1}{9}$$

$$\sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y}_n)^2 \quad \text{ici} \quad \sigma_y^2 = \frac{1}{9}\sum_{i=1}^{10} (y_i - \bar{y}_n)^2 = \frac{347.6}{9}$$

covariance empirique :

$$\mathrm{cov}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n)$$

$$\text{ici} \quad \mathrm{cov}(x,y) = \frac{1}{9}\sum_{i=1}^{10} (x_i - \bar{x}_n)(y_i - \bar{y}_n) = \frac{146.4}{9}$$

2)

La droite de régression expliquant y par x est :

$$y = \hat{a}_n + \hat{b}_n x \quad \text{avec} \quad \hat{a}_n = \bar{y}_n - \hat{b}_n x_n$$

$$\text{et} \quad \hat{b}_n = \frac{\Sigma (x_i - \bar{x}_n)(y_i - y_n)^2}{\Sigma (x_i - \bar{x}_n)^2}$$

$$\text{ici} \quad \hat{b}_n = \frac{146.4}{76.1}$$

$$\hat{a}_n = 113.2 - \frac{146.4}{76.1} \cdot 20,3$$

3)

(a) le coefficient de détermination : $R^2 = \dfrac{\Sigma(\hat{y}_i - \bar{y}_n)^2}{\Sigma(y_i - \bar{y}_n)^2}$

erreur quadratique : $\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = EQ$

a $\sum_{i=1}^{n} (y_i - \bar{y}_n)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_n)^2$

d'où $$R^2 = \frac{\frac{1}{n-1} \sum (y_i - \bar{y}_n)^2 - \frac{1}{n-1} \sum (y_i - \hat{y}_i)^2}{\frac{1}{n-1} \sum (y_i - \bar{y}_n)^2}$$

soit $$R^2 = 1 - \frac{EQ}{\sigma_y^2}$$

(b)

Dans le cas de la régression linéaire simple, on a :

$R^2 = (r)^2$ avec r le coefficient de corrélation

d'où $$(r)^2 = 1 - \frac{EQ}{\sigma_y^2}$$

(c)

si $|r|$ est proche de 1, cela signifie que l'erreur quadratique est proche de 0. Et donc que les $\hat{y}_i$ sont d'autant plus proches des y_i !

Erreurs :

- il faut être cohérent entre la définition des variances et covariances empiriques
- il faut connaître les définitions des coefficients de détermination et de corrélation.