TD 2
Les tests statistiques

Exercice 1

① $H_0 : a_1 = 0$ contre $H_1 : a_1 \neq 0$

$$t^* = \frac{\hat{a}_1 - 0}{\hat{\sigma}_{\hat{a}_1}} \sim t(T-k)$$

Si $|t^*| > t^{\alpha/2}_{(T-k)}$ alors on rejette H_0

Ici $t^* = \frac{0,572}{0,226} = 2,253$

$|t^*| > t^{\alpha/2}_{(10)} = 2,228$, donc on rejette H_0

La variable a_1 est significative (statistiquement $\neq 0$).

② $IC_{a_1} = [\hat{a}_1 - t^{\alpha/2}_{T-k} \times \hat{\sigma}_{\hat{a}_1} , \hat{a}_1 + t^{\alpha/2}_{T-k} \times \hat{\sigma}_{\hat{a}_1}]$

$= [0,07 ; 1,08]$

95 % de chance que la vraie valeur de a_1 soit entre [0,07 ; et 1,08.

③ $H_0 : a_1 = 0,5$ contre $H_1 : a_1 \neq 0,5$

$$t^* = \frac{\hat{a}_1 - 0,5}{\hat{\sigma}_{\hat{a}_1}} \sim t(T-k)$$

Si $|t^*| > t^{\alpha/2}_{(T-k)}$ alors on rejette H_0

$t^* = 0,32$

$$|t^*| < t^{\alpha/2}_{(10)} = 2,28,$$

Donc on ne peut pas rejeter H_0.

④ $SCT = \sum_{t=1}^{n} (y_t - \bar{y})^2 = 230,67$

$SCR = \sum e^2 = 140,79$

$SCE = \sum (\hat{y}_t - \bar{y})^2 = 89,8$

$SCT = SCE + SCR$

Coef de détermination

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 0,39$$

On explique 39 % des variations de y

La qualité du modèle d'ajustement n'est pas très satisfaisante. Elle mériterait d'être améliorée.

⑤ $\hat{y}_{13} = 30,20$

$\hat{y}_{14} = -11,02 + 0,572 \times 62 = 24,48$

Intervalle d'une prévision

$$y_t = \hat{y}_t \pm t^{\alpha/2}_{T-k} \times \hat{\sigma}_{e_t}$$

$$\hat{\sigma}_{e_{13}} = \hat{\sigma}^2_{\varepsilon} \sqrt{\frac{1}{T} + \frac{(x_{13}-\bar{x})^2}{\sum_{t=1}^{T}(x_t-\bar{x})^2} + 1}$$

$\hat{\sigma}^2_{e_{13}} = 17,02 \Rightarrow \hat{\sigma}_{e_{13}} = 4,13$

$\hat{\sigma}^2_{e_{14}} = 14,65 \Rightarrow \hat{\sigma}_{e_{14}} = 3,83$

Exercice 2

①

	SC	DL	CM
Modèle	SCE = 345,5	k-1 ici 2	SCE/(k-1) ici : 172,8
Résidu	SCR = 31,5	T-k ici : 7	SCR/(T-k) ici : 4,5
Total	SCT = 376,9	T-1 ici 9	

② Coef de déter : $\frac{SCE}{SCT} = 0,917 = R^2$

Ø. explique 92 % des variations de y.

$$\bar{R}^2 = 1 - \frac{SCR/(T-k)}{SCT/(T-1)} = 0,89$$

③ $H_0 : a_1 = a_2 = \dots = a_k = 0$

contre

$H_1 : \exists\, a_k \neq 0$

$$F^* = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} \sim F(k-1, T-k)$$

Si $F^* > F_{(k-1, T-k)}$ alors je rejette H_0

$F^* = 38,41$

$F^* > F_{(k-1, T-k)} = 4,74$ donc on rejette H_0. Le modèle est globalement significatif.

