# La méthode des moindres carrés ordinaires :

Économétrie : La régression linéaire simple.

Alexis VIALATTE

April 12, 2021

# Sommaire

# Sujet

Énoncé : En Mondelēzie, dans une école de tricotage de rien ou tricotage de R en langage jeune, l'enseignant M. Favard a demandé à ses étudiants de ne pas comsommer de chocolat de la région de la Côte d'Or pour ne pas attraper la maladie de Pacques et risquer de ne pas pouvoir assister à l'examen. A-t-il raison ? Est-ce que la consommation de chocolat de la région de la Côte d'Or augmente le risque d'attraper la maladie de Pacques ? Pour répondre à ces questions, nous disposons la base de données suivante concernant 10 étudiants de Lécot :

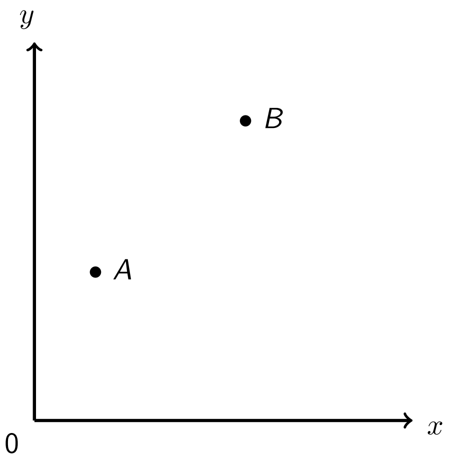
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 2 | 4 | 6 | 8 | 10 | 2 | 4 | 6 | 8 | 10 |
| $y_i$ | 2.5 | 6 | 7.5 | 10 | 2.5 | 1.5 | 4 | 4.5 | 6 | 9.5 |

Table: Base de données `lecot`

# Sujet

La base de données contient 10 individus notés $i$ où $x_i$ correspond au nombre de tablettes de chocolat consommées par le $i$ème individus et $y_i$ correspond au nombre de jours de sommeil causé par la maladie de Pacques pour le $i$ème individus *A priori*, $x_i$ est la variable explicative et $y_i$ est la variable expliquée.

# L'estimation sous la forme d'une droite



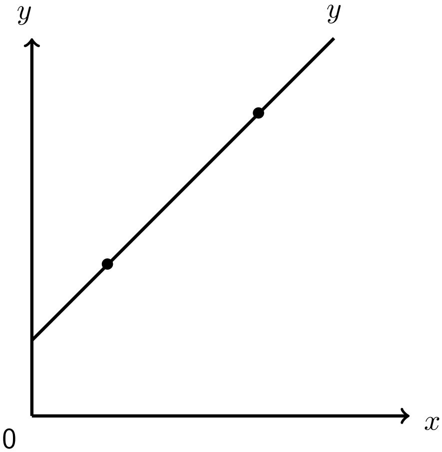
# L'estimation sous la forme d'une droite

L'équation d'une droite affine est donnée par $y = ax + b$ où $a$ est la pente de la droite $y$ et $b$ est l'ordonnée à l'origine, c'est à dire quand $x = 0$, de la droite $y$.

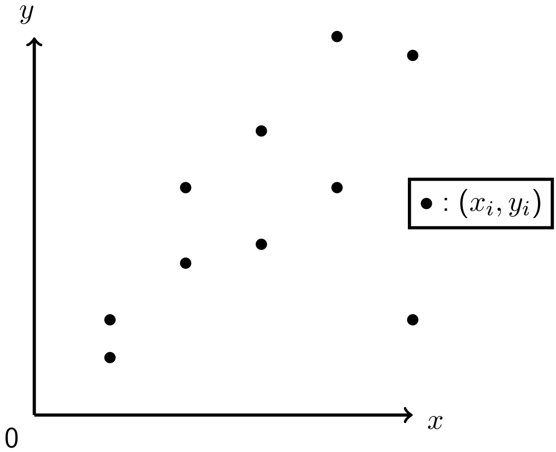
$$\left\{ \begin{array}{ll} A: & 2 = a + b \\ B: & 4 = 3a + b \end{array} \right. \iff \left\{ \begin{array}{ll} A: & a = 1 \\ B: & b = 1 \end{array} \right.$$

