1

TD4 - Examen blanc

Exercice 1

① Pour éviter un pb de multicolinéarité parfaite qui aurait rendu l'estimation des paramètre impossible

② Variable 1 : Significative (p-value < 0,05)

Toutes choses étant égale par ailleurs les établissement qui proposent un service ambulatoire atteignent mieux leurs objectif en terme de prévention des infections.

Leur indicateur est supérieur d'environ 41,5% par rapport aux établissement qui n'ont pas de service ambulatoire.

Les établissements du Rhône-Alpes remplissent moins bien leurs objectifs que les autres régions. y compris (IDF et Paca)

③ $\beta_0 = 88,19 - 11,08 = 77,11$

$\beta_1 = 41,45$

$\beta_2 = -3,41 + 11,08 = 7,66$

$\beta_3 = -2,68 + 11,08 = 8,4$

$\beta_4 = 11,08$

Partie 2

① Les 3 ~~var~~ ajoutées correspondent au terme d'intéraction.
~~var~~ 2.1 = 1 losque rhone + ambulation.
3.1 = 1 — Pacca + ambulatoire

②. Il n'y a pas de difference de score selon les régions pour les établissements qui n'on pas de service ambulatoire

- Parmi les établissements qui sont dans la région centre, les etab qui ont un service ambulatoire remplissent mieux leur objectif que ceux qui n'en n'ont pas.
- Parmis les étab en IDF, les étab qui ont un service ambulatoire réussisse moins bien leur objectif que ceux qui n'en n'ont pas.
- Pour Paca, ceux qui ont un service ambulatoire remplissent mieux leur objectif que les autres.

③ Score moyen IDF Ambu : 48,5

④ Vrai pour ceux qui ont de l'ambulatoire (très rigoureusement on ne peut pas comparer IDF ou Rhone Alpes) faux pour les autres non ambulatoire. (pas d'effet région, p-value trop faible)

⑤ Du plus mauvais au meilleur.

- IDF - Ambu
- Rhone-Alpes - Ambu
- Pas de Ambu
- PACCA - Ambu
- Ambu et région centre.

Sous-réserve de significativité

2)

(6) On ajoute le nb de patient accueilli + var croisée entre le nombre de patient et la variable ambulatoire.

On fait un test de significativité pour les coefficients que l'on vient d'ajouter.

i.e. un test de student : $H_0 : a_0 = 0$ contre $H_1 : a_1 \neq 0$

stat de desc $t = \frac{a_1}{\hat{\sigma}_{\hat{a}_1}}$ | Rejet de H_0 si $|t^*| > t_{\alpha/2, T-k}$

Exercice 2

(1) $R^2 = 0,28$ $\bar{R}^2 = 0,279$

$\bar{R}^2$ prend en compte le nb de variable

(2) $\frac{SCR}{T-k} = 53,13$

(3) $H_0 : a_1 = \dots = a_k = 0$ contre $H_1 : \exists a_i \neq 0$

$$F = \frac{R^2}{1-R^2} \frac{T-k}{k-1} = 218,2 \geq 1,88$$

On rejette H_0.

Le modèle est globalement significative.

(4) Pour l'age : $t = 8,73 > 1,96$. On rejette H_0

Pour enfant : $t = 0,9 < 1,96$. On ne rejette pas H_0

③ Par 1 de revenu en plus, le salaire augmente de 0,00005

Toute chose étant égale par ailleurs. Le salaire horaire augmente par 1,76 € par niveau d'éducation supplémentaire.

. Pas différence entre ville 2 et ville 3.
Un individu de ville de type 1 a un salaire supérieur à 3,26 euros par rapport à un individu de ville 4.

. Les enfants n'ont pas d'incidence.

⑤ $\pm t \times \sigma = 0,137$.

⑥ $y =$ cons $+ 5 \times$ educ $+ 1900 \times$ heure $+ 35 \times$ age

⑦ H_0 : $\beta_4 = 1,5$ et $\beta_5 = \beta_7$

H_1 : $\beta_4 \neq 1,5$ ou $\beta_5 \neq \beta_7$

Modèle contraint.

Salaire $= \beta_1 + \beta_2$ revenu $+ \beta_3$ Age $+ 1,5$ Education $+ \beta_5$ Ville 1 $+ \beta_6$ Ville 3
$+ \beta_5$ Ville$_4$ $+ \beta_8$ Enfant 6 $+ \beta_9$ Heures $+ \beta_{10}$ sexe $+ u$

$\Rightarrow$ Z $= \beta_1 + \beta_2$ revenu $+ \beta_3$ Age $+ \beta_5 X + \beta_6$ Ville$_3$ $+ \beta_8$ Enfant
$+ \beta_9$ heures $+ \beta_{10}$ sexe $+ u$

Avec $\begin{cases} Z = \text{Salaire} - 1,5\ \text{Education} \\ X = \text{Ville 1} + \text{Ville 4} \end{cases}$

3)

a. $f = \dfrac{(SCR_c - SCR)/2}{SCR/(8T-k)} \sim F(2, 5048)$

$\overset{\shortparallel}{s^2}$

Si $f^* > F(2, 5043) = 3,000$ alors on rejette H_0

(b) $f^* = \dfrac{(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)/2}{s^2} \sim F(2, 5043)$

Avec le sujet, on rejette H_0 $(f = 3,21)$

Les 2 contraintes ne sont donc pas vérifiée.

(8) (a) L'introduction du sexe permet de déterminer si il existe une diff avec le sexe

Mais ne permet pas de mesurer si l'impact sur les autres var sont impactées.

(b) oui. On a l'effet de chaque variable en fonction du sexe

Mais il faut faire un test de chow (Stabilité des paramètres)

$H_0 = \beta_t = \beta_H = \beta_F$ contre $H_1 : \beta_H \neq \beta_F$

$f = \dfrac{[SCR_T - (SCR_H + SCR_F)]/k_T}{(SCR_H + SCR_F)/(N_H - k_H + N_F - k_F)} \sim f(k, N-2k)$

$f^* = 59,31 > F(9, 5035) = 1,88$. On rejette H_0 (l'hyp de stabilité)

il est pertinent de séparer homme et femme.

Exercice 3

(1) Éviter un pb de multicolinéarité.

(2) $12,68 + 2,78 = 15,48$ €.

(3) $6,72 - 2,78 \simeq 3,94$

(4) $\beta_0 = 12,68 + 2,78 \simeq 15,46$.

$\beta_1 = -2,78$

$\beta_2 = 3,94$

(5) On rajoute BAC_3 ou BAC_5.
+ Test des contrainte linéaire de Fisher.

(6)

On intègre sexe et possession diplome
+ Var croisée.
+ Test de significativité