

# Índice

1. Indicaciones:	2
2. DIFERENCIAS EN DIFERENCIAS	2
3. PROPENSITY SCORE MATCHING	5
Referencias	6

## 1. Indicaciones:

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

## 2. DIFERENCIAS EN DIFERENCIAS

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.

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.

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.

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.

La base de datos panel\_progresas.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.

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.

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.

```
. cd "C:\Users\Juan Rivas\Documents\Drive\CIES"
C:\Users\Juan Rivas\Documents\Drive\CIES
. clear all
. notes: Cargamos la base de datos
. use panel_progres_V1.dta, clear
. notes: Listamos los primeros 10 datos de folio e ID
. list folio ID periodo in 1/10
```

	folio	ID	periodo
1.	11038	3	Octubre 1997
2.	11038	3	Noviembre 1999
3.	11038	3	Mayo 1999
4.	11038	3	Marzo 1998
5.	11038	3	Octubre 1998
6.	11038	4	Octubre 1997
7.	11038	4	Marzo 1998
8.	11038	4	Mayo 1999
9.	11038	4	Octubre 1998
10.	11038	4	Noviembre 1999

```
. notes: Construimos una variable local para almacenar las variables a ordenar
. local vars "folio ID periodo"
. sort `vars`
. notes: Observamos las primeras 10 observaciones de las variables de interés
. list folio ID periodo in 1/10
```

	folio	ID	periodo
1.	11038	3	Octubre 1997
2.	11038	3	Marzo 1998
3.	11038	3	Octubre 1998
4.	11038	3	Mayo 1999
5.	11038	3	Noviembre 1999
6.	11038	4	Octubre 1997
7.	11038	4	Marzo 1998
8.	11038	4	Octubre 1998
9.	11038	4	Mayo 1999

```
10. | 11038    4  Noviembre 1999 |
```

```
. notes: Creamos el identificador de grupos
. egen Id = group(folio ID)
. notes: Generamos el número de observaciones por individuos
. label var Id "Identificador por individuos"
. bys folio ID (periodo): gen Periodo = _n
. notes: Volvemos a listar la base con las nuevas variables
. list Id Periodo in 1/10
```

	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.

```
.
. notes: Nos generamos una dummy por cada periodo de la encuesta
. tab periodo, nol
```

Fecha de encuesta	Freq.	Percent	Cum.
1	21,113	20.00	20.00
2	21,113	20.00	40.00
3	21,113	20.00	60.00
4	21,113	20.00	80.00
5	21,113	20.00	100.00
Total	105,565	100.00	

```
. forvalues i=1/5 {
2.     g enc`i`=0
3.     replace enc`i` = 1 if periodo==`i`
4. }
(21,113 real changes made)
(21,113 real changes made)
(21,113 real changes made)
(21,113 real changes made)
(21,113 real changes made)
```

c) Genere las variables post, tratamiento e interacción.

```
. **** Generamos variable dummy cuando el tratamiento comienza. El programa comenzó a intervenir
> a los beneficiarios en junio de 1998
. gen time = (periodo>=3) & !missing(periodo)
. rename part_ treated
. **** Creamos la interaccion entre ambos
. gen did = time*treated
. **** Le indicamos a stata que trabajamos con un panel
. xtset Id Periodo
```

Panel variable: Id (strongly balanced)  
Time variable: Periodo, 1 to 5  
Delta: 1 unit

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.

```
. gen poLB = 0
. bys folio: replace poLB = 1 if pobre == 1 & periodo==1
(14,748 real changes made)
. egen polb= max(poLB > 0), by(folio)
. tab polb
```

polb	Freq.	Percent	Cum.
0	31,825	30.15	30.15
1	73,740	69.85	100.00
Total	105,565	100.00	

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).

```
. keep if polb!=0
(31,825 observations deleted)
```

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.

```
. xtset Id Periodo
Panel variable: Id (strongly balanced)
Time variable: Periodo, 1 to 5
Delta: 1 unit
```

```
. * Con reg
. reg everschool time treated did
```

