

Taller de Stata para evaluación de impacto

Laura H. Atuesta Becerra

**Profesora-investigadora
Centro de Investigación y Docencia Económicas
CLEAR**

Programa social y base de datos



- El Programa se llama Canasta y fue aplicado en Colombia a familias de bajos ingresos.
- Objetivo: mejorar la nutrición de los niños de bajos ingresos a través de la provisión de despensa a las familias seleccionadas.
- Variable de participación=1 si la familia recibió despensa.
- Variables de resultados:
 - Altura según la edad de los niños menores de 10 años
 - Peso según la edad de los niños menores de 10 años

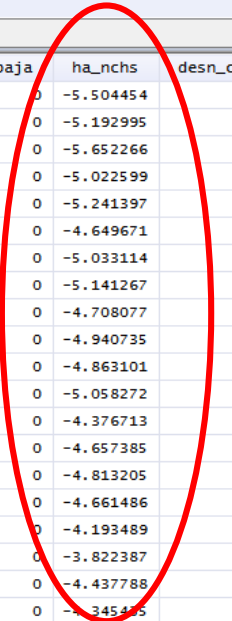
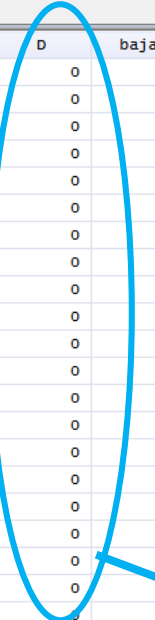
1. Modelo de diferencias (aleatoriedad)



- Supongamos que la selección de los beneficiarios se hizo de manera aleatoria.
- La única diferencia entre los dos grupos es la intervención del programa.
- El efecto se estima por medio de una regresión lineal usando:
 - Variable dependiente: estatura según la edad
 - Variable independiente: participación en el programa



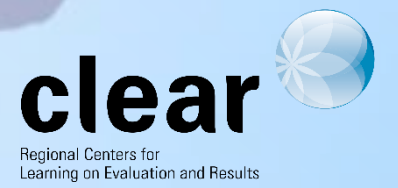
personas[1]											
personas	hombre	orden_n	D	baja	D_baja	ha_nchs	desn_cr	ocupado_jefe	educa_jefe	ingresos_h-e	
1	7	0	1	0	1	-5.504454	1	1	0	9	
2	4	1	2	0	0	-5.192995	1	1	3	0	
3	4	0	1	0	0	-5.652266	1	0	6	59	
4	4	0	1	0	0	-5.022599	1	1	5	148	
5	5	0	1	0	1	-5.241397	1	1	11	124	
6	3	0	1	0	1	-4.649671	1	1	11	61	
7	7	0	1	0	1	-5.033114	1	1	11	49	
8	4	0	1	0	0	-5.141267	1	1	5	39	
9	7	1	1	0	0	-4.708077	1	1	7	673	
10	7	0	1	0	0	-4.940735	1	0	7	29	
11	4	1	1	0	0	-4.863101	1	1	11	137	
12	6	0	1	0	1	-5.058272	1	0	5	63	
13	3	0	1	0	0	-4.376713	1	1	5	79	
14	5	0	2	0	1	-4.657385	1	1	0	0	
15	6	1	1	0	1	-4.813205	1	1	7	134	
16	5	0	2	0	1	-4.661486	1	1	3	9	
17	3	1	1	0	1	-4.193489	1	0	10	105	
18	3	0	1	0	0	-3.822387	1	1	11	94	
19	12	1	1	0	1	-4.437788	1	1	5	29	
20	3	0	1	0	0	-3.345485	1	0	11	172	
21	6	1	1	0	1	-4.2787	1	1	11	182	
22	3	0	1	0	1	-4.308841	1	1	0	19	
23	4	1	1	0	1	-4.230914	1	1	7	42.37	
24	3	0	1	0	0	-3.817977	1	1	11	19	
25	7	0	1	0	0	-3.680956	1	1	11	105	
26	5	1	2	0	1	-4.130428	1	1	10	54.2	
27	3	1	1	0	1	-3.914186	1	1	10	44	
28	4	1	1	0	1	-3.644102	1	1	5	42.3	
29	2	1	1	0	0	-3.598712	1	1	9	99	
30	3	1	1	0	0	-3.817464	1	1	11	88	
31	4	1	1	0	0	-3.622314	1	1	11	85.8	
32	4	1	1	0	0	-3.935029	1	0	11	79	
33	3	0	1	0	1	-3.642786	1	1	3	85	
34	4	1	1	0	0	-3.390474	1	1	11	69	
35	7	0	1	0	0	-3.801394	1	1	2	47.7	
36	4	1	1	0	1	-3.746871	1	1	6	43	
37	8	0	1	0	0	-3.811333	1	1	5	61	
38	5	0	1	0	0	-3.514508	1	1	6	36	
39	3	0	1	0	1	-3.592781	1	1	2	21	
40	8	1	2	0	0	-3.045791	1	1	2	54	
41	6	1	1	0	1	-3.481305	1	1	5	0	
42	5	1	1	0	1	-3.458964	1	1	3	47	
43	5	0	1	0	0	-3.119109	1	1	1	0	
44	5	1	1	0	0	-3.184862	1	1	7	0	



