

PROJET ECONOMETRIQUE SOUS R

Ce projet est à remettre au plus tard le 2 juillet minuit. Vous devrez également présenter les résultats de l'étude réalisée lors d'une soutenance de 15 minutes (5 minutes de présentation et 10 minutes de questions) qui aura lieu le 6 juillet.

Etude à réaliser :

- *Objectif :*

Analyser l'écart de salaire mensuel net entre la population immigrée et la population née en France.

- *Données :*

Les données à mobiliser sont issues des enquêtes *Emploi* réalisées par l'Insee en 2015 et 2016.

Les données pour l'année 2015 sont disponibles dans le fichier *enq_emploi_salarie_2015.dta* et celles pour l'année 2016 dans le fichier *enq_emploi_salarie_2016.dta*. Ces données sont en format STATA. Ces deux fichiers sont disponibles sur Moodle.

Au sein de ces deux bases, seule la population salariée a été conservée. En outre, pour l'ensemble des variables disponibles, les valeurs manquantes et aberrantes ont été supprimées.

Les différentes variables présentes dans les deux bases sont décrites dans le tableau présenté en annexe.

Questions :

Vous répondrez à ces questions en utilisant le logiciel R et vous devrez joindre à vos réponses le programme correspondant.

1. La population immigrée rassemble les immigrés de première et seconde génération. Les immigrés de première génération sont les personnes nées étrangères à l'étranger, indépendamment de la nationalité et de l'origine de leurs parents. Les immigrés de seconde génération correspondent aux individus nés français en France dont au moins un parent est né étranger à l'étranger.

En considérant l'ensemble des salariés enquêtés en 2015 et 2016, la proportion d'immigrés s'élève à :

- a. 12,22%
- b. 16,49%
- c. 23,12%

2. A combien s'élève l'écart de salaire moyen entre la population née en France et la population immigrée ? Cette différence est-elle significativement différente de zéro ?

3. Comparer à l'aide d'un tableau les caractéristiques de la population immigrée avec celles de la population née en France.

Les caractéristiques à comparer sont les suivantes : âge, sexe, statut marital, ancienneté, niveau de diplôme, catégorie socio-professionnelle, nombre d'heures, salaire, nature des horaires et travail de nuit.

Les éventuelles différences significatives observées peuvent-elles en partie expliquer l'écart de salaire moyen entre ces deux populations ? Expliquer.

4. La différence obtenue à la question 2 se modifie-t-elle « toutes choses égales par ailleurs » (c'est-à-dire à âge, sexe, statut marital, ancienneté, niveau de diplôme¹, csp, nombre d'heures, salaire, nature des horaires et travail de nuit équivalents²). Dans le modèle à estimer, vous exprimerez le salaire en logarithme.

Comment interprétez-vous le coefficient associé au sexe ainsi que celui associé à l'ancienneté ?

5. a – Estimer, au sein de chaque population, une équation de salaire mesurant l'effet de chaque caractéristique sur le salaire (exprimé en logarithme), « toutes choses égales par ailleurs ». Présenter les résultats obtenus sous forme d'un tableau.

b – Estimer l'écart entre le salaire moyen des salariés nés en France et le salaire moyen que les salariés immigrés auraient obtenu s'ils étaient rémunérés comme les salariés nés en France. Interpréter.

c – Estimer l'écart entre le salaire moyen que les salariés immigrés auraient obtenu s'ils étaient rémunérés comme les salariés nés en France et leur salaire moyen. Interpréter.

6. Quels sont les principaux éléments explicatifs de l'écart de salaire moyen observé entre la population née en France et la population immigrée ? Pour répondre à cette question, vous mettrez en œuvre une méthode de décomposition détaillée de type Oaxaca-Blinder avec une pondération correspondant au processus de formation des salaires des salariés nés en France. Seule la partie « expliquée » de l'écart de salaire entre les deux populations sera décomposée (sans procéder à une normalisation pour les variables qualitatives à plus de deux modalités). Vous pourrez présenter les résultats sous forme d'un tableau ou d'un graphique.

¹ Vous regrouperez au sein d'une même catégorie les salariés ayant le brevet des collèges et ceux n'ayant aucun diplôme.

² Concernant la nature des horaires, vous distinguerez, dans la spécification des modèles, les 3 modalités de la variable *horaic*. En revanche, concernant le travail nuit (variable *nuitc*), vous regrouperez dans une même modalité le fait de travailler la nuit plus de la moitié du temps et le fait de travailler la nuit moins de la moitié du temps.

Annexe

Nom	Description	Modalités
lnais	Lieu de naissance de l'enquêté	lnais=1 si naissance en France (y compris DOM, COM et POM) lnais=2 si naissance à l'étranger
nfrred	Nationalité de l'enquêté	nfrred=1 si français de naissance, y compris par réintégration nfrred=2 si Français par naturalisation, mariage, déclaration ou option à sa majorité nfrred=3 si étranger
naiper	Lieu de naissance du père	naiper=1 si naissance en France (y compris DOM, COM et POM) naiper=2 si naissance à l'étranger
nfrp	Nationalité du père	nfrp=1 si français de naissance, y compris par réintégration nfrp=2 si Français par naturalisation, mariage, déclaration ou option à sa majorité nfrp=3 si étranger
naimer	Lieu de naissance de la mère	naimer=1 si naissance en France (y compris DOM, COM et POM) naimer=2 si naissance à l'étranger
nfrm	Nationalité de la mère	nfrm=1 si français de naissance, y compris par réintégration nfrm=2 si Français par naturalisation, mariage, déclaration ou option à sa majorité nfrm=3 si étranger
salaire	Salaire mensuel net retirée de la profession principale, corrigée des valeurs aberrantes et manquantes	salaire= montant en euros
heures	Nombre d'heures travaillées en moyenne par mois dans l'emploi principal (heures supplémentaires comprises), corrigé des valeurs aberrantes et manquantes	heures = nombre d'heures travaillées par mois
datant	Année de début dans l'entreprise	datant=année de début dans l'entreprise
sexe	Sexe	sexe=1 si homme sexe=2 si femme
age	Age	age = âge en années

Nom	Description	Modalités
ddipl	Diplôme le plus élevé obtenu	ddipl=1 si diplôme supérieur à Bac+2 ddipl=3 si Bac+2 ddipl=4 si Bac ou brevet ou diplôme équivalent ddipl=5 si CAP, BEP ou autre diplôme de ce niveau ddipl=6 si brevet des collèges ddipl=7 si aucun diplôme ou CEP
couple	Statut marital	couple =1 si vit en couple couple=0 si vit seul
annee_enq	Année d'enquête	annee_enq=année d'enquête
cstotr	Catégorie socio-professionnelle	cstotr=3 si cadre cstotr=4 si profession intermédiaire cstotr=5 si employé cstotr=6 si ouvrier
horaic	Nature des horaires	horaic=1 si horaires à peu près semblables d'une semaine sur l'autre horaic=2 si horaires alternés : 2×8, 3×8, équipes... horaic=3 si horaires variables d'une semaine sur l'autre.
nuitc	Travail de nuit (entre minuit et 5 heures du matin)	nuitc=1 si travail de nuit plus de la moitié du temps nuitc=2 si travail de nuit moins de la moitié du temps nuitc=3 si pas de travail de nuit