UNIVERSITÉ PARIS EST-CRÉTEIL

UNITE DE FORMATION ET DE RECHERCHE

MASTER EN STATISTIQUES ET DE L'ECONOMETRIES POUR LA RECHERCHE, L'ANALYSE ET LE TRAITEMENT DE L'INFORMATTION

# M1 - SUJET 2 – LOGICIEL R


DEVOIR DE

BENGONE AKOU LAJOIE

SOUS LA SUPERVISION DE :

EMMANUEL DUGUET

PROFESSEUR CHERCHEUR

2024-2025

# SOMMAIRE

## TABLE DES ILLUSTRATIONS

## METADONNEES

Variables Présentes dans la base de données mroz

| Notation | Description | Unité |
|---|---|---|
| inlf | Indique si la femme est active en 1975 (=1 si oui, =0 sinon) | Binaire (0/1) |
| hours | Nombre d'heures travaillées par la femme en 1975 | Heures |
| kidslt6 | Nombre d'enfants de moins de 6 ans dans le foyer | Nombre d'enfants |
| kidsge6 | Nombre d'enfants âgés de 6 à 18 ans dans le foyer | Nombre d'enfants |
| age | Âge de la femme | Années |
| educ | Nombre d'années de scolarité complétées par la femme | Années |
| wage | Salaire horaire estimé de la femme | Dollars/heure |
| repwage | Salaire horaire rapporté lors de l'entretien en 1976 | Dollars/heure |
| hushrs | Nombre d'heures travaillées par le mari en 1975 | Heures |
| husage | Âge du mari | Années |
| huseduc | Nombre d'années de scolarité complétées par le mari | Années |
| huswage | Salaire horaire du mari en 1975 | Dollars/heure |
| faminc | Revenu total de la famille en 1975 | Dollars |
| mtr | Taux marginal d'imposition fédéral appliqué à la femme | Pourcentage (%) |
| motheduc | Nombre d'années de scolarité complétées par la mère | Années |
| fatheduc | Nombre d'années de scolarité complétées par le père | Années |
| unem | Taux de chômage dans le comté de résidence | Pourcentage (%) |
| city | Indique si la femme vit dans une zone urbaine (=1 si oui, =0 sinon) | Binaire (0/1) |
| exper | Expérience réelle de la femme sur le marché du travail | Années |
| nwifeinc | Revenu familial hors contribution de la femme | Milliers de dollars |
| lwage | Logarithme naturel du salaire horaire | Logarithme (sans unité) |
| expersq | Carré de l'expérience sur le marché du travail | Années au carré |

*Tableau 1: Métadonnées*

## REPONSES AUX QUESTIONS

Bien que la base de données mroz contienne au total 22 variables comme présentées dans les métadonnées, cette partie s'intéressera uniquement à 4 d'entre elles : la variable d'intérêt (« lwage »), la variable de traitement (« kidslt6 ») et les variables explicatives (« exper » et « educ ») .

L'objectif de cette analyse est d'évaluer l'impact de la présence des enfants en bas âge sur les rémunérations des femmes. C'est dans cette optique, que nous allons effectuer des statistiques descriptives, utiliser des tests paramétriques et non paramétriques pour comparer les populations ou encore développer un modèle linéaire de Rubin.

### 1.1.1 Création d'une variable de traitement nommée <<d>> et vérification des instructions données

Au regard des fréquences des modalités de la variable d'intérêt (« kidslt6 »), mon choix s'est porté sur la création d'une variable de traitement binaire qui prend la modalité (d=0) si la femme n'a aucun enfant de moins de 6 ans et (d=1), si elle a au moins un enfant âgé d'au moins 6 ans. Cette dernière aura comme avantage de constituer deux échantillons suffisamment représentatifs pour inférer les résultats obtenus à partir des échantillons.

| **Nombre d'enfant de moins de 6 ans** | **(d=0)** | **(d=1)** | **(d=2)** | **(d=3)** |
|---|---|---|---|---|
| **Nombre d'observation** | 606 | 118 | 26 | 3 |
| **Pourcentage** | 80.48 | 15.67 | 3.45 | 0.40 |