Variable de resultados

Así es como se ve la base de datos:
Número de variables: 11
Número de observaciones: 4,000

Variable de participación





#	Command	_rc
1	use "C:\Users\lau...	
2	browse	
3	reg ha_nchs D	

```

(R)
-----
Statistics/Data Analysis

Special Edition

13.1 Copyright 1985-2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

25-student Stata lab perpetual license:
Serial number: 401306201775
Licensed to: Rafael Garduno
Centro de Investigacion y Docencia Economicas

Notes:
1. (/v# option or -set maxvar-) 5000 maximum variables

. use "C:\Users\laura\Dropbox\evaluacion de politicas publicas 2016\talleres de evaluacion
. browse
. reg ha_nchs D

Source | SS df MS | Number of obs = 4000
-----+-----+-----+| F( 1, 3998) = 34.32
Model | 54.9104328 1 54.9104328 | Prob > F = 0.0000
Residual | 6396.72078 3998 1.59998018 | R-squared = 0.0085
Total | 6451.63121 3999 1.61931113 | Adj R-squared = 0.0083
| | | | Root MSE = 1.2649

ha_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]
-----+-----+-----+-----+-----
D | .2343298 .0399998 5.86 0.000 0.1559079 .3127516
_cons | -0.460041 .0282841 -29.91 0.000 -0.9014567 -.7905515
    
```

Los niños que participaron en el programa tienen una estatura de 0.23 desviaciones estándares por encima de la media MAYOR que los niños que no participaron en el programa.

Este resultado es estadísticamente significativo, lo que significa que este comportamiento se puede generalizar a la población.

2. Modelo de diferencias en diferencias (eventos fortuitos)



- Supongamos que en algunos municipios del país encontraron petróleo y esto incrementó las regalías. El gobierno decidió implementar el Programa Canasta en estos municipios.
- Se comparan los resultados nutricionales de los niños en municipios en donde encontraron petróleo y donde no encontraron petróleo.
- Puede haber diferencias pre-existentes entre los dos grupos. Puede que las poblaciones no sean iguales.
 - Controlamos por estas diferencias a través de la metodología dif-en-dif.

2. Modelo de diferencias en diferencias (eventos fortuitos)



- Pasos:
- Generamos una nueva variable que sea la diferencia entre la variable de resultados del periodo 2 y el periodo 1:

```
gen delta_ha_nchs= ha_nchs2- ha_nchs1
```

- Estimamos la regresión usando esta nueva variable como variable dependiente.

```
reg delta_ha_nchs D ingresos_hogar_jefe1 educa_jefe1
```

Stata/SE 13.1 - C:\Users\laura\Dropbox\evaluacion de politicas publicas 2016\talleres de evaluacion de impacto\2 experimentos naturales\experimentos_naturales

File Edit Data Graphics Statistics User Window Help

Review # Command _rc

```

1 use "C:\Users\lau...
2 gen delta_ha_nc...
3 reg delta_ha_nch...

