Econométrie TD2

- Exercice 1

$n = 12$

$y_t = a_0 + a_1 x_t + \varepsilon_t \quad t = 1, \dots, n$

$\hat{y}_t = -11{,}017 + 0{,}572\, x_t$

$\hat{\sigma}_{\hat{a}_0} = 14{,}03$ et $\hat{\sigma}_{\hat{a}_1} = 0{,}226$

1) $\alpha = 5\%$ $\quad H_0 : a_1 = 0 \quad H_1 : a_1 \neq 0$

Statistique de test : $t^* = \dfrac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = \dfrac{0{,}572}{0{,}226} = 2{,}531 \sim t(T-k) = t(12-2) = t(10)$

Règle de décision : rejet de H_0 si $|t^*| > t^{\alpha/2}_{T-k} = t^{0{,}025}_{10} = 2.228$

On a $|t^*| = 2{,}531 > t^{\alpha/2}_{T-k} = 2.228$ donc on rejette H_0.

Cette hypothèse est importante à tester car elle sert à vérifier que la variable explicative a bien un impact significatif sur la variable à expliquer ✓ elle conditionne l'interprétation qui pourra être faite de l'effet de x sur y.

2) $\alpha = 5\%$

Intervalle de confiance : $IC = \hat{a}_1 \pm t^{\alpha/2}_{T-k} \times \hat{\sigma}_{\hat{a}_1} = 0{,}572 \pm \underset{2.228}{\cancel{2{,}531}} \times 0{,}226 = \cancel{[0{,}000 ; 1{,}144]}$ $[0{,}07 ; 1{,}08]$

3) $\alpha = 5\%$; $H_0 : a_1 = 0{,}5 \quad H_1 : a_1 \neq 0{,}5$

Statistique de test : $t^* = \dfrac{\hat{a}_1 - 0{,}5}{\hat{\sigma}_{\hat{a}_1}} = \dfrac{0{,}572 - 0{,}5}{0{,}226} = 0.3186 \sim t(10)$ ✓

Règle de décision : rejet de H_0 si $|t^*| > t^{0{,}025}_{10} = 2.228$ ✓

On a $|t^*| = 0.3186 < t^{0.025}_{10} = 2.228$ donc on ne rejette pas H_0. ✓

Ainsi, $a_1 = 0{,}5$ avec un seuil de confiance de 95%.

4) $SCT = \sum_{t=1}^{n} (y_t - \bar{y})^2 = 230{,}667$

$SCR = \sum_{t=1}^{n} e_t^2 = 140{,}799$

$SCE = \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2 = 89{,}741$

$R^2 = \dfrac{SCE}{SCT} = 0{,}389$ $\qquad \bar{R}^2 = 1 - \dfrac{SCR/(T-k)}{SCT/(T-1)} = 0{,}329$

Le modèle permet d'expliquer environ 39% des variations de y. Le modèle mériterait d'être amélioré par l'ajout de variables supplémentaires.

Le coefficient de détermination mesure la qualité de l'ajustement réalisée au moyen des estimateurs. Il s'agit du ratio de la variance expliquée sur la variance totale.

Le coefficient de détermination étant faible, le modèle explicatif n'est pas de qualité.

5) $x_{13} = 72$ et $x_{14} = 62$

$\hat{y}_{13} = -11{,}017 + 0{,}572\,(72) = 30{,}167$

$\hat{y}_{14} = -11{,}017 + 0{,}572\,(62) = 24{,}447$

Intervalles de confiance ?

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

$$\hat{\sigma}_{e_t}^{2} = \hat{\sigma}_{\varepsilon}^{2} \sqrt{\frac{1}{n} + \frac{(x_t - \bar{x})^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2} + 1}$$

On a $\hat{\sigma}_{\varepsilon}^2 = 14{,}079$ (séance préc.) et $t^{0{,}025}_{10} = 2.228$

$$\hat{\sigma}_{e_{13}}^{2} = 14{,}079 \sqrt{\frac{1}{12} + \frac{(72 - 61{,}8)^2}{274{,}29} + 1} = 17{,}02 \Rightarrow \sigma_{e_{13}} = 4{,}13$$

$$\hat{\sigma}_{e_{14}}^{2} = 14{,}079 \sqrt{\frac{1}{12} + \frac{(62 - 61{,}8)^2}{274{,}29} + 1} = 14{,}65 \Rightarrow e_{14} = 3{,}83$$

$CI\, y_{13} = \hat{y}_{13} \pm y^{0.025}_{10} \times \hat{\sigma}_{e_{13}} = 30{,}167 \pm 2.228 \times 4{,}13 = [21{,}00 ; 39{,}41]$

$CI\, y_{14} = [15{,}94 ; 33{,}02]$

- Exercice 2

1)

	SC	d.d.l.
Modèle	345,515 /	2
Résidus	31,469 /	7
Total	376,9 /	9

2) $R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 0{,}917$ /

Le modèle permet d'expliquer environ 92% des variations de y.

