

# **TRABAJO FINAL DE EVALUACIÓN DE IMPACTO PARA LA GESTIÓN PÚBLICA**

Juan Manuel Rivas Castillo  
Darha Chavez Vasquez  
José Emilio Chicasaca Huamani

08 de agosto de 2023

# Índice

I. DIFERENCIAS EN DIFERENCIAS	II
II. PROPENSITY SCORE MATCHING	VII
Referencias	IX

## **Indicaciones:**

1. El trabajo consiste en realizar dos de los tres ejercicios propuestos en este documento.
2. La entrega consiste en el do file (STATA) comentado. También debe entregar los resultados (tablas, gráficos, etc.) explicados en un archivo Word o PDF.
3. El trabajo se puede realizar en grupos de máximo 4 integrantes.

# I. DIFERENCIAS EN DIFERENCIAS

1. El programa de transferencias condicionadas PROGRESA, conocido también como Oportunidades o Programa de Desarrollo Humano Oportunidades, es una iniciativa social implementada en México con el propósito de combatir la pobreza y promover el desarrollo humano en comunidades vulnerables del país.
2. El objetivo principal de PROGRESA es reducir la pobreza extrema y mejorar las condiciones de vida de las familias más desfavorecidas. Para lograrlo, el programa otorga transferencias económicas mensuales a las familias seleccionadas, siempre y cuando cumplan con ciertos requisitos y condiciones establecidas. Los beneficiarios son familias que viven en áreas rurales o urbanas con altos índices de pobreza y marginalidad, y que cumplen con los criterios definidos por el programa.
3. Entre las condiciones se encuentran aquellas relacionadas con la asistencia escolar de los niños. Los beneficiarios deben asegurar que sus hijos asistan regularmente a la escuela y que se mantengan inscritos para recibir el apoyo económico. De esta manera, PROGRESA busca fomentar la educación como una herramienta para romper el ciclo de pobreza. Además de la educación, el programa también pone énfasis en la salud y nutrición de las familias; los beneficiarios deben participar en programas de salud que incluyen chequeos médicos regulares, vacunación de los niños y acceso a servicios básicos de atención médica.
4. El programa PROGRESA se somete a evaluaciones periódicas para medir su efectividad y asegurar el cumplimiento de los objetivos establecidos. Esto permite ajustar y mejorar el programa con el tiempo, buscando una mayor eficacia en el combate contra la pobreza y la promoción del desarrollo humano en México.
5. La base de datos panel\_progresa\_V1.dta contiene una submuestra de niños y jóvenes con la cual se puede medir el efecto que tiene la transferencia monetaria sobre la tasa de asistencia escolar.
6. El programa comenzó a intervenir a los beneficiarios en junio de 1998. Hizo 2 levantamientos de información previos a la intervención (octubre de 1997 y marzo de 1998) y 3 levantamientos posteriores a la intervención (octubre de 1998, mayo de 1999 y noviembre de 1999). El tratamiento fue asignado de forma aleatoria a nivel de centros poblados o localidades. No obstante, solamente una fracción de los habitantes de una localidad participante se terminó beneficiando.
7. El propósito de este ejercicio es estimar el efecto del programa sobre la tasa de asistencia escolar mediante el estimador de diferencias en diferencias. Para lograr esto se necesitan 3 variables: una variable de tiempo post (después/antes de la intervención), una variable de tratamiento (si el hogar fue tratado o no) y un término de interacción de ambas, el cual es nuestro parámetro de interés.

- a). Organice (comando sort) la base en función de las variables "folio" (identificado de hogar), "ID" (identificador del individuo dentro de su hogar) y "periodo". Genere un identificador para cada individuo juntando las variables ID y folio.

La base de datos *panel\_progresas\_V1.dta* cuenta con 105565 observaciones y 24 variables. Respecto a los grupos formados con las variables folio e ID estos son un total de 21113. Además, debido a la estructura de panel de datos de la base se construyó la variable Periodo. Los códigos empleados se muestran a continuación

```
/* Código empleado para resolver la consulta a */
/-----/
clear all // limpiamos sesión
cd // Asigamos la carpeta de trabajo
use panel_progresas_V1.dta, clear // Cargamos la base de datos
list folio ID periodo in 1/10 // Listamos los primeros 10 datos de folio e ID
local vars 'folio ID periodo' // Guardamos en una local las variables de interés
sort 'vars' // Ordenamos la base según la variable de interés
list folio ID periodo in 1/10 // Observamos las primeras 10 observaciones de
las variables de interés
egen Id = group(folio ID) // Creamos el identificador de grupos
label var Id 'Identificador por individuos' // Etiquetamos la variable Id
bys folio ID (periodo): gen Periodo = _n // Generamos el número de observa-
ciones por individuos
list Id Periodo in 1/10 // Volvemos a listar la base con las nuevas variables
/-----/
```

=====

Estructura del Panel de Datos de las primeras 10 observaciones

	Id	Periodo
1.	1	1
2.	1	2
3.	1	3
4.	1	4
5.	1	5
6.	2	1
7.	2	2
8.	2	3
9.	2	4
10.	2	5

- b). Genere una variable dummy para cada ronda de encuesta. Es decir, convierta la variable categórica que muestra la base a una variable binaria para cada ronda de encuesta.