```

Statistics/Data Analysis

Special Edition

13.1 Copyright 1985-2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

25-student Stata lab perpetual license:
Serial number: 401306201775
Licensed to: Rafael Garduno
Centro de Investigacion y Docencia Economicas

Notes:
1. (/v# option or -set maxvar-) 5000 maximum variables

```

. use "C:\Users\laura\Dropbox\evaluacion de politicas publicas 2016\talleres de evaluacion de im
. gen delta_ha_nchs= ha_nchs2- ha_nchs1
. reg delta_ha_nchs D ingresos_hogar_jefel educa_jefel

```

Source	SS	df	MS			
Model	41.3464811	3	13.7821604	Number of obs =	4000	
Residual	362.01623	3996	.090594652	F(3, 3996) =	152.13	
Total	403.362711	3999	.100865894	Prob > F =	0.0000	
				R-squared =	0.1025	
				Adj R-squared =	0.1018	
				Root MSE =	.30099	

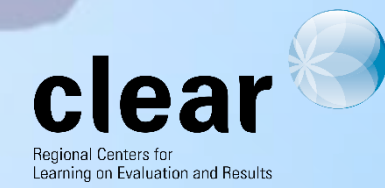
delta_ha_nchs	Coef.	Std. Err	t	P> t	[95% Conf. Interval]
D	.2016052	.0095246	21.17	0.000	.1829316 .2202788
ingresos_hogar_jefel	.0001276	.0000504	2.26	0.024	.0000017 .0002382
educa_jefel	-.0006636	.0012317	-0.54	0.590	-.0030784 .0017512
_cons	.1922858	.0109107	17.62	0.000	.1708948 .2136769



Creación de la nueva variable que nos va a servir como variable de resultados.

Nuevo impacto del programa: los niños que recibieron el Programa tienen una estatura de 0.20 desviaciones estándares por encima de la media más arriba que los niños que no recibieron el programa.

Resultados estadísticamente significativos



3. Modelo de emparejamiento (cuando solo lo que observamos importa...)



- Primero tenemos que seleccionar las variables para hacer el emparejamiento. Podemos hacer varias pruebas en Stata para ver qué variables son significativas.
 - Ingresos del jefe de hogar
 - Número de personas en el hogar
 - Orden en el cual se encuentra el niño
 - Educación del jefe de hogar
 - Si el jefe de hogar trabaja en empleo formal
 - Sexo del jefe de hogar
- Para hacer esta selección podemos regresar la variable de participación como variable dependiente y las características como variables independientes. Todas deben ser significativas en explicar la participación.

```

Review
# Command _rc
1 use "C:\Users\lau...
2 dprobit D ingres...

(R)
-----
Statistics/Data Analysis
Special Edition

13.1 Copyright 1985-2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

25-student Stata lab perpetual license:
Serial number: 401306201775
Licensed to: Rafael Garduno
Centro de Investigacion y Docencia Economicas

Notes:
1. (/v# option or -set maxvar-) 5000 maximum variables

. use "C:\Users\laura\Dropbox\evaluacion de politicas publicas 2016\talleres de eva...
. dprobit D ingresos_hogar_jefe personas orden_n educa_jefe ocupado_jefe hombre

Iteration 0: log likelihood = -2771.4388
Iteration 1: log likelihood = -2739.1529
Iteration 2: log likelihood = -2739.1378

Probit regression, reporting marginal effects
Number of obs = 4000
LR chi2(6) = 64.60
Prob > chi2 = 0.0000
Pseudo R2 = 0.0117
Log likelihood = -2739.1378

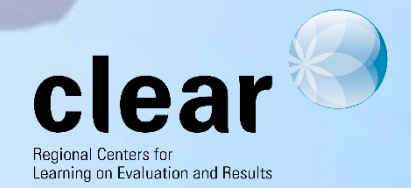
-----
+-----+-----+-----+-----+-----+-----+
| D | dF/dx | Std. Err. | z | P>|z| | x-bar [ 95% C.I. ] |
+-----+-----+-----+-----+-----+-----+
| ingresos_e | .0003089 | .0000954 | 3.24 | 0.001 | 72.9454 | .000122 | .000496 |
| personas | -.0217877 | .0042754 | -5.10 | 0.000 | 3.04525 | .030167 | -.013408 |
| orden_n | .0912844 | .0246782 | 3.70 | 0.000 | 1.00675 | .042916 | .139653 |
| educa_e | -.0060159 | .0021695 | -2.77 | 0.006 | 6.57575 | -.010268 | -.001764 |
| ocupad_e* | .0814328 | .0205396 | 3.93 | 0.000 | .809 | .041176 | .12169 |
| hombre* | .016465 | .015883 | 1.04 | 0.300 | .511 | -.014665 | .047595 |
+-----+-----+-----+-----+-----+
| obs. P | .488 |
| pred. P | .4877964 (at x-bar) |
+-----+-----+-----+-----+