*Tableau 2: Fréquence du nombre d'enfants de moins de 6 ans chez les femmes*

### 1.1.2 Statistiques descriptives et comparaison entre les populations selon la variable de traitement

Les statistiques descriptives et les tests statistiques ont été effectués dans le cadre de cette partie. Les statistiques descriptives ont servi à fournir un aperçu de la base de données globale et des sous base de données en présentant les différentes valeurs des paramètres de tendance centrale et de dispersion. Les tests statistiques quant à eux, ont été utilisés pour dresser une comparaison exhaustive entre les 2 populations concernées, c'est-à-dire entre les femmes ayant au moins un enfant âgé de moins de 6 ans et celles n'ayant pas.

#### 1.1.2.1 STATISTIQUES DESCRIPTIVES

##### *1.1.2.1.1 Résultats de la base de données globale*

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| **wage** | 428 | 4.18 | 3.31 | 0.13 | 25.00 |
| **educ** | 753 | 12.29 | 2.28 | 5 | 17 |
| **exper** | 753 | 10.63 | 8.07 | 0 | 45 |
| **d** | 753 | 0.20 | 0.40 | 0 | 1 |

*Tableau 3 : Statistique descriptive de la base de données intégrale*

##### *1.1.2.1.2 Analyses de la base de données globale*

Comme on peut le constater, la base de données « mroz » contient au total 753 observations. Cependant, on observe un nombre important de données manquantes (325) sur certaines variables à l'instar de « wage » donc « lwage».

Aussi, au regard des écarts-types et des valeurs des paramètres min et max, on peut constater que les données sont très hétérogènes. C'est notamment le cas pour les variables « educ » et « exper » où les coefficients de variation valent respectivement 80.18% et 75.91%.

##### *1.1.2.1.3 Résultats de la population n'ayant pas d'enfant (« d=0 »)*

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| **wage** | 375 | 4.14 | 3.11 | 0.16 | 25.00 |
| **educ** | 606 | 12.19 | 2.28 | 5 | 17 |
| **exper** | 606 | 11.50 | 8.39 | 0 | 45 |
| **d** | 606 | 0.00 | 0.00 | 0 | 0 |

*Tableau 4 : Statistiques descriptives du groupe m0*

##### *1.1.2.1.4 Analyses de la population n'ayant pas d'enfant (« d=0 »)*

On observe des tendances similaires au sein du groupe de femmes n'ayant pas d'enfants âgés de moins de 6 ans noté (« d=0 » ). Il s'agit, autrement dit, d'un nombre important de données manquantes dans la variable de résultat et une forte variabilité.

#### *1.1.2.1.5 Résultats de la population n'ayant pas d'enfant (« d=0 »)*

| | **wage** | **educ** | **exper** |
|---|---|---|---|
| **wage** | 1 | 0.36 | 0.08 |
| **educ** | 0.36 | 1 | 0.01 |
| **exper** | 0.08 | 0.01 | 1 |

*Tableau 5: Matrice de covariance du groupe m0*

#### *1.1.2.1.6 Analyses de la population ayant au moins un enfant (« d=1»)*

S'agissant de l'analyse des variables, pour le groupe (d=0), on peut constater que les corrélations entre les variables sont relativement faibles ou quasiment nulles. Par exemple, le coefficient de corrélation des variables « exper » et « wage » est de 0.08. Donc, les deux variables sont linéairement indépendantes.

#### *1.1.2.1.7 Résultats de la population ayant au moins un enfant (« d=1 »)*

| **Statistic** | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| **wage** | 53 | 4.45 | 4.52 | 0.13 | 25.00 |
| **educ** | 147 | 12.69 | 2.25 | 7 | 17 |
| **exper** | 147 | 7.03 | 5.25 | 0 | 25 |
| **d** | 147 | 1.00 | 0.00 | 1 | 1 |

*Tableau 6:Statistique descriptive du groupe m1*

#### *1.1.2.1.8 Analyses de la population ayant au moins un enfant (« d=1 »)*

Bien que l'échantillon constitué à partir des observations des femmes ayant au moins un enfant âgé de moins de 6 ans soit plus petit que celui des femmes n'ayant aucun, on peut constater que les caractéristiques des échantillons restent les mêmes notamment en termes de variance et de données manquantes.

