Evaluation des Politiques Publiques
ENS Paris Saclay
Licence 3 d'Economie
2023-2024

Enseignant : Yao Thibaut Kpegli

Application 2 (Python) : évaluation d'impact avec la méthode de double différence

Cette application consiste à utiliser la méthode de double différence pour évaluer l'impact de la construction d'écoles secondaires sur le niveau d'éducation primaire des enfants. Elle se base sur des données Tanzaniennes provenant de la Banque Mondiale (Beegle et al. 2006). Ces données concernent des communautés et des ménages de la région de Kagera en Tanzanie.

Dans les années 1980, un grand nombre de nouvelles écoles ont été construites dans la région de Kagera. Cependant, des écoles secondaires n'ont pas été construites dans tous les villages, permettant la comparaison d'individus provenant de villages avec et sans écoles, quelques années plus tard.

Informations générales

La base de données concerne des individus vivant dans des villages avec et sans écoles secondaires.

Les villages où une nouvelle école secondaire a été construite seront appelés « groupe de traitement ». Les autres villages (sans école secondaire sur la période) seront appelés « groupe de contrôle ».

Pour chaque groupe, on utilisera deux classes d'âge d'individus. Les jeunes individus (âgés de 6 à 16 ans en 1985), et les plus vieux (âgés de 21 à 41 ans). L'idée est que l'impact des nouvelles écoles ne peut se faire sentir que sur les individus qui étaient encore en âge d'y aller au moment de leur construction. Ces deux cohortes d'individus, combinées à la construction d'écoles dans certains villages et non dans d'autres, permettront l'estimation de la double différence.

Le tableau suivant donne le descriptif de l'ensemble des variables de la base de données.

Variables	**Description**
treat	=1 si une école secondaire a été construite dans le village
ycohort	=1 si l'individu avait entre 6 et 16 ans en 1985
ocohort	=1 si la personne était âgée de 21 à 41 ans en 1985.
ycohortxtreat	termes d'interaction entre cohorte et groupe de traitement.
primary	=1 if la personne a fini l'école primaire
electric	=1si le village est électrifié.
pipwater	=1 si le village a accès à un réseau d'eau potable.
Distcapital	= distance entre le village et la capitale (Dar es Salam)
cluster	= identifiant du village
age1994	=Age de l'individu en 1994
male1994	=1 si la personne est un homme

I. Première partie : Statistiques descriptives

Cette partie vous permet de décrire votre échantillon. Quelle est l'importance du problème considéré dans votre échantillon (niveau d'éducation) ? Quelles sont les différences apparentes entre groupes de traitement et groupe de contrôle ?

1. Quel est l'âge moyen et sa distribution ? Quelle est la proportion d'hommes ? Quelle est la proportion d'individus ayant terminé leur étude primaire ?

2. Présentez séparément ces résultats pour le groupe de traitement et le groupe de contrôle. Pour chaque variable, faites un test statistique pour évaluer s'il existe des différences significatives (au seuil de 5%) entre les groupes de traitement et de contrôle. Commentez vos résultats. En particulier, si vous trouvez des différences significatives, indiquez en quoi cela peut affecter les résultats de l'évaluation d'impact.

3. Refaite l'analyse de la question 2 pour la cohorte des jeunes individus (âgés de 6 à 16 ans en 1985) et la cohorte des plus vieux (âgés de 21 à 41 ans en 1985). Que pouvez vous conclure sur l'hypothèse du ''common trend'' ?

II. Deuxième partie : double différence, comparaison de moyennes.

1. Calculez les moyennes (a, b , c et d) d'éducation primaire pour chacun des groupes :
a. Quel pourcentage de la cohorte « jeune » a fini l'école primaire dans les villages de contrôle ?
b. Quel pourcentage de la cohorte « âgée » a fini l'école primaire dans les villages de contrôle ?
c. Quel pourcentage de la cohorte « jeune » a fini l'école primaire dans les villages de traitement ?
d. Quel pourcentage de la cohorte « âgée » a fini l'école primaire dans les villages de traitement ?

2. Remplissez le tableau suivant

	Villages de contrôle	Villages de traitement	
Cohorte jeune	a	c	
Cohorte âgée	b	d	
Différences	a-b	c-d	(c-d) - (a-b)

III. Troisième partie : double différence avec une régression MCO

1. Faites une régression du niveau d'éducation primaire (complété/non-complété) sur la variable de traitement, la variable binaire de la cohorte des jeunes (*ycohort*) et la variable d'interaction entre le traitement et la cohorte des jeunes (*ycohortxtreat*). Comparez le résultat de régression avec celui obtenu dans la deuxième partie (tableau). Quelle est sa significativité statistique de l'effet ?

2. Faites la même régression qu'en 1, mais cette fois en ajoutant d'autres variables de contrôle. Est-ce que la valeur du coefficient sur la variable de traitement change ?

IV. Quatrième partie : Hétérogénéité

1. Etudiez l'hétérogénéité en fonction du sexe.

Sur la base de l'ensemble de vos analyses, pensez-vous que la construction d'écoles secondaires a eu un effet positif sur les résultats en école primaire ? Quels sont les éventuels problèmes avec ce type d'analyse ?

Note : cette application est tirée d'un MOOC de Luc Behaghel, Anne-Sophie Robilliard et Philippe de Vreyer