

Problem Set 11: Matching

Fecha de entrega: 15 de noviembre de 2023

Gastón García Zavaleta y Tomás Pacheco

Instrucciones

- Agreguen los nombres y apellidos de los integrantes en el encabezado del archivo que entregan.
- El título de los archivos entregados debe tener la siguiente estructura:
PS11_Garcia_Gomez_Gonzalez_Perez.pdf
- **UN ÚNICO** miembro del grupo debe subir el PDF y *do-file* al Campus Virtual en la sección "Problem Set 11".
- El código lo tendrán que entregar de dos formas distintas. Primero deberán exportar el *do-file* como PDF y unirlo al informe ([ayuda](#)). Luego, deberán entregar el código en formato .do en el lugar indicado en el Campus Virtual.
- El directorio de trabajo debe tener tres carpetas: programs, input y output. El *do-file* debe ser entregado de tal manera que los tutores podamos correr el código **solamente** modificando el global con el *main directory* (**esto es muy importante**).
- Recuerden que siempre deben presentar la información en tablas y gráficos estéticos. Nunca presenten una tabla o figura sin dedicarle unas líneas a comentar lo que se observa.

Este problem set está basado en el trabajo de Yuri Soares y Micaela Sviatschi titulado "The Impact of Improving Access to Justice on Conflict Resolution". Este paper estudia el impacto de mejorar el acceso a la justicia sobre la resolución de conflictos. En particular, se evalúa el impacto de una intervención cuyo objetivo fue mejorar la cobertura del sistema judicial en zonas alejadas de los grandes centros urbanos. Para ello se crearon distintos centros de justicia (*justice modules*) en zonas remotas preparados para ofrecer los servicios de justicia más importantes.

La base de datos "base_censo.dta" contiene información sobre distintos distritos de Perú. La variable *treated* indica si en ese distrito se abrió o no un centro de justicia.

1. Analicen si los distritos tratados y no tratados son similares en función de ciertas características observables. Para ello, hagan un test de medias sobre la variable población (*pobl_1999*), principal vía de acceso (*via1*) y pobreza (*ranking_pobr*). Comenten los resultados observados. ¿Hay algún problema? ¿Cómo podríamos usar matching para solucionarlo?
2. Calculen el propensity score usando las siguientes variables explicativas: *ind_abs_pobr*, *ldens_pob*, *prov_cap*, *pob_1*, *pob_2*, *pob_3*, *pob_4*, *km_cap_prov*, *via3*, *via5*, *via7*, *via9*, *region_2*, *region_3*, *laltitud*, *tdesnutr*, *deficit_post*, *deficit_aulas*.
3. Grafique la distribución del propensity score para los distritos tratados y para los no tratados. Comenten.
4. Generen una variable binaria que valga 1 si la observación *i* está dentro del *common support*. Quédense sólo con las observaciones que pertenecen al soporte.
5. Corran las siguientes líneas para *matchear* distritos tratados y no tratados:

```
psmatch2 treated if cs1==1, p(pscore) noreplacement  
gen matches=_weight  
replace matches=0 if matches==.
```
6. Comparen nuevamente la distribución del propensity score de ambos grupos, pero considerando solo la submuestra "matches==1". Comenten las diferencias con lo observado en el inciso 3.
7. Repitan los tests de medias del inciso 1 pero usando solo la submuestra "matches==1". ¿Qué observan?
8. Si fueran a correr una regresión usando la muestra *matcheada*, ¿cuál sería el supuesto de identificación?