

Fiche 7 - TISD - Master Pro

Régression logistique

Tran Viet Chi, chi.tran@univ-lille1.fr, bureau 316 (bâtiment M3).

Exercice 1 (Décision de travail des femmes)

Les données de la table `activ.sas7bdat` sont issues de l'enquête emploi 2005 de l'INSEE et obtenu à partir de la base utilisée dans la fiche 3.

- La variable `ACT` vaut 1 lorsque la femme est active, 0 sinon.
- la variable `COHAB2` vaut 1 lorsque la femme ne vit pas en couple, 0 sinon,
- la variable `CSCJ` vaut 1 lorsque le chef de ménage est chef d'entreprise, cadre ou profession libérale, 0 sinon.
- la variable `EMPNBH` est le nombre d'heures travaillées dans une semaine.

Partie A Décision d'activité

1. On introduit une variable fictive et inobservable ACT^* que l'on appelle "besoin de travailler", telle que si $ACT^* > 0$ alors $ACT=1$ (la femme est active). On souhaite modéliser cette variable ACT^* comme fonction linéaire de l'âge (`AGE`), de fait de vivre en couple ou non (`COHAB2`), du nombre d'enfants de moins de 18 ans (`NBENF18`), du nombre de personnes dans le logement (`NBINDE`), de la CSP du chef de ménage (`CSCJ`), en supposant que les résidus suivent la loi logistique. Pour cela, utiliser la `PROC LOGISTIC` de SAS.

2. Commenter la valeur des paramètres. Tester la significativité du nombre d'enfants (`NBENF18`) dans cette décision.

3. Refaire la régression en supposant maintenant que les résidus suivent une loi normale. On utilisera la `PROC PROBIT`. Comparer les résultats avec les précédents.

Partie B Temps partiel

On peut considérer le choix d'un emploi à temps partiel comme une alternative entre le temps plein et l'inactivité. Nous nous intéressons maintenant à la modélisation du choix du nombre d'heures travaillées, que nous voyons comme une approximation du "besoin de travailler" introduit précédemment (le seuil n'est alors plus 0, mais un seuil propre à chaque femme). Le problème bien évidemment est que ce nombre d'heures travaillé n'est disponible que pour les femmes actives.

1. Estimer les coefficients de la régression linéaire du nombre d'heures travaillées (`EMPNBH`) sur les variables de la question 1 en n'utilisant que les observations pour lesquelles la variable `EMPNBH` est disponible.

2. Faire la régression Tobit de ce modèle et comparer avec la question 1. Pour cela, utiliser la `PROC LIFEREG`. Pour utiliser cette procédure, il faut créer deux variables qui indiquent s'il y a une censure à gauche et/ou à droite. Dans le cas où il n'y a pas de censure, la variable prend la valeur observée. S'il y a une censure, on remplace la valeur par `.`. Dans notre cas, le fait que l'on n'observe pas la durée de travail pour les femmes inactives est une censure à gauche. Il n'y a pas de censure à droite.

```
data malib.activ;
set malib.activ;
if act=1 then lower=empnbh; else lower=.;
if act=1 then upper=empnbh; else upper=empnbh;
run;
```

```
proc lifereg data=malib.activ;
model (lower,upper)=cohab2 age nbenf18 nbinde cscj;
run;
```

3. Comparer avec les résultats de la question **A. 1** et de la question **B. 1**.

Partie C Synthèse

Rédiger en quelques phrases une synthèse destinée à des non-statisticiens.