

Application 5 (Stata) : évaluation d'impact par la méthode d'appariement sur score de propension

Cette application consiste à évaluer, à l'aide de la méthode d'appariement sur score de propension, l'impact des enlèvements à des fins de participation forcée à une rébellion armée sur l'accumulation de capital humain et la participation au marché du travail. Pour cela, nous utilisons les données de l'article « Blattman, C., & Annan, J. (2010). The consequences of child soldiering. *The review of economics and statistics*, 92(4), 882-898. »

Les données ont été collectées en 2005-2006 dans deux districts de la région Acholi au Nord de l'Ouganda dont la population a été soumise à de nombreux enlèvements à des fins de participation forcée à la rébellion armée conduite par la LRA (Lord's Resistance Army). Il s'agit de données d'une enquête du projet SWAY (Survey of War Affected Youth).

Informations générales

La base de données concerne 741 hommes nés entre 1975 et 1991. Parmi ces hommes, ceux qui ont été enlevés seront appelés « groupe de traitement ». Les autres seront appelés « groupe de contrôle ». Le tableau suivant donne le descriptif des variables clés.

Variable	Description
<i>abd</i>	=1 s'il a été enlevé
<i>illiterate</i>	= 1 si analphabète
<i>educ</i>	Nombre d'années d'éducation atteint
<i>emp_mo</i>	= 1 s'il a travaillé lors des 4 dernières semaines
<i>birthyear</i>	Année de naissance
<i>age</i>	Age de l'enquêté
<i>C_ach</i>	= 1 si né dans l'Acholibur
<i>C_akw</i>	= 1 si né dans l'Akwang
<i>C_ata</i>	= 1 si né dans l'Atanga
<i>C_kma</i>	= 1 si né dans le K. Matidi
<i>C_oro</i>	= 1 si né dans l'Orom
<i>C_pad</i>	= 1 si né dans le Pader
<i>C_paj</i>	= 1 si né dans le Pajule
<i>C_pal</i>	= 1 si né dans le Palabek
<i>fthr_ed</i>	Education du père
<i>mthr_ed</i>	Education de la mère
<i>orphan96</i>	= 1 si les parents sont décédés avant 1997
<i>w10</i>	= 1 s'il faisait partie des 10% les plus pauvres en 1996 (capital/terre)
<i>w25</i>	= 1 s'il faisait partie des 25% les plus pauvres en 1996 (capital/terre)
<i>w75</i>	= 1 s'il faisait partie des 25% les plus riches en 1996 (capital/terre)
<i>w90</i>	= 1 s'il faisait partie des 10% les plus riches en 1996 (capital/terre)
<i>land_access</i>	= 1 s'il a accès à des champs

I. Première partie : Statistiques descriptives

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

1. Présentez les statistiques descriptives au niveau individuel dans l'échantillon. Quel est l'âge moyen et sa distribution ? Quel est le niveau d'éducation moyen et sa distribution ? Quelle est la situation vis-à-vis de l'emploi ?
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 est-ce que cela peut affecter les résultats de l'évaluation d'impact.
3. Sur la base de votre analyse de la question 2, pensez-vous que les individus du groupe de traitement sont différents de ceux du groupe de contrôle ? Quelles sont les caractéristiques de ceux qui ont été enlevés ?

II. Deuxième partie : score de propension

1. Estimer un score de propension (probabilité d'avoir été enlevé) avec un modèle logit en utilisant les variables « *birthyear* » et le lieu de naissance « *C_ach, C_akw, ..., C_pal* »).
2. Quel le score moyen des individus qui ont été enlevés ? Quel est le score moyen de ceux qui ne l'ont pas été ? Commentez les résultats.
3. Représentez sur le même graphique la courbe de densité du score des deux groupes. Que pensez-vous du chevauchement des scores ? Pensez-vous que la méthode d'appariement sur le score de propension est appropriée ?

III. Troisième partie : matching

1. Procédez à un appariement des données et testez l'égalité des moyennes entre groupe de traitement et groupe de contrôle en utilisant :
 - la méthode du plus proche voisin
 - la méthode des 3 plus proches voisins
 - la méthode de noyau (Kernel)

IV. Quatrième partie : impact

1. Estimez l'impact d'avoir été enlevé sur les différentes variables de résultats (analphabétisme, éducation, insertion sur le marché du travail) en utilisant :
 - la méthode du plus proche voisin
 - la méthode des 3 plus proches voisins
 - une approche non-paramétrique
2. Commentez les résultats

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