

Les points sont proportionnels à la difficulté des questions. Justifiez toutes vos réponses et étapes du raisonnement. Tous les tests sont demandés à 5% sauf si spécifié dans la question.

Partie 1. Régression

1. Lire le fichier *mroz.txt*. Ne sélectionner que les observations pour lesquelles la variable *wage* est strictement positive.
2. Faire les statistiques descriptives du salaire, de l'âge et de l'éducation pour l'ensemble des femmes puis, pour les femmes dont le salaire du mari est supérieure à la médiane de l'échantillon, puis pour les femmes dont le salaire du mari est inférieur à la médiane de l'échantillon
3. Faire l'histogramme de la variable *wage*. Calculer le log de *wage* et faire l'histogramme. Comparez les deux histogrammes et commentez
4. Calculer les corrélations *motheduc* et *fatheduc*. Commentez. Il y a-t-il un problème de multicollinéarité si l'on utilise ces variables comme variables explicatives ?
5. Faites un graphique en nuage de point entre *wage* et *educ*, *wage* et *exper*, *wage* et *fatheduc*. Commentez. S'agit-il d'un effet "toute chose étant égale par ailleurs ?"
6. Quelle est l'hypothèse fondamentale qui garantit des estimateurs non biaisés ? Expliquer le biais de variable omise.
7. Faire la régression de *wage* en utilisant les variables explicatives une constante, *city*, *educ*, *exper*, *nwifeinc*, *kidslt6*, *kidsgt6*. Commentez l'histogramme des résidus.
8. Faire la régression de *lwage* sur une constante, *city*, *educ*, *exper*, *nwifeinc*, *kidslt6*, *kidsgt6*. Comparer l'histogramme obtenu à celui de la question 7.
9. Tester l'hypothèse de non significativité de *nwifeinc* avec un seuil de significativité de 1%, 5% et 10% (test alternatif des deux côtés). Commentez les p-values.
10. Tester l'hypothèse que le coefficient associé à *nwifeinc* est égal à 0.01 avec un seuil de significativité de 5% (test à alternatif des deux côtés)
11. Tester l'hypothèse jointe que le coefficient de *nwifeinc* est égal à 0.01 et que celui de *city* est égal à 0.05.
12. Faites une représentation graphique de la manière dont le salaire augmente avec l'éducation et l'expérience professionnelle. Commentez

13. Tester l'égalité des coefficients associés aux variables *kidsgt6* et *kidslt6*. Interprétez.
14. En utilisant le modèle de la question 7, faire le test d'hétéroscédasticité de forme linéaire en donnant la p-valeur. Déterminer la ou les sources d'hétéroscédasticité et corriger avec les méthodes vues en cours. Comparer les écarts-types des coefficients estimés avec ceux obtenus à la question 7. Commenter.
15. Tester le changement de structure de la question 8 entre les femmes qui ont plus de 43 ans et les autres : test sur l'ensemble des coefficients. Refaire le test avec 3 groupes (mutuellement exclusifs) : les femmes de moins de 30 ans, entre 30 et 43 ans, plus de 43 ans. Donnez les p-valeurs
16. Construire les variables binaires correspondant à l'âge des femmes de la question 15. Refaire la question 8 en ajoutant ces variables et en utilisant comme référence les femmes qui ont moins de 30 ans. Interprétez les paramètres associés aux variables binaires. Faire le test de non significativité de l'ensemble des variables binaires. Donnez les p-valeurs.