CC - Entrainement 1h30 cette année, 2h l'année dernière

- Exercice 1

Les thermes d'erreur $\sim \mathcal{N}(0,1)$

1. Calc. proba. p_i que l'enfant i ait souffert d'un pb d'anémie au cours des 12 derniers mois en fonction de ses caractéristiques (x_i) et la proba. qu'il n'en ait pas souffert.

$p_i = \mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i^* > 0) = \mathbb{P}(X_i\beta + \varepsilon > 0) = \mathbb{P}(\varepsilon > -X_i\beta) = 1 - \mathbb{P}(\varepsilon < -X_i\beta) = \cancel{1} - [\cancel{1} - \mathbb{P}(\varepsilon < X_i\beta)]$
$= \mathbb{P}(\varepsilon < X_i\beta) = \Phi(X_i\beta)$ où Φ est la fonction de répartition de la loi normale.
$1 - p_i = 1 - \Phi(X_i\beta)$

2. Ecrire la vraisemblance du modèle.

$L(Y,\beta) = \prod_{i=1}^{n} [p_i]^{y_i} [1-p_i]^{n-y_i}$

3. Interpréter les résultats

TCEEPA, un enfant est moins probable d'avoir souffert d'amenie au cours des 12 derniers mois si son ménage fait parti du programme.
TCEEPA, un enfant est de moins en moins probable de souffrir d'aménie au fur et à mesure qu'il grandit.
TCEEPA, un garçon a une proba. plus grande de souffrir d'aménie qu'une fille.
TCEEPA, le niveau de revenu du ménage, le nombre de personnes habitant le logement, le nb de pièces du logement, l'électricité et l'eau potable n'ont pas d'impact sur la proba qu'un enfant souffre d'aménie.

4. $\hat{p}_i$, fille, 15 ans, 565 pesos, 3 pers., 4 pièces, pas eau potable, pas électricité, programme

$X_i\hat{\beta} = -0.084 - 0.199 - 0.159(15) + 0.000004 \times 565 + 0.012(3) - 0.039(4) = -2.786$
$\hat{p}_i = F(X_i\hat{\beta}) = F(-2.786) = 0.00267$
Comment trouver la fonction de répartition de la loi normale? Normal^t tableau

5. Cette proba. est-elle fortement élastique au revenu du ménage. Justifier.

L'élasticité de $\hat{p}_i$ par rapport au revenu du ménage :
$\frac{\partial \mathbb{P}(Y_i=1)}{\partial rev_i} \times \frac{rev_i}{\mathbb{P}(\hat{Y}_i=1)} = \beta_{rev}\, \phi(X_i\hat{\beta}) \times \frac{rev_i}{\mathbb{P}(\hat{Y}_i=1)} = 0.000004 \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-2.786)^2} \times \frac{565}{0.00267} = 0.00697$

Soit pour une différence de 1% du revenu, la probabilité augmente de 0.00697.
La probabilité n'est donc pas fortement élastique au revenu du ménage.

6. Tableau confronte les prédictions aux Obs. réelles
Essayer de refaire après correction de l'Exo 6
(a) Comment a été construit ce tableau ?
(b) Quelle(s) conclusion(s) pouvez-vous en tirer ?

• Exercice 2

1. Tester si le modèle est globalement significatif.

Hypothèses : $H_0 : \beta_i = 0, \forall i = 1, .., 16$ $\quad H_1 : \exists \beta_i \neq 0, i = 1, \dots, 16$
Statistique de test : $F^* = \frac{R^2}{1-R^2} \cdot \frac{T-k}{k-1} = \frac{0.1378}{1-0.1378} \cdot \frac{35272-16}{16-1} = 375.650$

$F^* \sim F(k-1, T-k) = F(15, \infty) \Rightarrow F(15, \infty, 0.05) = 1.67$
Règle de décision : rejet de H_0 si $F^* > F(15, \infty)$
$\Rightarrow$ Le modèle est globalement significatif

2. Calc. la variance résiduelle
Vérifier que c'est juste
$SCR = \frac{499234.74}{35,255} = 14.161$ ✓

3. Calc. l'IC associé à la variable prestation et interpréter.

$[\hat{\beta}_i \pm t^{\alpha/2}_{(T-k)} \hat{\sigma}_{\hat{\beta}_i}]$
$[-0.439611 \pm 1.96 \times 0.0084198] = [-0.456 ; -0.423]$
TCEEPA, une augmentation de 1 € de l'allocation hebdomadaire fait diminuer le congé entre 0.456 et 0.423 semaines, càd. entre 3.192 et 2.961 jours.

