

Logiciel SAS

TD 3

À préparer pour le mardi 7 novembre Modélisation micro-économétrique

Le but de cette séance de TD est de vous initier à la modélisation micro-économétrique en ajustant une équation de salaire sur données individuelles.

1) Recherchez, sur le site internet de l'INSEE, le « fichier détail » relatif à l'enquête emploi en continu de 2012. Nous allons utiliser l'enquête de 2012, la dernière qui nous permette d'estimer une équation de salaire. En effet, l'INSÉÉ signale sur son site

Avertissement : à partir de 2013, l'enquête emploi a été renouvelée. À cette occasion, l'Insee a simplifié le fichier détail de l'enquête emploi téléchargeable sur le site, en réduisant le nombre de variables diffusées.

Lisez la documentation relative à l'enquête de 2012. Sur la page intitulée « Téléchargement des fichiers », on trouve les indications suivantes

Le fichier au format dBase concerne l'ensemble des individus répondants. Il contient 548 variables et 422 133 observations. Il est scindé en trois tables en raison du nombre important de variables.

et

Télécharger le fichier national au format dBase (format zip : 80 518 506 octets)

Quel est le format *dBase* que la documentation évoque ? Comment peut-on lire des tables au format *dBase* avec le logiciel SAS ? Ces indications parlent, d'une part, de trois tables et, d'autre part, d'un seul fichier à télécharger. Ces indications sont-elles ou non contradictoires ?

2) Téléchargez sur votre micro-ordinateur le fichier `eec12_indiv12_dbase.zip` (le fichier national au format *dBase*). Combien de fichiers ce fichier a-t-il archivé ? Extraire du fichier `eec12_indiv12_dbase.zip` les fichiers `indiv121.dbf`, `indiv122.dbf` et `indiv123.dbf`. Pour quelle raison ces fichiers ont-ils l'extension « `dbf` » ? Pour quelle raison le nom de ces fichiers commence-t-il par « `indiv` » ? Pour quelle raison le nom de ces fichiers comporte-t-il le nombre « `12` » ? Pour quelle raison le nom de ces fichiers se termine-t-il par le chiffre « `1` », « `2` » ou « `3` » ?

3) Recopiez ces trois fichiers dans le répertoire accessible par le serveur SAS. Protégez ces trois fichiers contre l'écriture.

4) Commentez le fragment de programme SAS suivant :

```
PROC import
  DATAFILE = "/folders/myfolders/indiv121.dbf"
  OUT = EE2012_1
  DBMS = DBF
  REPLACE ;
RUN ;
```

Quel est le rôle du mot-clé `DBMS` ? de l'indicateur `REPLACE` ? Quelle est le nom de la table SAS créée en sortie ? Combien de variables et d'observations cette table comporte-t-elle ?

5) Écrivez les PROC SAS qui permettent de lire les deux autres tables *dBase* `indiv122.dbf` et `indiv123.dbf`.

6) Les tables SAS `EE2012_1`, `EE2012_2` et `EE2012_3` comportent-elles le même nombre d'observations ?

7) Quel est le but du fragment de programme SAS suivant :

```
DATA EE2012 ;
  MERGE EE2012_1 EE2012_2 EE2012_3 ;
RUN ;
```

Cette façon de procéder est-elle sûre ?

Sur le site, on trouve l'explication suivante :

Pour pouvoir croiser des variables appartenant à des fichiers différents, l'appariement des 3 fichiers de données dBase se fait à l'aide des variables IDENT+NOI+TRIM, seul véritable identifiant d'un enregistrement. A cet effet, ces 3 variables sont présentes dans chacun des 3 fichiers. Aussi, les 3 variables de pondération sont présentes dans les 3 fichiers.

8) Donnez le sens des variables ident, noi et trim. Pourquoi ces variables sont-elles communes aux trois tables initiales ?

9) Quel est le but du fragment de programme SAS suivant :

```
DATA EE2012 ;
  MERGE EE2012_1 (IN=_i1) EE2012_2 (IN=_i2) EE2012_3 (IN=_i3) ;
  BY ident noi trim ;
  IF NOT (_i1 AND _i2 AND _i3) THEN ABORT ;
RUN ;
```

Quelle est la contrainte liée à l'utilisation de l'instruction BY avec l'instruction MERGE ? En quoi cette façon de procéder est-elle sûre ?

10) On s'intéresse aux variables suivantes :

Variable	Explication	Nature
naia	Année de naissance	CHAR
salred	Salaire mensuel net redressé des non réponses (y compris les primes mensualisées et redressées des non-réponses)	NUM
nbheur	Nombre d'heures correspondant au salaire déclaré	NUM
sexe	Sexe	CHAR
fordat	Année de fin d'études initiales	CHAR

Donnez les raisons (économiques) de l'intérêt porté à ces variables.

11) On veut ajuster l'équation de salaire

$$\ln \left(\frac{S_i}{H_i} \right) = a df_i + b \text{âge}_i + c \text{âge}_i^2 + d f_i + e + u_i$$

où S_i est le salaire mensuel de l'individu i exprimé en euros, H_i le nombre d'heures travaillées dans le mois, df_i la durée de la formation initiale exprimée en année, âge_i l'âge exprimé en année, âge_i^2 l'âge au carré, f_i une indicatrice repérant les femmes et u_i le terme d'erreurs. Pourquoi la variable à expliquer est-elle en logarithme ? Quelle valeur attend-on pour a et pour b . Quels signes attend-on pour les coefficients c et d ?