Para responder a esta consulta nos generamos un bucle que genere una variable dummy en función de cada uno de los periodos. Los detalles de código empleado se presentan a continuación:

```
/* Código empleado para resolver la consulta b */
/-----/
tab periodo, nol
forvalues i=1/5
  g enc'i'=0
  replace enc'i' = 1 if periodo==i'
/-----/
```

- c). Genere las variables post, tratamiento e interacción.

Dado que el programa comenzó a intervenir a los beneficiarios en junio de 1998, se generó una variable dummy que toma el valor de 1 a partir del período 3 (octubre de 1998) y 0 para períodos anteriores. Además, identificamos a la variable **\*\*part\_efectiva\*\*** para identificar a los tratados y controles. Con la multiplicación de las dos variables se generó la variable de interacción y; finalmente, se configuró la base de datos como de tipo panel.

```
/* Código empleado para resolver la consulta c */
/-----/
gen time = (periodo =3) & !missing(periodo)
lab var time "Descripción"
rename part_ treated
gen did = time*treated
un panel
xtset Id Periodo
/-----/
```

- d). Genere una dummy si el hogar es considerado pobre en la primera línea de base (octubre de 1997). Hacer esto por "folio" usando el comando bysort. Es decir, si al menos 1 individuo es considerado pobre, todos los individuos de ese folio también sean considerados pobres.

Tal como solicita esta consulta lo que hacemos es identificar si el individuo es pobre en la línea base y si ocurre ello se extrapolan todos los resultados al folio

```

/* Código empleado para resolver la consulta d */
/-----/
gen poLB = 0
bys folio: replace poLB = 1 if pobre == 1 & periodo==1
egen polb= max(poLB 0), by(folio)
/-----/

```

- e). Como la población objetivo del programa son los pobres, quédese solo con los individuos que se consideran pobres en la primera línea de base (octubre de 1997).

**Tomando en cuenta la extrapolación realizada en la consulta anterior, lo hacemos es quedarnos con todos los folios que cumplen con esta condición**

```

/* Código empleado para resolver la consulta e */
/-----/
keep if polb!=0
/-----/

```

- f). Estime el efecto medio del tratamiento aplicando el método de doble diferencia. No olvide mostrar el valor de la 1era diferencia y de la 2da diferencia.

**Tomemos en consideración que la ecuación a estimar es la siguiente dado el método de doble diferencia:**

$$attendschool_{it} = \alpha + \beta_1 * time + \beta_2 * treated + \beta_3 * (time * treated) + \epsilon_{it}$$

La estimación anterior se realiza empleando Mínimos Cuadrados Ordinarios (MCO). Sin embargo, debido a la estructura de panel de datos se puede considerar la estimación mediante un panel de datos de efectos fijos. No obstante, el estimador within solo permite estimar el efecto promedio entre el grupo de tratamiento y el control como un todo pero no los efectos individuales por unidad. Dado ello, emplearemos MCO para responder a la consulta

```

/* Código empleado para resolver la consulta f */
/-----/
xtset Id Periodo
reg attendschool time treated did
xtreg attendschool time treated did, fe // xtreg
xtreg attendschool time##treated, fe // Forma alternativa
diff everschool, t(treated) p(time) // con diff
/-----/

```

Cuadro 1: Método de Diferencias en Diferencias

	MCO	EF	diff
time	-0.0480*** (0.00399)	-0.0520*** (0.00266)	-0.0480*** (0.00399)
treated	0.0120** (0.00399)	0 (.)	0.0120** (0.00399)
did	0.0352*** (0.00516)	0.0361*** (0.00343)	
_diff			0.0352*** (0.00516)
_cons	0.879*** (0.00309)	0.888*** (0.00130)	0.879*** (0.00309)
<i>N</i>	73098	73098	73098

Standard errors in parentheses

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ 

Como podemos observar del resultado anterior tanto la estimación mediante MCO y el comando diff nos brindan resultados similares con un efecto inicial de 0.0120, efecto que no es estimado por el modelo de panel de datos de efectos fijos; además, se tiene un efecto posterior de -0.0480; mientras que en panel de datos de efectos fijos es de -0.0580. Finalmente, el resultado de DiD es de 0.0352; es decir, el programa aumentó en 3.6 % la asistencia a la escuela.