#### *1.1.2.1.9 Résultats de la population ayant au moins un enfant (« d=1 »)*

| | **wage** | **educ** | **exper** |
|---|---|---|---|
| **wage** | 1 | 0.27 | -0.03 |
| **educ** | 0.27 | 1 | 0.004 |
| **exper** | -0.03 | 0.004 | 1 |

*Tableau 7 : Matrice de covariance du groupe m1*

#### *1.1.2.1.10 Analyses de la population ayant au moins un enfant (« d=1 »)*

Dans cette population, les relations entre les différentes variables sont beaucoup plus faibles que celles calculées dans le groupe (« d=0 »). Cela peut être observé par les coefficients de corrélation des $\rho_{(educ,wage)}$ $et$ $\rho_{(exper,wag)}$ qui passent respectivement de 0.36 et 0.08 dans le groupe (« d=0 ») à 0.27 et -0.03 dans le groupe (« d=0 »).

Bien que plusieurs facteurs puissent expliquer ces légères différences entre lesdits populations, il paraît avant toute chose légitime de se poser comme question, si ces différences observées entre les deux populations au moyen des analyses descriptives sont suffisantes pour conclure une différence significative entre les échantillons.

C'est dans le but de répondre à cette question, que nous allons utiliser des tests statistiques paramétriques ou non paramétriques.

### 1.1.2.2 TESTS STATISTIQUES POUR COMPARER LES DEUX POPULATIONS

Comme évoqué ci-dessus, des tests statistiques ont été réalisés dans le but de savoir si les deu x populations sont identiques ou s'il existe des différences significatives entre elles de façon p lus robuste.

#### *1.1.2.2.1 Résultats des tests non paramétriques de Kolmogorov-Smirov*

Le premier test utilisé est le test de Kolmogorov-Smirov qui permet de comparer les distributi ons des probabilités des deux populations sans supposer que les deux populations suivent une l oi de distribution a priori.

| **Asymptotic two-sample Kolmogorov-Smirnov test** | | | |
|---|---|---|---|
| **Data** | m0$lwage and m1$lwage | m0$educ and m1$educ | m0$exper and m1$exper |
| **D** | 0.1115 | 0.077793 | 0.22102 |
| **P-value** | 0.6107 | 0.4712 | 1.911e-05*** |
| **Alternative hypothesis** | two-sided | | |
| **Note:** | *p<0.1; **p<0.05; ***p<0.01 | | |

*Tableau 8: Tests de Kolmogorov-Smirov*

#### *1.1.2.2.2 Interprétations des tests non paramétriques de Kolmogorov-Smirov*

Deux critères ont été mobilisés pour interpréter les résultats de ces tests, il s'agit de : la statistique de test $D$, qui représente la distance maximale entre les fonctions de distribution cumulée des groupes et la p-valeur qui détermine si le test est significatif ou non, autrement dit si on peut rejeter H0. L'hypothèse nulle ici, suppose que les distributions des échantillons sont identiques.

- Comparaison de la distribution des salaires : Comme la statistique D est égale à 0.1115 et la p-valeur vaut 0.6107(>0.05), alors on ne peut pas rejeter l'hypothèse nulle. Autrement dit, il n'y a pas de preuve statistiquement significative pour dire que les deux échantillons proviennent de distributions différentes. Alors, on peut supposer que les deux échantillons suivent la même distribution.
- Comparaison de la distribution du niveau d'éducation : Avec D=0.077793 et une p-valeur de 0.47120(>0.05), on ne peut toujours pas rejeter H0 qui suppose une différence de distribution entre les populations. Ainsi, on peut conclure que là encore, les échantillons sont distribués de façon similaire.
- Comparaison de la distribution de l'expérience : Une statistique D=0.2210 et une p-valeur de 1.911e−05(<0.05) indiquent une différence fortement significative entre les distribution des deux populations. Par conséquent, H0 peut être rejetée.

Il est important de noter, qu'on peut utiliser la puissance de test pour améliorer l'analyse de ces tests statistiques non paramétriques qui intègrent au calcul, un risque de second espèce $\beta$ ou effectuer d'autres tests comme les tests paramétriques qui sont réputés plus robustes.