```



Regresamos la participación (D) explicada por las características del hogar.

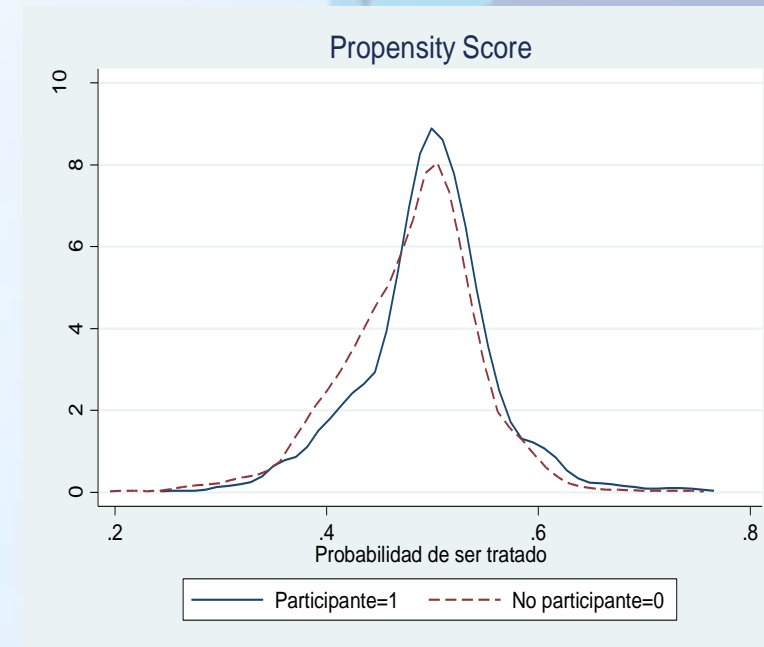
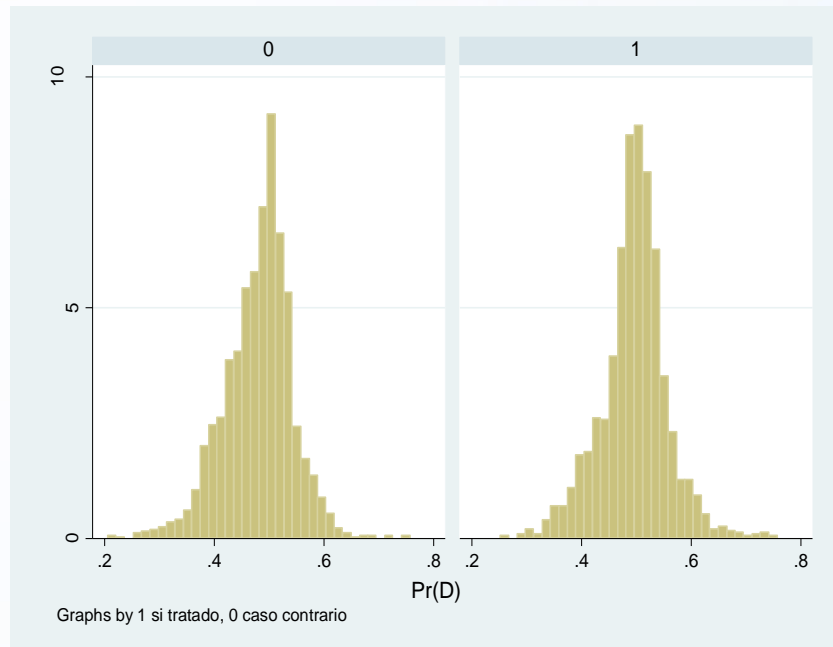
Todas las características son estadísticamente significativas en explicar la participación. Por ejemplo, una persona más en el hogar disminuye la probabilidad de participación en 2%.