$\hat{y}_{11} = 35,14$

$\hat{y}_{12} = 22,01$

Intervalle $\quad y_t = \hat{y}_t \pm t_{T-k}^{\alpha/2} \times \hat{\sigma}_{et}$

Exercice 3 :

1. $R^2 = \dfrac{SCE}{SCT} = 0,328$

$\bar{R}^2 = 1 - \dfrac{SCR/(T-k)}{SCT/(T-1)} = 0,30$

$T = 267$

$k = 11$

Le $\bar{R}^2$ prend en compte le nb de variable et permet de comparer 2 modèles qui n'ont pas le même nb de var. Ici le modèle n'est pas satisfaisant.

2. Variance res ds le tableau :

$$Var_{res} = 7,64$$

3. $H0 : \beta_1 = \dots = \beta_{10} = 0$

vs $H1 : \exists\, \beta_i \neq 0$

Test de Fisher

$$F^* = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} \sim F(k-1, T-k)$$

④ Test de student

$H_0 : a_i = 0$ et $H1 : a_i \neq 0$

$$t^* = \frac{\hat{a}_i}{\hat{\sigma}_{\hat{a}_i}} \sim t(T-k)$$

. femme $t^* = |2,16| > 1,96$ On rejette H_0
variable significative.

. Bac es $t_* = |0,32| < 1,96$ On ne rejette pas

. CC $t^* = |7,40| > 1,96$ On rejette H_0

Explication : Lorsque l'on est une femme, on obtient par rapport aux hommes 0,79 pts de point en plus toutes choses étant égales par ailleurs

En moyenne les étudiants qui ont fait un bac S ont une note supérieur d'environ 1,69 point par rapport aux étudiants ayant fait un bac technologique ou es, toutes choses étant égales par ailleurs.

(5) $IC_{\text{réorientation}} = \hat{a}_{\text{réo}} \pm t_{T-k}^{\alpha/2} \times \hat{\sigma}_{\hat{a}_{\text{réo}}}$

$= 0{,}31 \pm 1{,}96 \times 0{,}48$

$= [-0{,}63 \, , \, 1{,}25]$

À 95 %, le coefficient de la variable réorientation est compris entre $-0{,}63$ et $1{,}25$

NB : $0 \in IC_{\text{réo}}$, la variable n'est donc pas significative.

(6) $\hat{y}_i = 4{,}5948 + 1{,}6852 + 0{,}3062 + 0{,}1509$
$= 6{,}74$

7.a Pour éviter un pb de multicolinéarité qui aurait rendu l'estimation par MCO impossible.

7.b. H_0 : $\beta_5 = \beta_6 = 0$ contre H_1 : $\beta_5 \neq 0$
ou $\beta_6 \neq 0$

Modèle général : Modèle (1)
Modèle contraint : Modèle (1) sans les variables de mention du bac.

nb contrainte

Sous H_0 : $F^* = \dfrac{(SCR_c - SCR_G)/c}{SCR_G/(T-k)} \sim F(c, T-k)$

si $F^* > F_{0,05}(c, T-k)$, on rejette H_0

$F^* = \dfrac{(2031,23 - 1955,98)/2}{1955,98/(267-11)} = 4,92$

$F^* > F_{0,05} = 3,00$, on rejette H_0

8.a. Pour mesurer l'écart de l'effet du cc entre les 2 campus, on a créé une variable croisée entre le cc et le campus.
Cette variable permet de mesurer l'écart de l'effet du cc entre Aix et Marseille

$aix\_controle\_continu = \begin{cases} 1 & \text{si Aix + CC} \\ 0 & \text{sinon} \end{cases}$

c. En moyenne le CC aura un impact supérieur de 0,94 pts pr les étudiants de Aix par rapport à ceux de Marseille

Eff CC Marseille 2,75. , pour Aix 2,75 + 0,94

3

1.e **Test de chow**

$H_0 : \beta^T = \beta^{Aix} = \beta^{Marseille}$ contre $H1 : \beta^{Aix} \neq \beta^{Marseille}$

Sous H_0 : $F^* = \dfrac{(SCR_T - (SCR_A + SCR_M))/k}{(SCR_A + SCR_M)/(N-2k)} \sim F(k, N-2k)$

$F^* > F_{0,05}(k, N-2k)$, on rejette H_0

$F^* = \dfrac{(1955,98 - (921,06 + 881,62))/11}{(921,06 + 881,62)/(267-22)}$

$= 1,89$

$F^* > F_{0,05} = 1,79$, on rejette H_0