- g). Proponga y aplique mejoras a la estimación realizada: puede incluir covariables, clusterizar errores, corregir heterocedasticidad, etc.

Para corregir potenciales problemas con los errores estándar se realiza una estimación de tipo Bootstrapping a partir de 50 estimaciones, luego de ello se incluyen un conjunto de covariables asociadas a asistir a la escuela; finalmente, se corrigen los errores estándar de las estimaciones mediante bootstrapping



```

/* Código empleado para resolver la consulta g */
/-----/
* Bootstrapped std. err.: diff everschool, t(treated) p(time) bs rep(50)
* DiD con variables de control (covariates) diff everschool, t(treated) p(time) cov(
edad hhsized educfather educmother nodad nomom sex)
diff everschool, t(treated) p(time) cov( edad hhsized educfather educmother nodad
nomom sex) report
diff everschool, t(treated) p(time) cov( edad hhsized educfather educmother nodad
nomom sex) report bs
/-----/

```

**La inclusión de covariables y el trabajar con bootstrapping vemos que no genera modificaciones en el efecto final.**

Cuadro 2: Método de Diferencias en Diferencias

	M1	M2	M3
time	-0.0480*** (0.00397)	0.00963* (0.00410)	0.00963* (0.00409)
treated	0.0120** (0.00370)	0.00758 (0.00405)	0.00758* (0.00369)
_diff	0.0352*** (0.00558)	0.0335*** (0.00522)	0.0335*** (0.00485)
edad		-0.0429*** (0.000465)	-0.0429*** (0.000500)
hhsiz		-0.000640 (0.000607)	-0.000640 (0.000681)
educfather		0.00731*** (0.000563)	0.00731*** (0.000446)
educmother		0.00719*** (0.000575)	0.00719*** (0.000503)
nodad		0 (.)	0 (0)
nomom		0 (.)	0 (0)
sex		-0.0240*** (0.00251)	-0.0240*** (0.00219)
_cons	0.879*** (0.00267)	1.281*** (0.00764)	1.281*** (0.00789)
<i>N</i>	73098	58533	58533

Standard errors in parentheses

\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

## II. PROPENSITY SCORE MATCHING

1. CANASTA es un programa de apoyo alimentario que buscaba proporcionar una canasta básica de alimentos a familias en situación de vulnerabilidad y pobreza. El objetivo principal del programa es mejorar la nutrición y la alimentación de los hogares más necesitados, contribuyendo así a mejorar su calidad de vida.
2. El programa está dirigido a familias de escasos recursos, en especial a aquellas que se encontraban en situación de pobreza o pobreza extrema, y cuyo acceso a una alimentación adecuada era limitado. La selección de beneficiarios se basaba en criterios socioeconómicos establecidos por las autoridades estatales. Por tanto, la asignación del programa no es aleatoria. En este ejercicio se busca estimar el efecto del programa Canasta sobre la estatura según la edad de los niños beneficiarios. En ese sentido, los investigadores tienen diferentes herramientas para balancear la muestra, de modo que los grupos sean comparables. Una de esas estrategias es Propensity Score Matching (PSM).
3. En la base de datos `.emparejamiento_base` encontrará información de una muestra de 4 mil niños entre tratados y no tratados. Se le pide implementar el PSM siguiendo los siguientes pasos:
  - a). Explore la base de datos identificando la variable tratamiento, la variable de interés y las variables que pueden determinar si el niño accede o no al programa.
  - b). Estime la probabilidad predicha de participación en el programa para cada individuo.
  - c). Realice el emparejamiento con el vecino más cercano sin reemplazo y estime el efecto del programa.
  - d). Realice el emparejamiento con el vecino más cercano con reemplazo y estime el efecto del programa.
  - e). Realice el emparejamiento con los 5 vecinos más cercanos sin reemplazo y estime el efecto del programa.
  - f). Realice el emparejamiento con los 5 vecinos más cercanos con reemplazo y estime el efecto del programa.
  - g). Realice el emparejamiento con kernel y estime el efecto del programa.
  - h). Compare la calidad del emparejamiento de los métodos aplicados. ¿Cuál realiza el mejor emparejamiento?
  - i). Compare los efectos estimados de los diferentes métodos. ¿Cambian mucho? j) ¿Cuál de los métodos elegiría usted para presentar en una investigación? ¿Por qué?

## Referencias