4. Var. d'âge et de région n'ont pas été directement introduites. Expliquer la démarche et les conséquences pour l'interprétation des résultats.
Vérifier que tous les éléments sont là.
Les variables d'âge et de région telles que décrites dans l'énoncé sont des variables qualitatives. On ne peut pas les introduire directement dans le modèle. Il nous faut créer des variables dichotomiques pour chaque modalité et introduire celle-ci. Par souci de colinéarité, on exclu une des modalité qui sera considérée comme la référence. Ainsi, l'interprétation des résultats se fait toujours par rapport à la classe de référence.

5. Que pouvez-vous dire de l'hypothèse faite sur le salaire à partir des res.

Hypothèse : plus le salaire du parent est élevé, plus le coût d'opportunité du congé parental (perte de revenu) est élevé et donc plus le prestataire est incité à écourter la durée de son congé parental pour retrouver son niveau de revenu initial en retournant travailler.

* D'après les résultats, un salaire supérieur de 1€ fait diminuer le congé en moyenne de 0.0387 semaines pour un homme et 0.0105 semaines pour une femme. Ceci est en accord avec l'hypothèse.

6. Commenter littérairement l'ensemble des résultats du modèle.

TCEEPA, une allocation hebdomadaire de 1€ supérieur fait augmenter la durée du congé de 0.063 semaines en moyenne pour un homme et 0.019 semaines en moyenne pour une femme.

TCEEPA la durée du congé pour une femme est de 3.77 semaines de plus en moyenne que pour un homme

TCEEPA, un homme de moins de 25 ans, entre 25 et 29 ans, et entre 30 et 34 ans, prend un congé réduit de 3.25, 1.95, et 0.485 semaines de moins, en moyenne respectivement, qu'un parent homme plus de 35 ans

TCEEPA, un parent vivant dans la région Est prend en moyenne un congé de 0.228 semaines de plus qu'un parent vivant ailleurs.

TCEEPA, un parent qui prend un congé seul l'allonge en moyenne de 1.64 semaines par rapport à un parent dont le conjoint prend également un congé.

* TCEEPA, une femme de moins de 25 ans, entre 25 et 29 ans, et entre 30 et 34 ans, prend un congé réduit de 0.207, 0.133, et 0.0341 semaines de moins en moyenne, respectivement, qu'une femme de plus de 35 ans ans.

TCEEPA, le statut de salarié n'a pas d'impact sur la durée du congé.

7. mère, 25 ans, N, non salariée, conjoint en congé, rev = 1550 €, prestation = $\underbrace{50\% + 50\text{€}}_{243.75\text{€}}$

$$y_i = X_i\hat{\beta} = 18.11694 - 0.0386821 \times 1550 + 0.0629598 \times 243.75$$
$$- 1.951943 - 0.2869399 + 0.0281324 \times 1550 - 0.0439611 \times 243.75 + 1.819358$$
$$= 5.976$$

Une telle personne prendrait en moyenne un congé de 5.976 semaines.

8. $H_0 : \beta_{région\_1} = \beta_{région\_2}$

Formule du test qui a permis de calc. le F. Préciser à quoi correspondent les éléments de la formule. Conclusion sur l'hypothèse, interprétation du résultat.

Voir avec la prof.

$$F^* = \frac{(R\hat{\beta}-q)'[R(X'X)^{-1}R'](R\hat{\beta}-q)/2}{S^2} \sim F(c, T-k)$$

$R = [0$

9. Expliquer la mise en oeuvre du test de Chow dans le but de savoir si les déterminants de la durée du congé parental sont les mêmes pour les mères et pères (régression(s) effectuée(s), variable(s) explicative(s), statistique(s) de test, règles de décision). A votre avis quel va être le résultat de ce test ? Justifier

Voir avec la prof.

8. Test de contraintes linéaires de Fisher

$H_0 : \beta_{region\_1} = \beta_{region\_2}$ $\quad H_1 : \beta_{region\_1} \neq \beta_{region\_2}$

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

où SCR_g (resp. SCR_c) est la somme carré des résidus du modèle généralisé (resp. contraint)

Si $F^* > F$ on rejette H_0

Ici, on ne peut pas rejetter H_0 car la p-value est supérieure à 0.05

Ainsi, au seuil de 95% les coefficients de region 1 et region 2 ne sont pas différents.

9. Estimer 3 régressions différentes sur l'échantillon total, sur l'échantillon des mères et sur l'echantillon des pères.

Test de Chow

$H_0 : \beta_T = \beta_M = \beta_P$

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

$F^* > F \Rightarrow$ rejet de H_0

Si H_0 rejette, les déterminants de la durée du congé parental ~~est~~ sont différents pour les mères et les pères

A priori, H_0 serait rejeté car dans le modèle, les variables croisées avec le sexe sont significatives et les coefficients sont importants.