



UNIVERSITÉ DE NANTES

---

**Master 1 EKAP - Travaux dirigés d'introduction au logiciel R - TD n°5**

---

Responsable du cours : Camille Aït-Youcef

Année : 2020 - 2021

## **Exercice 7 : R et les régressions linéaires multiples**

L'objectif de cet exercice est d'utiliser les commandes permettant de régresser des variables en fonction de plusieurs autres variables mais aussi d'interpréter les sorties de résultats du logiciel R et de tester les estimateurs du modèle. La base de données qui est exploitée regroupe des données financières. Elle correspond à la base de données de l'exercice 5 de la fiche de TD 3.

### **Description de la base de données :**

**ident** : Identifiant

**gender** : sexe

**fsize** : taille de la famille

**marr** : statut marital ("1" si marié, "0" sinon)

**age** : âge

**inc** : revenu familial annuel

**nettfa** : patrimoine financier net

**pira** : participation à certains régimes de retraite

**e401k** : éligibilité au système d'épargne par capitalisation

**p401k** participation au système d'épargne par capitalisation

Les variables de richesse et de revenu sont exprimées en milliers de dollars.

### Questions

1. Importer la base de données *completset.csv* dans Rstudio et nommer la base de données *completset*.
2. Créer les variables *age2* et *inc2* qui correspondent aux variables *age* et *inc* au carré. Ajouter ces variables au dataframe *completset*.

3. Soit le modèle suivant (en utilisant les données de la base *completset*) :

$$nettf_a = \beta_0 + \beta_1 inc + \beta_2 age + u$$

Que pouvez vous dire concernant la significativité individuelle et la significativité globale des paramètres ?

4. Evaluer l'hypothèse de normalité des résidus avec des graphiques (normal Q-Q plot et histogramme) et appliquer le test de normalité de Jarque Bera (pour appliquer le test vous devez télécharger la librairie *normtest*).
5. Créer la variable *nettfadev* exprimant le patrimoine financier net en dollars puis estimer le modèle suivant :

$$nettfadev = \beta_0 + \beta_1 inc + \beta_2 age + u$$

6. Expliquer l'évolution des coefficients par rapport aux coefficients estimés à la question 3.
7. Soit le modèle suivant :

$$nettf_a = \beta_0 + \beta_1 inc + \beta_2 age + \beta_3 age2 + \beta_4 e401k + u$$

Interpréter les sorties du logiciel.

8. La relation entre le niveau de patrimoine financier et les variables explicatives varie-t-elle entre les ménages d'une seule personne et les autres types de ménage ? Quel test devez vous utiliser pour répondre à la question. La fonction à utiliser se trouve dans la librairie `strucchange`.

9. Créer le dataframe dont les observations ne concernent que les ménages composés d'une seule personne. A partir de cette base de données estimer le modèle suivant :

$$netffa = \beta_0 + \beta_1 inc + \beta_2 age + \beta_3 age2 + \beta_4 e401k + u$$

Les erreurs du modèle sont-elles hétéroscédastiques ? Vous appliquerez le test de Breuch-Pagan pour répondre à la question. Par ailleurs, il est nécessaire d'installer la librairie `lmtest` en amont pour répondre à la question.