



UNIVERSITÉ DE NANTES

Master 1 EKAP - Travaux dirigés d'introduction au logiciel R - TD n°4

Responsable du cours : Camille Aït-Youcef

Année : 2020 - 2021

Exercice 6 : 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. 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. Présentez un tableau montrant la répartition des individus selon la possession d'un compte épargne retraite (*pira*), puis la répartition selon, à la fois, la possession d'un compte épargne retraite et le sexe.
3. Représentez la distribution des âges des individus enquêtés avec un histogramme par tranche de 10 ans.
4. Caractérisez la distribution (médiane, moyenne, écart-type, variance) du revenu en utilisant la fonction *summary*
5. Combien de ménages d'une personne sont présents dans la base de données ?
6. Créez une nouvelle base de données permanente *onepers* constituée uniquement de ménages d'une personne.
7. Soit le modèle suivant : $nettfa = \beta_0 + \beta_1 inc + \beta_2 age + \beta_3 homme + \beta_4 femme + u$
Utilisez la méthode des MCO pour estimer ce modèle pour les ménages d'une seule personne.
Que constate-t-on ?
8. Soit le modèle suivant : $nettfa = \beta_0 + \beta_1 inc + \beta_2 age + \beta_3 homme + u$
Interprétez vos résultats.
9. Soit le modèle suivant : $nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u$
Interprétez vos résultats

10. Représentez un nuage de points avec *age* en abscisses et *inc* en ordonnées. Que constatez-vous ?
11. Calculez la corrélation entre les deux variables en utilisant la commande *cor.test*
12. Si vous faites une régression simple de *nettf* sur *inc*, trouvez-vous un coefficient estimé pour la variable *inc* très différent de l'estimation réalisée précédemment ? Expliquez votre résultat.