(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

3. Modelo de emparejamiento (cuando solo lo que observamos importa...)

- Después sacamos el soporte común a través de un comando que se llama “dprobit”.
- También podemos dibujar los histogramas del soporte común y su densidad:



3. Modelo de emparejamiento (cuando solo lo que observamos importa...)



- Finalmente, calculamos el efecto del programa por medio de un comando llamado "psmatch2".

```
set seed 50
```

```
drawnorm orden
```

```
sort orden
```

```
global X "personas orden_n ocupado_jefe educa_jefe ingresos_hogar_jefe"
```

```
psmatch2 D $X, outcome(ha_nchs2) n(1) com
```



```

opbox\evaluacion de politicas publicas 2016\talleres de impacto\3 metodo de emparejamiento\emparejamiento_base.d
statistics User Window Help
r(110);
.
. sort orden
.
. global X "personas orden_n ocupado_jefe educa_jefe ingresos_hogar_jefe"
.
. psmatch2 D $X, outcome(ha_nchs2) n(1) com

Probit regression              Number of obs   =      4000
                              LR chi2(5)        =      63.52
                              Prob > chi2         =      0.0000
                              Pseudo R2          =      0.0115

Log likelihood = -2739.6749

```

D	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
personas	-.0547807	.0107191	-5.11	0.000	-.0758897	-.0337717
orden_n	.2290269	.0618843	3.70	0.000	.1077359	.3503178
ocupado_jefe	.2053767	.0522936	3.93	0.000	.102883	.3078703
educa_jefe	-.0151349	.00544	-2.78	0.005	-.0257972	-.0044727
ingresos_hogar_jefe	.0007741	.0002392	3.24	0.001	.0003014	.0012429
_cons	-.1307518	.104657	-1.25	0.212	-.3358158	.0743722

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
ha_nchs2	Unmatched	-.642856900	-.970360915	.332497022	.03380015	9.84
	ATT	-.643655504	-.988453676	.344798172	.047459317	7.27

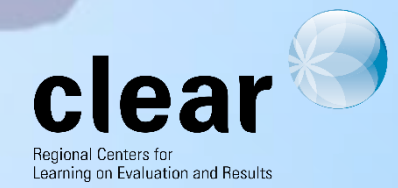
Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support			Total
	Off suppo	On suppor		
Untreated	0	2,048		2,048
Treated	2	1,950		1,952
Total	2	3,998		4,000



Primer paso para seleccionar el soporte común. Son los mismos resultados que observamos anteriormente.

Impacto del programa: los niños que recibieron el programa están 0.34 desviaciones estándares por encima de la media más arriba que los niños que no recibieron el programa.



4. Variables instrumentales



Asumimos que existen dos instrumentos válidos para explicar la participación en el programa Canasta.

1. El primero es la distancia de los hogares a las oficinas de inscripción.
2. El segundo es el número de oficinas disponible en cada municipio.

