

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 garanti 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 *exper* 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 *educ* est égal à 10% avec un seuil de significativité de 5% (test à alternatif des deux côtés)
11. Tester l'hypothèse jointe que le rendement de l'éducation est de 10% et que celui de l'expérience professionnelle est de 5%.
12. De combien augmente le salaire en pourcentage avec 10 années d'expérience ?
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. Corriger le problème par rapport à la variable la plus importante en utilisant la méthode des MCG. 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 moins de 30 ans, entre 30 et 43 ans, plus de 43 ans (3 groupes mutuellement exclusifs). Donnez les p-valeurs.
16. A partir de la variable *kidslt6*, créer un ensemble de variables binaires pour le nombre d'enfants de moins de 6 ans. Refaire la question 8 avec ces variables et en utilisant comme référence les femmes qui ont des enfants de plus de 6 ans. Ces catégories sont-elles mutuellement exclusives ? Interprétez les paramètres associés aux variables binaires. Faire le test de non significativité de l'ensemble des variables binaires. Donnez les p-valeurs
17. A partir de l'échantillon global, faire une régression de *inlf* sur une constante, *city*, *educ*, *age*, *kidslt6*, *kidsgt6*. Interprétez les coefficients estimés.
18. Estimer le modèle probit de *inlf* sur une constante, *city*, *educ*, *age*, *kidslt6*, *kidsgt6*. Faire le test de non significativité jointes des coefficients associés à *kidslt6* et à *kidsgt6*. Comparez le résultat du test à celui de la question 18.
19. Calculer les effets partiels pour l'ensemble des variables explicatives :  $\partial p(y=1)/\partial x_k$  ( $k = 1, \dots, K$ ), où  $K$  est le nombre de variables explicatives. Comparer vos résultats à ceux obtenus à la question 17. Commentez.
20. Faire le test de non significativité jointes des coefficients associés à *kidslt6* et à *kidsgt6* en utilisant la méthode du rapport de vraisemblance. Comparez aux résultats de la question 18.