REGRESSION LINEAIRE UNIVARIÉE

$$Y = aX + b + \varepsilon$$

Estimateur des moindres carrés

$$\min_{a,b} \underbrace{\sum_{i=1}^{n} (y_i - a x_i - b)^2}_{J(a,b)}$$

On dérive $J(a,b)$

$$\begin{cases} \dfrac{\partial J}{\partial b}(a,b) = -2 \displaystyle\sum_{i=1}^{n} (y_i - a x_i - b) \\ \dfrac{\partial J}{\partial a}(a,b) = -2 \displaystyle\sum_{i=1}^{n} x_i (y_i - a x_i - b) \end{cases}$$

On doit alors résoudre un système linéaire de 2 équations à 2 inconnues.

$$a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i$$

$$a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

On note que $\sum_{i=1}^{n} 1 = n$, la première équation s'écrit donc aussi

$$a \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i}_{m_x} + b = \underbrace{\frac{1}{n} \sum_{i=1}^{n} y_i}_{m_y}$$

et on peut écrire la seconde

$$a \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i^2}_{C_{xx}} + b \underbrace{\frac{1}{n} \sum_{i} x_i}_{m_x} = \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i y_i}_{C_{xy}}$$

d'où

$$\begin{cases} a\, m_x + b = m_y & (1) \\ a\, C_{xx} + b\, m_x = C_{xy} & (2) \end{cases}$$

$(2) - M_x(1)$

$$\hookrightarrow \quad a\,(C_{xx} - M_x^2) = C_{xy} - M_x M_y$$

et $\quad a = \dfrac{C_{xy} - M_x M_y}{C_{xx} - M_x^2}$

(1) : $\quad b = M_y - a M_x$

Remarque : $C_{xx} - M_x^2$ est la variance de X

$C_{xy} - M_x M_y$ " " covariance de (X, Y)