$$y = x + 1 \tag{1}$$

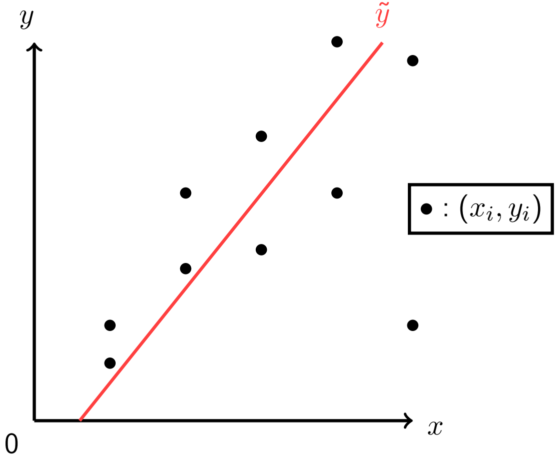
# L'estimation sous la forme d'une droite



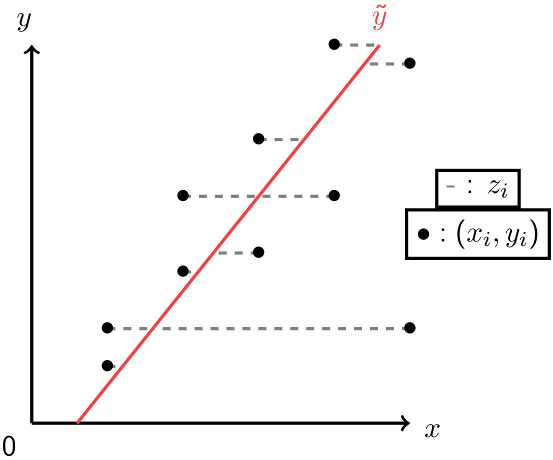
# La base de données



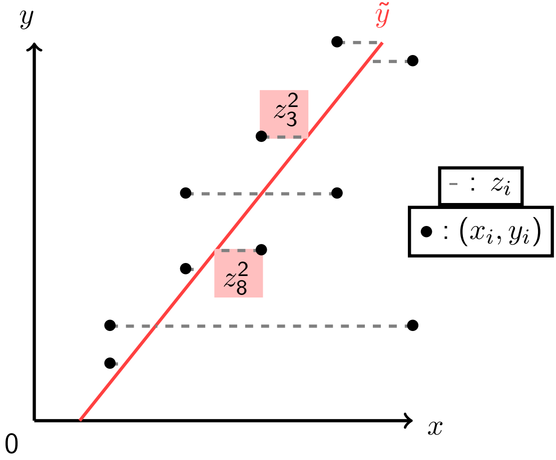
# La droite supposée $\tilde{y}$



# Les erreurs de la droite supposée $\tilde{y}$



# Les erreurs au carré de la droite supposée $\tilde{y}$



# La méthode des moindres carrés ordinaires

L'équation de la régression linéaire simple est donc $y_i = \alpha + \beta x_i + z_i$ donc l'équation de l'erreur est $z_i = y_i - \alpha - \beta x_i$.

$$\begin{cases} \hat{\alpha} = ? \\ \hat{\beta} = ? \end{cases}$$

# La méthode des moindres carrés ordinaires

Nous appliquons la méthode des **MCO** et nous calculons les dérivés de premier ordre en fonction de $\alpha$ et de $\beta$ :

$$Min\left(\sum_{i=1}^{I} {z_i}^2\right) = Min\left[\sum_{i=1}^{I}\left(y_i - \alpha - \beta x_i\right)^2\right] \qquad (2)$$

$$\begin{cases} \frac{\partial \text{MCO}}{\partial \alpha} = 2\sum_{i=1}^{I}\left(y_i - \alpha - \beta x_i\right) \\ \frac{\partial \text{MCO}}{\partial \beta} = 2\sum_{i=1}^{I} x_i\left(y_i - \bar{y} + \beta\bar{x} - \beta x_i\right) \end{cases} \Longleftrightarrow \begin{cases} \frac{\partial \text{MCO}}{\partial \alpha} = \sum_{i=1}^{I} y_i - \alpha\sum_{i=1}^{I} - \beta\sum_{i=1}^{I} x_i \\ \frac{\partial \text{MCO}}{\partial \beta} = \sum_{i=1}^{I} x_i y_i - \bar{y}\sum_{i=1}^{I} x_i + \beta\bar{x}\sum_{i=1}^{I} x_i - \beta\sum_{i=1}^{I} {x_i}^2 \end{cases}$$

