

TP ANALYSE DE LA VARIANCE

Analyse de l'impact de la construction d'écoles secondaires sur le niveau d'éducation primaire

Dans cet exercice, il vous est demandé d'analyser des données provenant de la Banque Mondiale (Beegle et al. 2006). Ces données concernent des communautés et des ménages de la région du Sud dans un pays africain. En utilisant la double différence, il est demandé d'analyser l'impact de la construction d'écoles secondaires sur le niveau d'éducation primaire des enfants. En effet, un grand nombre de nouvelles écoles ont été construites dans les années 1990 dans la région du Sud. 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. Deux niveaux d'analyse sont donc pertinents : la comparaison entre villages, et la comparaison entre individus.

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 1995), 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.

Variables clés :

treat =1 si une école secondaire a été construite dans le village

ycohort=1 si l'individu avait entre 6 et 16 ans en 1995

ocohort=1 si la personne était âgée de 21 à 41 ans en 1995.

ycohortxtreat = termes d'interaction entre cohorte et groupe de traitement.

primary =1 if la personne a fini l'école primaire

electric =1 si 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

age1994=Age de l'individu en 1994

sex1994=1 si la personne est un homme

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) ? Quels sont les différences apparentes entre groupes de traitement et groupe de contrôle.

1. Présentez les caractéristiques moyennes des individus des cohortes jeunes en 1995 pour le groupe de traitement et le groupe de contrôle. Pour chaque variable, faite 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 est-ce que cela peut affecter les résultats de l'évaluation d'impact.
2. Présentez ces mêmes statistiques descriptives au niveau village (cela veut dire n'utiliser qu'une observation par village) pour les variables d'accès à l'électricité, à l'eau potable, et de distance à la capitale selon que le village soit traité ou pas. Commentez vos résultats notamment la significativité des écarts. Indiquez en quoi est-ce que cela peut affecter les résultats de l'évaluation d'impact.
3. Sur la base de vos analyses des questions 1 et 2, pensez-vous que les villages de traitement sont différents des villages de contrôle ? Pourquoi pensez-vous que les écoles ont d'abord été construites dans les villages de traitement ?

Deuxième partie : simple différence

1. Comparez les niveaux d'éducation des cohortes de jeunes (en 1995), entre villages de traitement et de contrôle. Pour cela, faites une régression du niveau d'éducation primaire (complété/non-complété) sur la variable de traitement, en n'utilisant que les observations de la cohorte jeune. Quelle est la valeur du paramètre que vous obtenez pour la variable de traitement ? Quelle est sa significativité statistique ?

Note : De simples régressions MCO sont le moyen le plus couramment utilisé pour estimer l'impact d'une intervention. Si la variable dépendante est binaire, certains choisissent néanmoins d'utiliser les MCO (modèle de probabilité linéaires), quand d'autres préfèrent utiliser des modèles de type Logit ou Probit. Vous avez le choix d'utiliser l'estimateur que vous jugez le plus approprié. Cependant, si les résultats sont similaires de ceux obtenus avec les MCO, ces derniers sont en général préférés car les coefficients sont plus directement interprétables.

2. Sur la base de ces résultats, peut-on dire que la construction d'écoles secondaires a eu un effet direct sur l'éducation primaire ? Ou pourrait-on attribuer ces résultats à d'autres facteurs ?
3. 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 ?

Troisième partie : double différence, comparaison de moyennes.

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

2. Remplissez le tableau suivant.

	Villages de traitement	Villages de contrôle	
Cohorte jeune	A	C	
Cohorte âgée	B	D	
Première différence	a-b	c-d	Double différence : (a-b)-(c-d)

Quatrième partie : double différence en MCO

1. Estimez l'effet de la construction d'école en utilisant une régression MCO. Quel est votre résultat pour l'effet du traitement ? Est-ce qu'il est significatif ? Comment se compare-t-il aux résultats trouvés dans la section précédente ?
2. Faites le même exercice uniquement selon le sexe des personnes
3. 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 ?