#### *1.1.2.2.3 Résultats des tests paramétriques de Student sur deux échantillons non appariés*

A la différence du test de Kolmogorov-Smirov qui compare les distributions des échantillons, le test de Student Welch permet de comparer les moyennes des populations et indiquent si ces dernières sont différentes ou non.

| **Welch Two Sample t-test** | | |
|---|---|---|
| **Data** | m0$educ and m1$educ | m0$exper and m1$exper |
| **t** | -2.4412 | 8.1069 |
| **df** | 224.58 | 350.04 |
| **p-value** | 0.01541 | 8.81e-15 |
| **Ninety-five percent confidence interval** | 0.91401531 -0.09750217 | 3.385026 5.553548 |
| **Mean in group nochildren** | 12.18812 | 11.503300 |
| **Mean in group children** | 12.69388 | 7.034014 |
| **Alternative hypothesis:** | true difference in means between group nochildren and group children is not equal to 0 | |
| **note:** | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

*Tableau 9: Tests de Student*

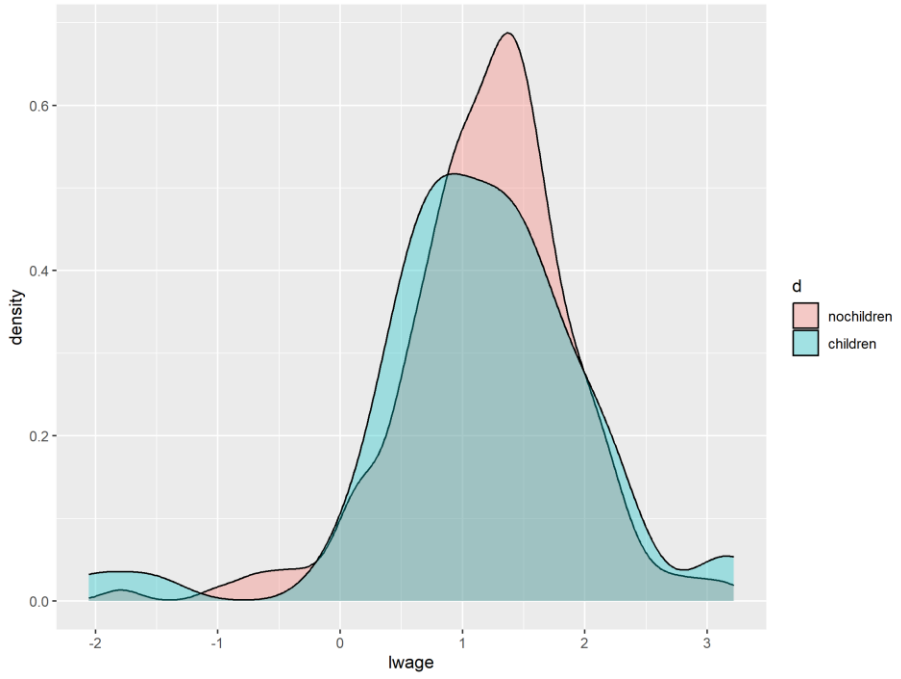
- Comparaison du niveau d'éducation par groupe : Comme la p-valeur = 0.01541, alors on peut rejeter H0. Autrement dit, il existe une différence statistiquement significative entre le niveau d'éducation moyen des femmes ayant des enfants et celui n'ayant pas. Par ailleurs, la moyenne du groupe « nochildren » qui vaut 12.18812 est inférieure à celle du groupe « children » qui est égale à est 12.69388.
- Comparaison du nombre d'année d'expérience par groupe : La statistique de t est égale à 8.1069 et la p-valeur vaut 8.81e−15 suggèrent que les moyennes associées aux nombre d'années d'expérience au sein des groupes sont significativement différentes.

En définitive, le test de Kolmogorov-Smirov et le test de Student effectués, suggèrent que les échantillons constitués à partir des groupes de femmes ayant au moins un enfant âgé d'au moins de 6 ans et celles n'ayant pas sont significativement différents en moyenne et en distribution au regard de la variable « exper ».