$\bar{R}^2 = 1 - \frac{SCR/(T-k)}{SCT/(T-1)} = 0{,}893$ /

3) Test de significativité globale $\alpha = 6\%$

$H_0 : a_1 = a_2 = 0$ $\quad H_1 : \exists a_i \neq 0,\ i = 1,2$

statistique de test : $F^* = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} = \cancel{8{,}563}\ 38{,}41 \sim F(2,7)$

Règle de décision : rejet de H_0 si $F^* > F(2,7) = 4{,}74$ /

On a $F^* = \cancel{8{,}563}\ 38{,}41 > F(2,7) = 4{,}74$ donc on rejette H_0. /

Le modèle est globalement significatif.

4) $x_{1,M} = 24$ $\quad x_{2,M} = 21$ $\quad \hat{y}_M = 35{,}142$ /

$x_{1,12} = 12$ $\quad x_{2,12} = 30$ $\quad \hat{y}_{12} = 22{,}014$ /

- Exercice 3

1) $R^2 = \frac{SCE}{SCT} = \frac{953{,}823086}{2909{,}8009} = 0.328$ /

$\bar{R}^2 = 1 - \frac{SCR/(T-k)}{SCT/(T-1)} = 1 - \frac{1955.97781/256}{2909.80091/266} = 0.302$ /

Le modèle permet d'expliquer environ 30% des variation de la moyenne générale. Il mériterait d'être amélioré par l'ajout de variables supplémentaires.

- Le coefficient de détermination non-ajusté le nombres de variables explicatives et augmentent mécaniquement avec ces dernières. Le coefficient ajusté quant à lui, prend en compte le nombre de variables explicatives et permet ainsi la comparaison de modèles de complexités différentes.

La qualité d'ajustement du modèle est faible car le coefficient de détermination est bas. /

2) ~~SCE~~ $S^2 = \frac{e'e}{T-k} = \frac{1955.97781}{267-11} = \frac{SCR}{T-k} = 7.64$ (déjà dans le tableau)

3) Test de significativité globale

$F^* = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} = 12.484$ /

$F(10,256) = 1.83 < F^* \Rightarrow$ rejet de H_0

Au seuil de 5%, $\exists B_i$ tq $B_i \neq 0$

Le modèle est globalement significatif.

4) $t^*_{femme} = \frac{0.7429867}{0.3428307} = 2.167$

$t^*_{bac.es} = \frac{0.1634863}{0.4989145} = 0.328$

$t^*_{cc} = \frac{3.369687}{0.4555728} = 1.535$

$t_{256}^{0.025} = 1.960 \Rightarrow H_0$ rejeté pour femme mais accepté pour bac.es et contrôle_continu.

Visiblement le contrôle continu n'est pas significatif pour les résultats.

5) $CI_r = \hat{a}_r \pm t_{256}^{0.025} \times \hat{\sigma}_{\hat{a}_r} = 0.3062411 \pm 1.960 \times 0.4828929 = [0.640 ; 1{,}253]$

L'intervalle de confidence est très étendu relativement à la valeur estimée du coefficient.

De plus, 0 appartient l'intervalle. On ne peut pas tirer de grandes conclusions si ce n'est la volatilité du coefficient et donc s'interrogé sur sa place dans le modèle.

TD 2

- Exercice 3

2) Estimation de la variance résiduelle :

$$s^2 = \frac{e'e}{T-k} = \frac{SCR}{T-k} = \frac{1955.97781}{256} = 7.64$$

4)

	t^*	$t_{256}^{0.025} = 1.960$
femme	2,167	H_1
baccalaureat_es	0,328	H_0
contrôle_continu	~~1,535~~ 7.397	~~H_0~~ H_1

D'après les résultats, la variable femme est significative. C'est-à-dire que le sexe est une variable explicative de la moyenne finale observée.
À l'inverse, le bac_es et le controle_continu ne sont pas des variables significatives.
La variable indicative bac_es a pour groupe de référence les élèves ayant fait un autre bac. Ainsi, on interprète qu'un bac_es ou un bac autre que es, s, et pro, n'a pas d'effet explicatif de la moyenne.
~~Finalement, ce test de significativité révèle que le controle continu n'a pas d'impact mesurable sur la moyenne~~.
→

5) $CI_r = \hat{a}_r \pm t_{256}^{0.025} \times \hat{\sigma}_{\hat{a}_r} = [-0,640\,;\,1,253]$
L'intervalle de confiance est très étendu. De plus, 0 appartient à l'intervalle donc le résultat n'est pas exploitable pour tirer des conclusions quant à la significativité de la variable dans le modèle. A priori, celle-ci n'est pas significative au seuil de 5%.