12) Quel est le but du fragment de programme SAS suivant :

```
LIBNAME MaBib "/folders/myfolders" ;
DATA MaBib.EqSal ;
  SET EE2012 ;
  IF (salred NE .) AND (nbheur NE .) ;
RUN ;
```

13) Quel est le but du fragment de programme SAS suivant :

```
PROC freq DATA = MaBib.EqSal ;  
  TABLE naia fordats sexe ;  
RUN ;
```

14) Quel est le but du fragment de programme SAS suivant :

```
DATA MaTableDeTravail ;  
  SET MaBib.EqSal (KEEP = salred nbheur naia fordats sexe) ;  
  lsh = LOG(salred / nbheur) ;  
  age = 2012 - naia ;  
  age2 = age**2 ;  
  dfi_mqt = (fordats EQ '1') ;  
  IF dfi_mqt THEN dfi = 0 ; ELSE dfi = fordats - naia ;  
  femme = (sexe EQ '2') ;  
RUN ;
```

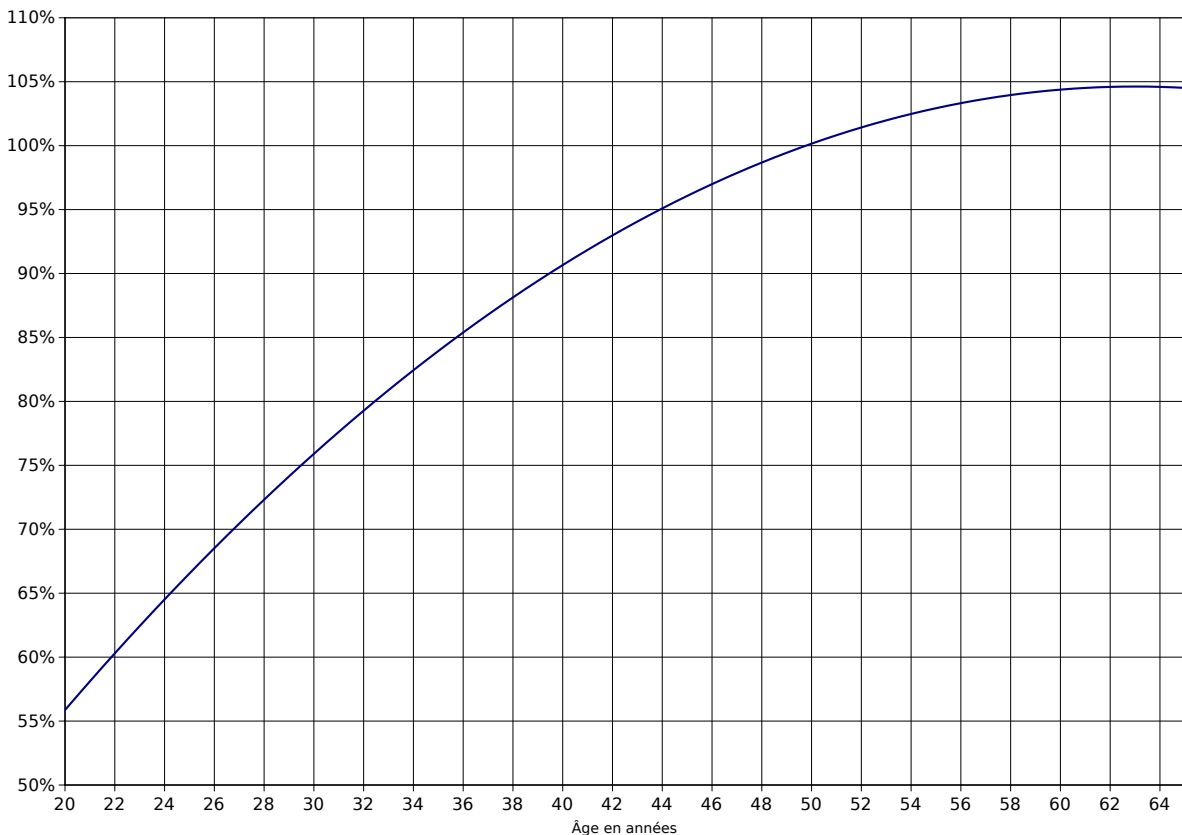
Expliquez notamment la façon dont la variable `dfi` est construite. Pourquoi la variable `femme`, qui est la variable indicatrice des femmes, est-elle calculée de cette manière ?

15) Quel est le but du fragment de programme SAS suivant :

```
PROC reg DATA = MaTableDeTravail OUTEST = OutEst ;  
  MODEL lsh = dfi age age2 femme dfi_mqt ;  
RUN ;
```

Expliquez notamment le rôle de la variable `dfi_mqt`. On trouve que le coefficient estimé de `dfi` est égal à 0,047 ; cette valeur est-elle sur-estimée ou sous-estimée ? On trouve que le coefficient estimé de `femme` est égal à -0,15 ; cette valeur est-elle sur-estimée ou sous-estimée ? Est-il possible d'interpréter directement les coefficients de `age` et de `age2` ?

FIGURE 1 : Effet relatif de l'âge sur le taux de salaire horaire en %



16) Comment peut-on obtenir le graphique de la figure 1 ?