

Examen OS:

1) Régression linéaire simple  
car il y a une seule variable explicative

$$2) \quad Y = 0,2268 X + 4,0213$$

3) Estimation :

$$(\text{Intercept}) : 4,0213$$

$$X : 0,2268$$

4) Lorsque on augmente une unité de la valeur  $X$   
 $Y$  augmente de 0,2268

$$5) \quad \begin{cases} H_0 : \beta_i = 0 & (\text{Pas de Signification}) \\ H_a : \beta_i \neq 0 & (\text{Significatif}) \end{cases}$$

6) ~~Donner~~ (Intercept) :

$$p\text{-value} = 0,796 > 0,05$$

$\Rightarrow H_0$  est acceptée

$\Rightarrow$  Pas de Signification

$$X : p\text{-value} = 5,81 \times 10^{-7} < 0,05$$

$\Rightarrow H_a$  est acceptée

$\Rightarrow$  il y a une Signification (le coeff est Significatif)

On a donc  $X$  a un effet Significatif sur  $Y$   
mais intercept (origine de la fct affine de régression linéaire)  
n'a pas d'effet significatif.



7)

$$\begin{cases} H_0 : \text{Tous les coeff sont nuls} \\ H_1 : \text{Au moins un coeff est non nul} \end{cases}$$

on a  $p\text{Value} = 0,00012 < 0,05$

$\Rightarrow H_1$  est acceptée

$\Rightarrow$  le modèle est globalement significatif  
il explique bien  $Y$ .

Ex 2 :

\* Test 1 :

$$\begin{cases} H_0 : A \sim N \\ H_1 : A \not\sim N \end{cases}$$

on a  $p\text{Value} = 0,27 > 0,05$

$\Rightarrow H_0$  acceptée

$\Rightarrow A$  suit la loi normale

\* Test 2 :

$$\begin{cases} H_0 : B \sim N \\ H_1 : B \not\sim N \end{cases}$$

on a  $p\text{Value} = 0,83 > 0,05$

$\Rightarrow H_0$  acceptée

$\Rightarrow B$  est la loi normale

\* Test 3 :

$$\begin{cases} H_0 : \sigma_A = \sigma_B \\ H_1 : \sigma_A \neq \sigma_B \end{cases}$$

on a  $p\text{Value} = 0,498 > 0,05$

$\Rightarrow H_0$  acceptée

$\Rightarrow$  les variances de A et B sont égales

• Test 4:

$$\begin{cases} H_0: \mu_A = \mu_B \\ H_a: \mu_A \neq \mu_B \end{cases}$$

$\therefore p\text{-value} = 0.9216 > 0.05$

$\Rightarrow H_0$  accepted

$\Rightarrow$  Les moyennes de A et B sont considérées égales

### Exercice 3 :

- 1) 153 observations , 6 Variables  
 ozone : int ; Solar.R : int ; Wind : num ,  
 Temp : int , Month : int , Day : int
- 2) Wind et ozone .
- 3) borne Supérieure =  $Q_3 + 1,5 (Q_3 - Q_1)$   
 borne Inférieure =  $Q_1 - 1,5 (Q_3 - Q_1)$   
 Les valeurs qui dépassent la borne Sup ou  
 qui sont en dessous de la borne inf  
 sont des valeurs aberrantes .
- 4) Generalized Imputation :
  - Remplacer les variables quantitative par { moyenne  
 médiane
  - Remplacer // qualitative par Mode  
 Mode c'est la valeur la plus fréquente

S                      Similar Case Imputation

Le plus proche  
voisin

↓

KNN

remplacer par une  
valeur aléatoire

↓

Random
- 5) Imputation par la médiane