Statistique
TD n°9

Ex 1

$$\min\left(\sum_{i=1}^{n}(Y_i - (b + a x_i))\right)^2 = \sum \varepsilon_i^2$$

Regression :

$Y = a_1 x_1 + \dots + a_m x_m + b + \varepsilon$

$m = 1$

$Y_i = b + a x_i + \varepsilon_i$

On observe $x_1 \dots x_n$

Estimer les paramètres $a, b \rightarrow \hat{a}, \hat{b}$

$\varepsilon \rightarrow \hat{\varepsilon}$

$\hat{Y} = \hat{b} + \hat{a} x$

Validité du modèle

$Y_i = b + a_1 X_{1i} + a_2 X_{2i} + \varepsilon_i$

$H_0 : a_i = 0 \ \forall i$

H_1 : il existe au moins un $a_i \neq 0$.

1) Test sur R^2 :

Si $F_{calculé} = \dfrac{n-m-1}{m} \times \dfrac{R^2}{1-R^2} > F_{\alpha, m, n-m-1}$

On rejette H_0 : il existe au moins un $a_i \neq 0$ et on valide le modèle proposé.

2) test des coef a_i ($H_0 : a_i = 0$, contre $a_i \neq 0$)

$$\frac{|\hat{a}_i|}{\sqrt{V(\hat{a}_i)}} \leq t_{\alpha, n-m-1}$$

écart-type

alors $\hat{a}_i$ est proche de 0.

Pour l'exercice 1 :

Y_t : prix moyen d'un kg de banane

X_t : prix du ticket de métro

$t = 1, \dots, 9$

$\quad 1975, \dots, 1983$

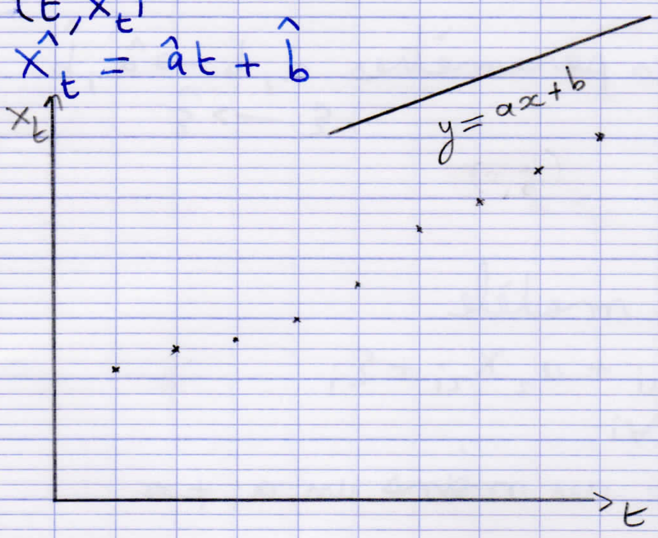
Modèle 1

$X_t = at + b + \varepsilon_t \quad (m = 1, n = 9)$

1) Nuage de points

(t, X_t)

$\hat{X}_t = \hat{a}t + \hat{b}$



$\Rightarrow$ on peut approximer par une droite : le modèle semble bon

2) Test de validité du modèle (1)

$$\frac{\hat{a}}{\sqrt{V(\hat{a})}} = \frac{0,505}{0,032} = 15,78125$$

à comparer avec $t_{0,05,7} = 1,833$

$\Rightarrow$ on rejette l'hypothèse $H_0 : a = 0$: le modèle proposé est significatif (il semble bon).

3) L'erreur commise pour l'année 79 $\rightarrow t = 5$

$\varepsilon_t = X_t - (at + b)$

On calcule une estimation de l'erreur

$\hat{\varepsilon}_t = X_t - (\hat{a}t + \hat{b})$

$\hat{\varepsilon}_t = x_5 - (5\hat{a} + \hat{b})$

$= 3,6 - (5 \times 0,505 + 1,364)$

$= 0,289$

4) Pour l'année 1985 $\rightarrow t = 11$

$\hat{x}_{11} = 11\hat{a} + \hat{b} = 11 \times 0,505 + 1,364 = 6,919 \simeq 6,75$

(5) Du point de vue mathématique c'est un modèle significatif.