Source	SS	df	MS	Number of obs	=	73,291
Model	5.51041576	3	1.83680525	F(3, 73287)	=	70.07
Residual	1920.99871	73,287	.026211998	Prob > F	=	0.0000
				R-squared	=	0.0029
				Adj R-squared	=	0.0028
Total	1926.50912	73,290	.026286112	Root MSE	=	.1619

everschool	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
time	.0205864	.0019263	10.69	0.000	.0168108	.0243619
treated	.0123945	.0019265	6.43	0.000	.0086186	.0161705
did	-.0084079	.0024894	-3.38	0.001	-.0132871	-.0035287
_cons	.9562489	.0014908	641.43	0.000	.953327	.9591709

```
.
. * Con xtreg
. xtreg everschool time treated did, fe
note: treated omitted because of collinearity.
```

```
Fixed-effects (within) regression      Number of obs   =   73,291
Group variable: Id                    Number of groups =   14,748
R-squared:                             Obs per group:
    Within   = 0.0062                      min =          2
    Between  = 0.0002                      avg  =         5.0
    Overall  = 0.0017                      max  =          5
```

everschool	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
time	.0196548	.0012939	15.19	0.000	.0171188	.0221907
treated	0 (omitted)					
did	-.007285	.0016721	-4.36	0.000	-.0105624	-.0040077
_cons	.9638256	.0006335	1521.49	0.000	.962584	.9650673
sigma_u	.13072244					
sigma_e	.10841536					
rho	.59247658	(fraction of variance due to u_i)				

```
. * Forma alternativa:
. xtreg everschool time##treated, fe
note: 1.treated omitted because of collinearity.

Fixed-effects (within) regression              Number of obs   =       73,291
Group variable: Id                           Number of groups =       14,748

R-squared:                                    Obs per group:
    Within = 0.0062                                min =           2
    Between = 0.0002                                avg  =          5.0
    Overall = 0.0017                                max  =           5

                                                F(2,58541)      =       183.58
corr(u_i, Xb) = -0.0105                      Prob > F         =       0.0000
```

everschool	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
1.time	.0196548	.0012939	15.19	0.000	.0171188	.0221907
treated Participa	0	(omitted)				
time#treated 1#Participa	-.007285	.0016721	-4.36	0.000	-.0105624	-.0040077
_cons	.9638256	.0006335	1521.49	0.000	.962584	.9650673
sigma_u	.13072244					
sigma_e	.10841536					
rho	.59247658	(fraction of variance due to u_i)				

```
.
. **** Paquete diff: nos hace todo esto sin escribir las regresiones
.
. *ssc install diff // Lo instalamos
.
. diff everschool, t(treated) p(time)
```

## Number of observations in the DIFF-IN-DIFF: 73291

	Before	After	
Control:	11794	17613	29407
Treated:	17604	26280	43884
	29398	43893	

6

Control	0.977			
Treated	0.981			
Diff (T-C)	0.004	0.002	2.53	0.011**
Diff-in-Diff	-0.008	0.002	3.38	0.001***

R-square: 0.00

\* Means and Standard Errors are estimated by linear regression

\*\*Inference: \*\*\* p<0.01; \*\* p<0.05; \* p<0.1

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

```
. * DiD sin variables de control
. diff everschool, t(treated) p(time)
```

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS  
Number of observations in the DIFF-IN-DIFF: 73291

	Before	After		
Control:	11794	17613	29407	
Treated:	17604	26280	43884	
	29398	43893		

---

Outcome var.	evers-l	S. Err.	t	P> t
Before				
Control	0.956			
Treated	0.969			
Diff (T-C)	0.012	0.002	6.43	0.000***
After				
Control	0.977			
Treated	0.981			
Diff (T-C)	0.004	0.002	2.53	0.011**
Diff-in-Diff	-0.008	0.002	3.38	0.001***

R-square: 0.00

\* Means and Standard Errors are estimated by linear regression

\*\*Inference: \*\*\* p<0.01; \*\* p<0.05; \* p<0.1

```
.
. * Bootstrapped std. err.:
. diff everschool, t(treated) p(time) bs rep(50)
(running regress on estimation sample)
```

Bootstrap replications (50)

1	2	3	4	5
---	---	---	---	---

..... 50

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS  
Number of observations in the DIFF-IN-DIFF: 73291