$$\begin{cases} \frac{\alpha\sum_{i=1}^{I}}{I} = \frac{\sum_{i=1}^{I} y_i - \beta\sum_{i=1}^{I} x_i}{I} \\ \frac{\beta\left(\sum_{i=1}^{I} {x_i}^2 - \bar{x}\sum_{i=1}^{I} x_i\right)}{I} = \frac{\sum_{i=1}^{I} x_i y_i - \bar{y}\sum_{i=1}^{I} x_i}{I} \end{cases} \Longleftrightarrow \begin{cases} \hat{\alpha} = \frac{1}{I}\sum_{i=1}^{I} y_i - \hat{\beta}\frac{1}{I}\sum_{i=1}^{I} x_i \\ \hat{\beta} = \frac{\frac{1}{I}\sum_{i=1}^{I}(x_i y_i - \bar{y}\bar{x})}{\frac{1}{I}\sum_{i=1}^{I}(x_i - \bar{x})^2} \end{cases}$$

# La méthode des moindres carrés ordinaires

Nous savons que $\bar{y} = \frac{1}{I}\sum_{i=1}^{I} y_i$, $\bar{x} = \frac{1}{I}\sum_{i=1}^{I} x_i$, $Cov(x,y) = \frac{1}{I}\sum_{i=1}^{I}(x_i y_i - \bar{y}\bar{x})$ et $Var(x) = \sum_{i=1}^{I}(x_i - \bar{x})^2$ donc nous avons :

$$\begin{cases} \hat{\alpha} = \bar{y} - \left(\frac{Cov(x,y)}{Var(x)}\right)\bar{x} \\ \\ \hat{\beta} = \frac{Cov(x,y)}{Var(x)} \end{cases} \tag{3}$$

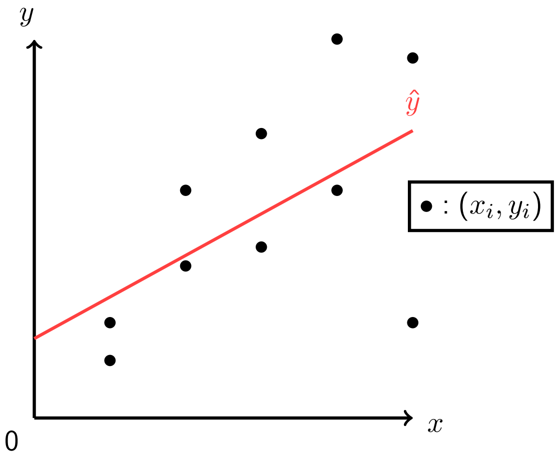
## Calculs

$\bar{x} = 3$, $\bar{y} = 2.7$, $Var(x) = 2.22$ et $Cov(x,y) = 1.22$ donc $\beta = 0.55$ et $\alpha = 1.05$.

Les calculs ont été effectués sur R.

$$y_i = 1.05 + 0.55 x_i + z_i \tag{4}$$

# La droite $\hat{y}$



# Alors ?

Conclusion : *A priori*, M. Favard a raison.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $\sum_{i=1}^{I} {z_i}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $z_i$ | -0.35 | 0.85 | 1.05 | 1.75 | -2.55 | -0.85 | -0.15 | -0.45 | -0.25 | 0.95 | 13.43 |

Table: Résidus de la droite $\hat{y}$

Néanmoins, pour effectuer une régression linéaire décemment, il faudrait tester plusieurs hypothèses :

- Tester l'hypothèse nulle de la constante du modèle (Student),
- Tester l'hypothèse nulle du coefficient du modèle (Student),
- Tester l'hypothèse significative du modèle (Fisher),
- Calculer le $R^2$ et le $R^2_{ADJ}$ et les analyser,
- Analyser les résidus et la somme des résidus au carré :
  - Tester la linéarité et l'autocorrélation (Durbin-Watson),
  - Tester la normalité (Shapiro-Wilk),
  - Tester l'homoscédasticité des variances (Harisson-McCabe)...