Cependant, en ce qui concerne la variable « educ », les tests indiquent que les échantillons sont distribués de façon similaire et que les moyenne de ces échantillons sont significativement différentes. Donc, les femmes ayant au moins un enfant d'âgé de moins de 6 ans ont un niveau d'éducation supérieur à celle n'ayant aucun.

Enfin, le test de Kolmogorov-Smirov réalisé à partir de la variable « lwage » montre que les échantillons suivent la même loi de probabilité.

### 1.1.3 Comparaison des deux populations en utilisant une représentation graphique




*Graphique 1: Densité de la variable salaire en fonction du statut de la femme*

#### 1.1.3.1 RESULTATS

Mon choix s'est porté sur les courbes de densité pour comparer graphiquement les deux populations. Plusieurs raisons justifient ce choix :

Tout d'abord, les courbes de densité permettent d'identifier les types de distribution statistiques des deux populations. Dans cette étude, les échantillons semblent être distribués selon une loi normale.

Ensuite, on peut comparer visuellement les deux groupes, en s'appuyant sur la forme et le chevauchement des distributions qui donnent des indications sur des paramètres tels que la moyenne, l'écart-type, le coefficients d'asymétrie et le coefficient d'aplatissement.

On observe par exemple, que la moyenne du logarithme de salaire du groupe des femmes n'ayant pas d'enfant est supérieure à celle du groupe de femme ayant des enfants. Autrement dit, les femmes qui n'ont pas d'enfants ont un salaire (**par transformation strictement croissant**) supérieur à la moyenne du groupe de femme ayant des enfants.

De plus, graphiquement, la distribution du groupe ayant des enfants est plus aplatie que celle du groupe de femme n'ayant pas d'enfants. De ce fait, on peut déduire que les salaires au sein du groupe des femmes ayant des enfants sont beaucoup plus hétérogènes que ceux du groupe des femmes n'ayant pas d'enfants.

### 1.1.4 Comparaison des valeurs moyennes des deux populations au moyen d'un test de Student et une régression par les moindres carrés ordinaires.

Pour comparer la moyenne de la variable résultat (lwage) entre les deux populations, nous allons utiliser deux méthodes différentes : un test de Student et une régression par moindres carrés ordinaires.

### 1.1.5 Test de Student

Une brève description de la nature d'un test de Student a déjà été effectuée dans la partie 1.1.2.2.3.

##### *1.1.5.1.1 4.1 RESULTATS DU TEST PARAMETRIQUE DE STUDENT DEUX ECHANTILLONS NON APPARIES*

| **Welch Two Sample t-test** | |
|---|---|
| **Data** | lwage by d |
| **t** | 0.45952 |
| **df** | 61.079 |
| **p-value** | 0.6475 |
| **95 percent confidence interval** | -0.1983866 0.3167744 |
| **Mean in group nochildren** | 1.197503 |

| | |
|---|---|
| **Mean in group children** | 1.138309 |
| **Alternative hypothesis:** | true difference in means between group nochildren and group children is not equal to 0 |
| **note:** | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

*Tableau 10:Test de Student entre lwage et d*

#### *1.1.5.1.2 Analyse du test paramétrique de Student deux échantillons non appariés*

- La statistique de test t, mesure la différence entre les moyennes des deux groupes par rapport à la variabilité des échantillons. Dans ce test, elle est égale à 0.45952. Cela signifie que les moyenne entre deux populations ne sont pas très différentes.
- De plus comme la p-valeur est égale à 0.6475, alors on ne peut pas rejeter l'hypothèse nulle, qui suppose pour rappel, que les moyennes des échantillons sont identiques.

Par conséquent, on peut conclure que le test de Student montre qu'il n'y a pas de différence statistiquement significative entre le salaire moyen du groupe des femmes ayant des enfants et celui des femmes n'ayant pas d'enfant.

De plus, on peut constater que le salaire moyen du groupe des n'ayant pas d'enfant (lwage=1.20) est supérieur à celui des femmes ayant au moins un enfant âgé de moins de 6 ans enfants. (lwage=1.40), cela représente une différente de 0.6 sur une échelle logarithmique. Bien que cette différence ne soit pas significative.

