Regression linéaire $Y = \alpha + \beta X$

modèle gaussien : $y_i = \hat{A} + \hat{B} x_i + \varepsilon_i$

$\alpha_{mes} \pm \Delta\alpha = ?$

$\beta_{mes} \pm \Delta\beta = ?$

Sources : - Cours de S. Le Digabel
- Cours de Ruch
- Protassov

On dispose de N couples (x_i, y_i) constituant un N-échantillon d'observations indépendantes.

On suppose que pour chaque observation/mesure on a :

$$\boxed{y_i = \hat{A} + \hat{B}\, x_i + \varepsilon_i}$$

où : les ε_i sont des réalisations indépendantes de $\varepsilon \sim \mathcal{N}(0, \sigma^2)$,

x_i de X,

y_i de Y,

et X et Y telles que : $E[Y \mid X] = \alpha + \beta X$.

$\hat{A}$ ($\sim$ moyenne m) est l'estimateur sans biais de α, c'est une V.A.

$\hat{B}$ " β, "

$\hat{Y}^*$ " $E[Y \mid X = x] = \alpha + \beta x$, "

On peut mq :

$$\boxed{\begin{aligned} &\hat{A} \sim \mathcal{N}\left(\alpha, \sigma^2\left(\frac{1}{N} + \frac{m_x^2}{S_{xx}}\right)\right) \\ &\hat{B} \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{S_{xx}}\right) \\ &\hat{Y}^* \sim \mathcal{N}\left(\alpha + \beta x, \sigma^2\left(\frac{1}{N} + \frac{(x - m_x)^2}{S_{xx}}\right)\right) \end{aligned}}$$

où
- $S_{xx} = \sum_{i=1}^{N} (x_i - m_x)$
- m_x est l'estimateur de la moyenne de x : $m_x = \frac{1}{N} \sum_{i=1}^{N} x_i$

et ainsi que :

$$\boxed{t_A = \frac{\hat{A} - \alpha}{S_{\hat{A}}} \sim T_{N-2} \quad \text{et} \quad t_B = \frac{\hat{B} - \beta}{S_{\hat{B}}} \sim T_{N-2}}$$

(T_{N-2} = loi de student à N-2 ddl)

où :

$$S_{\hat{A}} = \left[MS_\varepsilon\left(\frac{1}{N} + \frac{m_x^2}{S_{xx}}\right)\right]^{1/2} \quad \text{et} \quad S_{\hat{B}} = \left[\frac{MS_\varepsilon}{S_{xx}}\right]^{1/2}$$

où

$$MS_\varepsilon = \frac{SS_\varepsilon}{N-2} \quad \text{et} \quad SS_\varepsilon = \sum_{i=1}^{N} \varepsilon_i^2 \left(= \sum_{i=1}^{N} (y_i - y_i^*)^2\right)$$

On peut de plus mq: $MS_\varepsilon = \hat{\sigma}^2$ càd MS_ε est l'estimateur non biaisé de σ^2

et que:

$$S_{YY} = SS_\varepsilon + SS_R$$

où $SS_R := \sum_{i=1}^{N} (y_i^* - m_y)^2$, $\begin{cases} m_y = \hat{A} + \hat{B} m_x \\ y_i^* = \hat{A} + \hat{B} x_i \end{cases}$

$$\Rightarrow SS_R = \hat{B}^2 S_{xx}$$

or $\hat{B} = \dfrac{S_{xy}}{S_{xx}}$, où $S_{xy} = \sum_{i=1}^{N} (x_i - m_x)(y_i - m_y)$

$$= r \left(\frac{S_{YY}}{S_{xx}} \right)^{1/2} \quad \text{où} \quad r = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{S_{xY}}{(S_{xx} S_{YY})^{1/2}}$$

donc

$$SS_R = r^2 S_{YY}$$

d'où

$$S_{YY} = SS_\varepsilon + SS_R \Leftrightarrow \boxed{SS_\varepsilon = S_{YY}(1 - r^2)}$$

et donc:

$$\boxed{MS_\varepsilon = \frac{1 - r^2}{N-2} S_{YY}}$$

et ainsi:

$$\boxed{\begin{aligned} S_{\hat{A}} &= \left[\frac{1-r^2}{N-2} S_{YY} \left(\frac{1}{N} + \frac{m_x^2}{S_{xx}} \right) \right]^{1/2} \\ S_{\hat{B}} &= \left[\frac{1-r^2}{N-2} \cdot \frac{S_{YY}}{S_{xx}} \right]^{1/2} \end{aligned}}$$