6) Moyenne $= \beta_0 + \beta_2 + \beta_8 + \beta_9 = 4.594816 + 1.685185 + 0.3062411 + 0.1504356$
$= 6.74$

7) a) La série du bac comporte 4 modalités - S, ES, pro, autre - donc seules 3 variables indicatrices ont été introduites par soucis de colinéarité parfaite.
Le groupe de référence choisi correspond aux individus ayant fait un bac autre donc on n'a pas introduit de variable indicatrice pour cette modalité. En d'autres mots, si un individu a fait un bac autre, on égalise les autres variables indicatrices à 0.

b) $SCR_2 = 2031.231$

4) Toutes choses étant égales par ailleurs, les femmes ont en moyenne une moyenne finale 0.743 supérieure par rapport aux hommes.
Toutes choses étant égales par ailleurs, les personnes ayant fait le contrôle continu ont en moyenne une moyenne finale supérieure de 3.370 par rapport à ceux qui ne l'ont pas fait.
Il n'y a pas de différence significative en terme de note finale obtenue entre les étudiants qui ont fait un bac ES et ceux qui ont fait un bac autre

7) b) Test de contraintes linéaires de Fisher : $H_0 : \beta_S = \beta_6 = 0 \quad H_1 : \beta_S \neq 0$ ou $\beta_6 \neq 0$

$$F^* = \frac{SCR_c - SCR_g}{c} \times \frac{T-k}{SCR_g} \sim F(c, T-k) \simeq 3$$

$$F^* = \frac{2031.231 - 1955.98}{2} \times \frac{256}{1955.98} = 4.92 > F(2,256)$$

$\Rightarrow$ on rejette H_0

8) a) On sait que le contrôle continue a un impact sur la moyenne finale.
On veut déterminer si cet impact est différent en fonction du campus.

controle_continu
- 1 aix_controle_continu
  - 1 cc à Aix
  - 0 cc non à Aix
- 0 aix_controle_continu
  - 1 impossible
  - 0 ps cc, ps à Aix

c) (La différence entre ceux qui ont suivi le cc à Aix et ceux qui n'ont pas suivi le cc est de 3.69)
TCEEPA, les étudiants qui suivent le contrôle_continu à Aix obtiennent une note finale supérieure d'environ 0.94 par rapport à ceux qui suivent le cc à Marseilles.
(La diff entre ceux qui ont suivi le cc à Marseille et ceux qui ne l'ont pas suivi est de 2.796).

9) a) Tester si l'ensemble des variables ~~significatives~~ explicatives ont des effets différents en fonction du campus.

c) Tester la stabilité des paramètres : test de Chow
$H_0 : \beta^T = \beta^A = \beta^M$ contre $H_1 : \beta^A \neq \beta^M$

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

Si $F^* > F_{0.05}(k, N-2k)$

$$F^* = \frac{(1955,98 - (921,06 + 881,62))/11}{(921,06+881,62)/(262-2\times 11)} = 1,89$$

$F_{0.05}(k, N-2k) = 1.79 \Rightarrow F^* > F_{0.05}(k, N-2k) \Rightarrow$ rejet de H_0

On rejette l'hypothèse de stabilité des paramètres.
Il y a effectivement, une différences significative entre les paramètres estimés par le sous-échantillon d'Aix et le sous-échantillon de Marseille.
Il y a donc un intérêt à estimer ces paramètres de manière distincte.

[On voit que la p-value associée au test est inférieure à 5% donc le modèle est globalement significatif

8) b) $R^2 = SCE/SCT = \frac{993.833494}{2909.8009} = 0.342$

$F^* = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} = \frac{R^2}{1-R^2} \cdot \frac{256}{11} = \cancel{12.543}\ 12.025$

$F(11, 255) = 12.02 < F^* \Rightarrow$ rejet de H_0
Le modèle est globalement significatif.

pas la peine