### 1.1.5.2 4.2. MODELE DE REGRESSION LINEAIRE MULTIPLE PAR METHODE DES MOINDRES CARRES ORDINAIRES (MCO)

Pour comparer les valeurs moyennes des deux populations en utilisant une régression linéaire multiple par la méthode des Moindres Carrés Ordinaires (MCO), une variable indicatrice (« d ») a été introduite dans le modèle de régression, qui peut s'écrire comme suit :

$$lwage = \beta_0 d_1 + \beta_1 Educ + \beta_2 Exper + \mu \text{ où}$$

- lwage : est la variable d'intérêt
- d1 : la variable indicatrice qui prend la valeur « 1 » pour le groupe de femme ayant des enfants et « 0 » sinon.
- Educ et Exper : les variables de contrôles
- $\mu$ : le résidus
- $(\beta_0, \beta_1, \beta_2)$ :les coefficient de régression

#### *1.1.5.2.1 Les résultats de cette régression ont été résumés dans le table ci-dessous.*

| | *Dependent variable:* |
|---|---|
| | lwage |
| **dchildren** | -0.072 |

| | |
|---|---|
| | (0.101) |
| **educ** | 0.111*** |
| | (0.014) |
| **exper** | 0.015*** |
| | (0.004) |
| **Constant** | -0.398** |
| | (0.190) |
| **Observations** | 428 |
| **$R^2$** | 0.149 |
| **Adjusted $R^2$** | 0.143 |
| **Residual Std. Error** | 0.669 (df = 424) |
| **F Statistic** | 24.820*** (df = 3; 424) |
| ***Note:*** | *p<0.1; **p<0.05; ***p<0.01 |

*Tableau 11: Tableau de régression multiple*

#### *1.1.5.2.2 Interprétation de cette régression*

- La valeur du coefficient de régression d1 est égale à -0.072. Cela signifie que les femmes ayant des enfants gagnent en moyenne 7,2 % de moins que celles n'en ayant pas.
- Par ailleurs, la p-valeur associée à ce coefficient qui est égale à 0.476207 indique que les résultats cités ci-dessus ne sont pas statistiquement significatifs. Autrement dit, on ne peut pas conclure qu'avoir des enfants a un effet systématique sur le logarithme du salaire dans cette population.

### 1.1.6 Écriture d'un modèle linéaire de Rubin permettant de mesurer l'effet du traitement sur les traités.

L'écriture d'un modèle linéaire causal de Rubin permettant de mesurer l'effet de la variable de traitement sur les traités peut s'écrire comme :

$$y = \alpha_0 + (X - m)\beta_O + \gamma W + W(X - m)\delta + \mu \text{ où}$$

- $y$ : est la variable de résultat associée.
- $W$ : est la variable de traitement binaire.
- $X$ : la matrice des variables de confusion.
- $m$ : est le centrage.

Pour mesurer l'effet de traitement sur les traités (ATT), il faut utiliser un centrage par rapport à $\overline{X_1}$ , où $\overline{X_1}$ correspond au vecteur des moyennes de X sur le sous-échantillon défini par W = 1, c'est-à-dire les personnes traitées. (Dugret, 2024)

### 1.1.7 Estimation de l'effet moyen de la présence d'enfants en bas âge sur le salaire et comparaisons avec la différence de salaire entre les populations.

En se basant sur l'écriture du modèle de régression de Rubin pour mesurer l'effet du traitement sur les traités, on peut en déduire que le modèle de régression qui permet de mesurer l'effet d'avoir des enfants en bas âge sur le logarithme de salaire en tenant compte des variable de confusion que sont le niveau d'éducation et le nombre d'années d'expérience peut s'écrire comme suit :

$$lwage = \alpha_0 + (X - \overline{X_1})\beta_0 + \gamma W + W(X - \overline{X_1})\delta + \mu \text{ où}$$

- $y$ : est la variable de résultat associée au logarithme du Salaire.
- $W$ : est la variable de traitement binaire qui indique si une femme a au moins un enfant de moins de 6 ans (W=1) ou pas (W=0).
- $X$ : la matrice des variables de confusion (exper et educ ) qui influencent à la fois $y$ et $W$.
- Avec $m = \overline{X_1}$ pour mesurer l'effet moyen du traitement sur traités ATT sur le logarithme du salaire