	Before	After		
Control:	11794	17613	29407	
Treated:	17604	26280	43884	
	29398	43893		

Bootstrapped Standard Errors

---

Outcome var.	evers-l	S. Err.	t	P> t
Before				
Control	0.956			
Treated	0.969			
Diff (T-C)	0.012	0.002	5.11	0.000***
After				
Control	0.977			
Treated	0.981			
Diff (T-C)	0.004	0.001	2.91	0.004***
Diff-in-Diff	-0.008	0.003	2.99	0.003***

```

R-square:    0.00
* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1

.
. * DiD con variables de control (covariates)
. diff everschool, t(treated) p(time) cov( edad hhsiz educfather educmother nodad nomom sex)
DIFFERENCE-IN-DIFFERENCES WITH COVARIATES
DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS
Number of observations in the DIFF-IN-DIFF: 58679
      Before      After
Control: 9363      13986      23349
Treated: 14162     21168      35330
      23525      35154

Outcome var. | evers-1 | S. Err. | |t| | P>|t|
-----|-----|-----|-----|-----|
Before
  Control    0.900
  Treated    0.910
  Diff (T-C) 0.010    0.002    4.50    0.000***
After
  Control    0.917
  Treated    0.920
  Diff (T-C) 0.003    0.002    1.90    0.057*
Diff-in-Diff -0.006    0.003    2.29    0.022**

R-square:    0.01
* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1

. diff everschool, t(treated) p(time) cov( edad hhsiz educfather educmother nodad nomom sex) rep
> ort
DIFFERENCE-IN-DIFFERENCES WITH COVARIATES
DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS
Number of observations in the DIFF-IN-DIFF: 58679
      Before      After
Control: 9363      13986      23349
Treated: 14162     21168      35330
      23525      35154

Report - Covariates and coefficients:
Variable(s) | Coeff. | Std. Err. | t | P>|t|
-----|-----|-----|-----|-----|
edad        0.002    0.000    9.709    0.000
hhsiz       0.002    0.000    5.964    0.000
educfather  0.004    0.000   11.947    0.000
educmother  0.004    0.000   13.769    0.000
nodad       0.000    0.000    .         .
nomom       0.000    0.000    .         .
sex        -0.004    0.001   -2.968    0.003

Outcome var. | evers-1 | S. Err. | |t| | P>|t|
-----|-----|-----|-----|-----|
Before
  Control    0.900
  Treated    0.910
  Diff (T-C) 0.010    0.002    4.50    0.000***
After
  Control    0.917
  Treated    0.920
  Diff (T-C) 0.003    0.002    1.90    0.057*
Diff-in-Diff -0.006    0.003    2.29    0.022**

R-square:    0.01
* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1

```



```

. diff everschool, t(treated) p(time) cov( edad hhsize educfather educmother nodad nomom sex) rep
> ort bs
DIFFERENCE-IN-DIFFERENCES WITH COVARIATES
(running regress on estimation sample)
Bootstrap replications (50)
----- 1 ----- 2 ----- 3 ----- 4 ----- 5
..... 50
DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS
Number of observations in the DIFF-IN-DIFF: 58679
      Before      After
Control: 9363      13986      23349
Treated: 14162     21168      35330
      23525      35154
Report - Covariates and coefficients:

```

Variable(s)	Coeff.	Std. Err.	z	P> z
edad	0.002	0.000	8.980	0.000
hhsize	0.002	0.000	6.974	0.000
educfather	0.004	0.000	11.794	0.000
educmother	0.004	0.000	16.688	0.000
nodad	0.000	0.000	.	.
nomom	0.000	0.000	.	.
sex	-0.004	0.001	-2.913	0.004

Bootstrapped Standard Errors

Outcome var.	evers-1	S. Err.	t	P> t
Before				
Control	0.900			
Treated	0.910			
Diff (T-C)	0.010	0.002	4.46	0.000***
After				
Control	0.917			
Treated	0.920			
Diff (T-C)	0.003	0.001	2.72	0.007***
Diff-in-Diff	-0.006	0.003	2.29	0.022**

```

R-square:      0.01
* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1
.
.
.

```

### 3. PROPENSITY SCORE MATCHING

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.

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).

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