



UNIVERSITÉ DE NANTES

Master 1 EKAP - Travaux dirigés d'introduction au logiciel R - TD n°6

Responsable du cours : Camille Aït-Youcef

Année : 2020 - 2021

Exercice 8 : R et les régressions linéaires simple et 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 et de tester la validité du modèle. La base de données qui est exploitée provient de la librairie Wooldridge.

Description de la base de données :

salary : salaire des PDG en millier de \$

age : âge

college : 1 si l'individu est diplômé du supérieur

grad : 1 si l'individu a fini le lycée

comten : nombre d'année au sein de la société

ceoten : nombre d'année en tant que PDG dans la société

sales : chiffre d'affaire en millions

profits : profit en millions

mktval : valeur de marché de l'entreprise en million

lsalary : $\log(\text{salary})$

lsales : $\log(\text{sales})$

lmktval : $\log(\text{mktval})$

comtensq : comten^2

ceotensq : ceoten^2

profmarg : profits en pourcentage du chiffre d'affaire

Questions

1. Télécharger la librairie Wooldridge puis charger la base de données ceosal2 en tapant la commande suivante : `data('ceosal2')`.

2. Inspecter la base de données

3. Soit le modèle suivant (en utilisant les données ceosal2) :

$$lsal = \beta_0 + \beta_1 lsales + u$$

Interpréter les sorties du logiciel .

4. Soit le modèle suivant :

$$lsal = \beta_0 + \beta_1 lsales + \beta_2 lmktval + u$$

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

5. Pourquoi le paramètre β_1 a-t-il changé ?

6. Calculer le coefficient de corrélation entre lsales et lmktval. Qu'en déduisez-vous ?

7. Soit le modèle suivant :

$$lsal = \beta_0 + \beta_1 lsales + \beta_2 lmktval + \beta_3 profits + u$$

Pourquoi ne doit-on pas ajouter profits en log ? Que représente β_3 ?

8. Soit le modèle suivant :

$$lsal = \beta_0 + \beta_1 lsales + \beta_2 lmktval + \beta_3 ceoten + u$$

Interpréter les sorties du logiciel.

9. Soit le modèle suivant :

$$lsal = \beta_0 + \beta_1 lsales + \beta_2 lmktval + \beta_3 ceoten + \beta_4 ceoten^2 + u$$

Interpréter les sorties du logiciel.

10. Selon vous, quel est le “meilleur” modèle ? pour répondre à la questions vous utiliserez les critères suivants : BIC, AIC et R^2 ajusté.
11. En utilisant la fonction `stargazer()` contenue dans le package du même nom, afficher et exporter en sortie texte les résultats des régressions des modèle 4 et 5.
12. Evaluer l’hypothèse de normalité des résidus avec des graphiques (normal Q-Q plot et histogramme) et appliquer les tests de normalité de Jarque Bera (pour appliquer le test vous devez télécharger la librairie `normtest`) et le test de Shapiro (pour appliquer le test vous devez télécharger la librairie `lmtest`).
13. La propriété d’homoscédasticité est-elle vérifiée ? Vous utiliserez le test de Breusch-Pagan, de White (pour le test de White il est nécessaire de télécharger la librairie `skedastic` et d’utiliser la fonction `white_lm` ou de préciser un argument dans la fonction `bgtest` de la librairie `lmtest`) et de Golfeld-Quandt pour répondre à la question
14. Les résidus sont-ils autocorrélés ? Pour répondre à la question vous effectuerez un test de Durbin-Watson et un test de Breusch-Godfrey.