

## Tarea 1: Microeconometría, Magíster en Economía

En esta tarea trabajaremos con datos de PROGRESA, un programa de transferencias condicionadas en México. En caso de querer información acerca de la naturaleza de esta programa, una lectura potencial es “An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico’s PROGRESA Program” (Díaz y Handa) *J. Human Resources* Spring 2006 vol. XLI no. 2 319-345.

Para realizar esta tarea, utilizaremos los datos PROGRESA.dta (formato Stata, con etiquetas) o PROGRESA.csv (los mismos datos en formato csv sin etiquetas). Para esta tarea, puede utilizar el idioma de programación que quiere. Cuando revisaremos las respuestas en clase, lo haremos utilizando Stata. Si utilizan Stata, los comandos de `teffects` podrían ser bastante útiles para algunas sub-secciones.

La naturaleza de la aleatorización en PROGRESA fue la siguiente: cada comunidad estuvo asignada al azar a ser una comunidad “tratada” o una comunidad “de control”. En el caso que la comunidad está en el grupo de tratamiento, cualquier persona que vive bajo un umbral de pobreza pre-definida recibe una transferencia condicionada de dinero. Esta transferencia está condicionada a que los niño/as en estas familias asistan a colegio. Las familias que viven bajo el umbral de pobreza en las comunidades de control *no* reciben el tratamiento, y por ende son las unidades de comparación. En esta tarea, nos interesa comparar la probabilidad que una persona asista a colegio habiendo recibido la transferencia condicionada, versus el caso en que no recibieron la transferencia.

En la base de datos PROGRESA.dta (o PROGRESA.csv), existen dos periodos temporales indicados con la variable `t`. PROGRESA fue implementado después de periodo 1, y por lo tanto nos interesa evaluar su impacto enfocándonos en el periodo  $t = 2$ . Sin embargo, el periodo  $t = 1$  está incluido en los datos para poder permitir observar algunas variables en la línea base (antes del tratamiento). La base consiste de 28,120 niñas/os con un id individual (`iid`). El identificador único de su hogar es `hh`, y el identificador único de su comunidad es `cid`. El edad del/la niño/a es `age`, y la variable de interés (si asisten o no a colegio) es `enrolled`. Las personas *solamente* son elegibles para recibir el tratamiento si son clasificados como pobres (`pobre`), que está basado en una puntuación de vulnerabilidad (`score`) y un punto de corte regional (`cutoff1`). Existen tantas familias pobres como no pobres en la base de datos. En el caso de las familias no pobres, incluso si su comunidad está aleatorizada para recibir el programa PROGRESA, ellas *no* recibirán la transferencia. Por ende, *estas personas no son ni sujetos tratados ni sujetos no tratados* cuando se realiza la evaluación experimental.

**Instrucciones Generales** La entrega de esta tarea debería consistir de un archivo comprimido (formato .zip o .7z) que contiene: (a) un archivo pdf donde presentan todas sus respuestas y resultados, (b) el código necesario para replicar los resultados que generan/muestran, y (c) en caso de ser necesario, un archivo README explicando cómo replicar este código. La fecha de entrega es el día jueves 9 de septiembre, a más tardar a las 10:59 am. Esta fecha de entrega es estricta, dado que en la clase del día 9 de septiembre, conversaremos acerca de las respuestas a esta tarea. En caso de cualquier problema, por favor contactar directamente al profesor por correo electrónico.

La entrega debe ser por correo electrónico a [dclarke@fen.uchile.cl](mailto:dclarke@fen.uchile.cl), y copiando a [kathya.tapia@usach.cl](mailto:kathya.tapia@usach.cl).

La nota final del trabajo se calcula de la siguiente forma:

Pregunta A: Max 1,5 puntos

Pregunta B: Max 1,5 puntos

Pregunta C: Max 2 puntos

Pregunta D: Max 1 punto

Si el código entregado no replica los resultados entregados, se puede restar hasta 1 punto del total, y la nota final se calcula de acuerdo a:

$$\text{Nota} = 1 + \text{Pregunta A} + \text{Pregunta B} + \text{Pregunta C} + \text{Pregunta D} - \text{Cualquier descuento.}$$

La nota mínima es 1.0, y la máxima es 7.0. Cada persona debería entregar sus propias soluciones y código. En caso de preguntas adicionales por favor contactar a: [dclarke@fen.uchile.cl](mailto:dclarke@fen.uchile.cl).