Les résultats de cette régression ont été résumés dans le tableau ci-dessous :

#### 1.1.7.1 RESULTATS DU MODELE 1 DE RUBIN

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Y | | | |
| | (1) | (2) | (3) | (4) |
| **W** | -0.059 | | -0.072 | -0.074 |
| | (0.106) | | (0.101) | (0.101) |
| **c.exper** | | 0.016*** | 0.015*** | 0.014*** |
| | | (0.004) | (0.004) | (0.004) |
| **c.educ** | | 0.109*** | 0.111*** | 0.106*** |
| | | (0.014) | (0.014) | (0.015) |
| **W:c.exper** | | | | 0.017 |

| | | | | |
|---|---|---|---|---|
| | | | | (0.018) |
| **W:c.educ** | | | | 0.037 |
| | | | | (0.043) |
| **Constant** | 1.198*** | 1.198*** | 1.210*** | 1.212*** |
| | (0.037) | (0.038) | (0.042) | (0.043) |
| **Observations** | 428 | 428 | 428 | 428 |
| **$R^2$** | 0.001 | 0.148 | 0.149 | 0.153 |
| ***Note:*** | *p<0.1; **p<0.05; ***p<0.01 | | | |

*Tableau 12: Table de régression du model 1 du Rubin*

#### 1.1.7.2 INTERPRETATION

Au regard des résultats, on peut conclure que l'effet moyen de la présence d'enfants en bas âge sur le salaire est de -0.074. Ce qui signifie que selon le modèle, les femme ayant un enfant âgé de moins de 6 ans ont un salaire inférieur à hauteur de 7.4% que les enfants n'ayant pas d'enfants. Cependant, ces résultats ne peuvent pas être inférés car la p-value associée au coefficient de régression n'est pas significative.

Par ailleurs, les résultats empiriques calculés à partir du logarithme des salaires indiquent une différence de 6%. Donc, le modèle à surestimer cet effet.

### 1.1.8 7. Addition d'une nouvelle variable de contrôle et comparaison des résultats avec les précédents résultats

C'est dans le but d'améliorer la robustesse du modèle de Rubin, qu'une nouvelle variable a été introduite. Il s'agit de « exper2 » qui est une variable calculée à partir d'une fonction quadratique et de la variable « exper »

#### 1.1.8.1 RESULTATS DU MODELE 2 DE RUBIN

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Y | | | |
| | (1) | (2) | (3) | (4) |
| **W** | -0.059 | | -0.060 | -0.060 |
| | (0.106) | | (0.101) | (0.102) |
| **c.exper** | | 0.042*** | 0.041*** | 0.037** |
| | | (0.013) | (0.013) | (0.014) |
| **c.educ** | | 0.107*** | 0.109*** | 0.104*** |

| | | | | |
|---|---|---|---|---|
| | | (0.014) | (0.014) | (0.015) |
| **c.exper2** | | -0.001** | -0.001** | -0.001 |
| | | (0.0004) | (0.0004) | (0.0004) |
| **W:c.exper** | | | | 0.047 |
| | | | | (0.054) |
| **W:c.educ** | | | | 0.030 |
| | | | | (0.044) |
| **W:c.exper2** | | | | -0.002 |
| | | | | (0.003) |
| **Constant** | 1.198*** | 1.187*** | 1.198*** | 1.198*** |
| | (0.037) | (0.039) | (0.043) | (0.044) |
| **Observations** | 428 | 428 | 428 | 428 |
| **$R^2$** | 0.001 | 0.157 | 0.158 | 0.160 |
| ***Note:*** | *p<0.1; **p<0.05; ***p<0.01 | | | |

*Tableau 13 : Table de régression du model 2 de Rubin*

#### 1.1.8.2 INTERPRETATION DES RESULTATS MODELE DE RUBIN 2

Comme on peut le constater, la variable W n'est toujours pas significative bien que le modèle ait gagné en qualité, car le $R^2$ finale passe de 15.3% à 16%.

En ce qui concerne la comparaison entre la moyenne des salaire calculée à partir des données empirique et celle estimée par le modèle de Rubin2, la différence est maintenant de 0. En définitive, ce modèle est plus en accord avec les observations empiriques