1)

CLAVIER Mathieu

M1 Actuariat

26/11/2020

CC - Économétrie

Exercice 1

1) On propose le modèle suivant :

$$y_i = \begin{cases} 1 & \text{si l'enfant a été atteint d'anémie au cours des 12 derniers mois} \\ 0 & \text{sinon} \end{cases}$$

$z_i = x_i\beta + u_i$ où $u_i \sim \mathcal{N}(0,1)$

$$p_i = \mathbb{P}(y_i = 1) = \mathbb{P}(z_i > 0)$$

$$= \mathbb{P}(x_i\beta + u_i > 0)$$

$$= \mathbb{P}(u_i > -x_i\beta)$$

$$= 1 - \mathbb{P}(u_i < -x_i\beta)$$

$$= \Phi(x_i\beta)$$

Avec Φ la fonction de répartition de la loi normale

2) La fonction de vraisemblance est :

$$L(Y, Z, \beta) = \prod_{i=1}^{n} [\phi(x_i\beta)]^{Y_i} [1 - \phi(x_i\beta)]^{(1-Y_i)}$$

Ici on a $n = 3756$.

3) Interprétation du tableau :

Toutes choses égales par ailleurs :

- Les enfants bénéficiant d'un programme ont moins de chance de souffrir d'anemie
- L'augmentation de l'age diminue aussi les risques de souffrir d'anemie.
- Les hommes ont plus de risque de souffrir d'anémie.
- Toutes les autres variables sont non significatives car leur p-value est supérieur à 0,05.

4) $x_i\beta = -0{,}084 - 15 \times 0{,}159 + 565 \times 0{,}000004 + 0 \times 0{,}095$

$+ 3 \times 0{,}012 - 4 \times 0{,}039 + 0 \times 0{,}274$

$+ 0 \times 0{,}025 + 1 \times (-0{,}199)$

$= -2{,}786$

$\boxed{p_i = \Phi(x_i\beta) = 0{,}27\ \%}$

5) On calcule l'élasticité par rapport a une augmentation du revenu de 1%.

$$\frac{\partial \mathbb{P}(Y_i = 1)}{\partial Revenu_i} \times \frac{Revenu_i}{\widehat{\mathbb{P}(Y_i = 1)}}$$

Or : • $\dfrac{\partial \mathbb{P}(Y_i = 1)}{\partial Revenu_i} = \beta_{revenu} \times \underbrace{\phi(x_i\beta)}_{\frac{1}{\sqrt{2\pi}} e^{-(-2{,}786)^2/2}} = 3{,}29 \cdot 10^{-8}$

• $\dfrac{Revenu_i}{\widehat{P(Y_i = 1)}} = \dfrac{565}{0{,}0027}$

Donc $\boxed{Elasticité = 0{,}0069}$

Une augmentation de 1% du revenu entraine une augmentation du risque d'anémie de 0,0069 %.

6) a. Les lignes du tableau constituent aux valeurs prédites par le modèle.

Les colonnes sont les valeurs observée

Il y a 2 lignes et 2 colonnes (matrice 2x2) correspondant aux 2 cas 0 ou 1 (positif ou négatif à l'anémie).

b. Plus la matrice se rapproche d'une matrice diagonale et plus le modèle est bon.

Ici, ce n'est pas vraiment le cas. On constate notamment qu'il y a beaucoup de faux positifs. C'est à dire que l'on ne détecte pas les malades.

Lorsque le modèle indique un positif : il y a $\frac{310}{360+310} = 46\%$ de chance que ce ne soit pas le cas. Le modèle est légèrement meilleurs pour indiquer les personnes non malade (64,7% de personne bien prédites comme non malade).

Exercice 2

1) Pour la significativité globale, on réalise un test de Fisher :

$H0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ contre $H1 : \beta_i \neq 0$

Statistique de test :

$$F = \frac{R^2}{1-R^2} \frac{T-k}{k-1} \sim F(k-1, T-k) \quad \text{où} \begin{cases} k-1 = 16 \\ T-k = 35\,255 \end{cases}$$

Règle de décision :

Rejet de H0 si $|F^*| > F(k-1, T-k)$

Ici $|F^*| = 358,16 > 1,64$

On rejette donc H0.

$\boxed{\text{Le modèle est globalement significatif}}$

2) La formule de la variance résiduelle est :

$$\sigma_\varepsilon^2 = \frac{\sum_{i=1}^{n} e^2}{n-k} = \frac{SCR}{n-k} = \frac{499234,74}{35\ 255} = 14,16$$

3) L'intervalle de confiance est

$$\beta_{prestation} = \hat{\beta}_{prestation} \pm t_{T-k}^{\alpha/2} \times \hat{\sigma}_{\hat{\beta}_{prestation}} \qquad \left(\text{Ici } t_{T-k}^{\alpha/2} = 1,96\right)$$

Ici on a $[0,0465\ ;\ 0,795]$

4) Pour exploiter des variables catégorielles il faut réaliser des dichotomies. Il faut en réaliser (nb catégorie - 1) pour éviter un problème de multicolinéarité parfaite.

La catégorie non insérée en dichotomie sera la référence.

Exemple : L'individu de référence a 35 ans ou plus si age_1, age_2 et age_3 égale 0.

5) L'hypothèse est vérifiée par le modèle.
En effet, le nombre de semaine de congé parental diminue de 0,038 pour chaque euro de revenu supplémentaire.
(Toutes choses étant égales par ailleurs)

6) Toutes choses étant égales par ailleurs :

- le nb de semaine diminue de 0,0387 pour un euro supplémentaire de revenu
- le nb de semaine augmente de 0,06296 pour un euro de prestation supplémentaire
- le nb de semaine augmente de 3,77 si c'est une femme
- le nb de semaine diminue de -3,25 si la personne à moins de 25 ans par rapport à une personne de plus de 35ans.
- le nb de semaine diminue de -1,95 si la personne a entre 25 et 29 ans par rapport à une personne de plus de 35 ans
- ——— -0,485 ——— entre 29 et 34ans
  ———

8)

- il n'y a pas de différence significative entre le Nord et le sud ni entre l'Ouest et le sud

- il y un un impact de 0,22 à la hausse pour les personnes de l'Est par rapport à celles du sud

- Un exemple de variable croisée :

Les femmes agées de moins de 25 ans prennent plus de congé parental (augmentation de 3,04 nb de semaine) par rapport à un homme de plus de 35 ans.

9)

7) $18{,}11 + 1\times 3{,}77 - 1\times 3{,}25 + 1500\times(-0{,}03868) + (50 + 187{,}5)\times 0{,}068$

8) il s'agit

9) Test de Chow :

$H_0 : \beta_{n_1} = \beta_{n_2}$ contre $H1 : \beta_{n_1} \neq \beta_{n_2}$

Stat de test $F = \dfrac{[SCR_T - (SCR_{n_1} + SCR_{n_2})/k}{(SCR_{n_1} + SCR_{n_2})/(N-2k)} \sim F(k, N-2k)$

Rejet de H0 si $F^* > F_{0,05}(k, N-2k)$

Interprétation : Existe-t-il une différence significative entre les 2 coefficients