### **Preguntas:**

**(A) Estadísticas Descriptivas y Diseño Experimental** En base a los datos de PROGRESA incluidos con la tarea, conteste a las siguientes preguntas, proporcionando cualquier resumen de información necesario en el documento que entreguen para apoyar su respuesta.

1. ¿Estos datos forman un panel balanceado? Consideramos las variables `hhincome`, `age`, `female` y `enrolled` (ingreso total de hogar, edad del niño/a, indicador si el individuo es una niña, y estado de enrolamiento en colegio). Muestre estadísticas descriptivas de estas variables, y comente acerca de la cobertura de los datos. ¿Las variables varían mucho entre olas del panel?
2. ¿Qué porcentaje de lo/as niños/as viven en aldeas tratadas? ¿Qué porcentaje están expuestos al tratamiento?
3. ¿Está bien dirigido el programa en términos de pobreza? ¿Existen personas que no clasifican para el programa, pero aún así lograron entrar?
4. Describe (gráficamente o de alguna forma) la variable `score` y su distribución entre individuos clasificados como pobres y como no pobres. Explique la diferencia entre las distribuciones y su implicancia para el experimento.

**(B) Evidencia Experimental del Impacto de PROGRESA** En esta pregunta, examinamos los resultados experimentales de PROGRESA. Por lo tanto, en esta pregunta debería enfocarse solamente en el segundo periodo (salvo posibles variables de control medidas en la línea base), y en los grupos de tratamiento y control.

1. ¿Qué es el estimador de comparación de medias calculando el impacto sobre niño/as que recibieron

PROGRESA considerando como variable dependiente de interés “enrolled”? ¿Dirías que este estimador es un ATT, o un ATE, u otra cosa? ¿Por qué?

2. Existe balance en términos de las variables en la línea base (por ejemplo ingreso de hogar, edad, enrolamiento, puntaje de pobreza, ...)
3. ¿Cuáles son los supuestos necesarios para que el estimador de la parte B1 sea insesgado?
4. ¿Parecen razonables en este contexto? ¿Por qué sí o por qué no?

**(C) Análisis No-experimental** Supongamos que PROGRESA no fue un experimento aleatorio, sino que solamente contamos de las observaciones de los lugares tratados ( $tcomm=1$ ), y los “controles” tuvieron que ser generados a partir de las personas que no fueron elegibles a PROGRESA en estas aldeas. Además, supongamos que no supimos la regla de asignación a PROGRESA, ni tampoco el puntaje de pobreza ( $score$ ).

1. Utilizando las medidas disponibles en la línea base como ingresos de los hogares, edades y géneros, construya un estimador del propensity score. ¿Cómo afectan las variables  $x$  al cálculo del propensity score  $p(x)$ ?
2. Eligiendo algún cálculo del propensity score, inspeccione la distribución del propensity score entre los tratados y los controles. ¿Parece razonable el supuesto de overlap en este contexto?
3. Utilizando el propensity score, calcule el ATT y ATE utilizando emparejamiento. Resume cómo este cálculo depende de la cantidad de vecinos utilizados para generar los pares, y las variables utilizadas para generar el propensity score.
4. ¿Este cálculo es muy sensible a límites máximos de diferencia entre propensity score cuando se matchean las unidades tratadas con las unidades de control?
5. Calcule el ATT utilizando emparejamiento con la métrica de Mahalanobis, y con un emparejamiento exacto en algunas covariables (por ejemplo sexo e ingreso en la línea base redondeado a unidades de 500).
6. ¿Cómo comparen estos métodos de la parte 5 con un cálculo utilizando ponderación por propensity score?
7. Comente acerca de la comparación experimental (de la pregunta B) y los cálculos en base a métodos asumiendo selección sobre los observables (de esta pregunta).

**(D) Inferencia** Comparamos la inferencia de regresión con inferencia de permutaciones.

1. Utilizando el cálculo experimental de la pregunta B, realice un proceso de inferencia utilizando permutaciones de la variable PROGRESA dentro de la muestra (permutando sobre individuos). Compare el impacto real calculado del experimento con la inferencia utilizando permutaciones.

2. ¿Parece razonable el valor p-calculado dado lo que vimos en la inferencia en base a regresión?
3. Ahora, realice el mismo procedimiento que fue realizado en la parte D1, pero ahora permutando PROGRESA a nivel de aldea en vez de permutando la variable a nivel de individuo. ¿Cambia mucho la interpretación del valor-p en este caso?