4. Variables instrumentales



PASOS

1. En el primer paso estimamos el efecto del instrumento en la probabilidad de participar.

$$D_i = \beta_0 + \beta_1 Z_i + \beta_2 X_i + \varepsilon_i$$

2. SEGUNDA ETAPA: utilizamos los resultados de la primera etapa para estimar el impacto del programa.

$$Y_i = \gamma_0 + \gamma_1 \hat{D}_i + \gamma_2 X_i + u_i$$

PRIMERA ETAPA: REGRESAR D EN FUNCIÓN DE Z



```
Dropbox\evaluacion de politicas publicas 2016\talleres de evaluacion de impacto\4 variables instrumentales\variables_instrumenta
Statistics User Window Help
. global X "personas orden_n ocupado_jefe educa_jefe ingresos_hogar_jefe"
. ivreg2 ha_nchs $X (D=distancia), first

First-stage regressions
-----
First-stage regression of D:
-----
OLS estimation
-----
Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

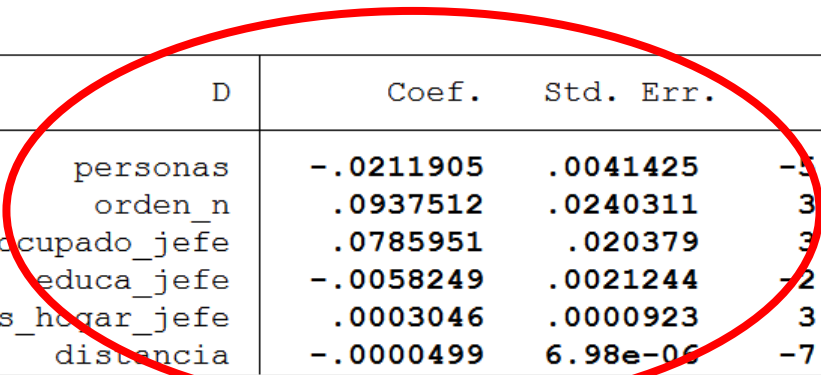
Total (centered) SS      =      999.424
Total (uncentered) SS  =      1952
Residual SS            =     971.3568451

Number of obs =      4000
F( 6, 3993) =     19.23
Prob > F      =     0.0000
Centered R2   =     0.0281
Uncentered R2 =     0.5024
Root MSE     =     .4932

+-----+-----+-----+-----+-----+-----+
|          D          | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
+-----+-----+-----+-----+-----+-----+
| personas            | -0.0211905 | .0041425 | -5.12 | 0.000 | -0.0293121  -0.013069 |
| orden_n             |  .0937512 | .0240311 |  3.90 | 0.000 |  .0466369   .1408655 |
| ocupado_jefe        |  .0785951 | .020379  |  3.86 | 0.000 |  .0386409   .1185493 |
| educa_jefe          | -0.0058249 | .0021244 | -2.74 | 0.006 | -0.0099899  -0.0016598 |
| ingresos_hogar_jefe |  .0003046 | .0000923 |  3.30 | 0.001 |  .0001237   .0004856 |
| distancia           | -0.0000499 | 6.98e-06 | -7.14 | 0.000 | -0.0000635  -0.0000362 |
+-----+-----+-----+-----+-----+-----+

```

Resultados de la primera etapa.
Variable dependiente: D
Variable independiente: Z y otras características de los hogares.
Z es significativo al 99% de confianza.



SEGUNDA ETAPA: ESTIMAR EL IMPACTO DEL PROGRAMA



\\laura\Dropbox\evaluacion de politicas publicas 2016\talleres de evaluacion de impacto\4 variables instrumentales\variables_instrumenta

Statistics User Window Help

IV (2SLS) estimation

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

Number of obs = 4000
F(6, 3993) = 4.64
Prob > F = 0.0001
Centered R2 = 0.0464
Uncentered R2 = 0.0567
Root MSE = .5416

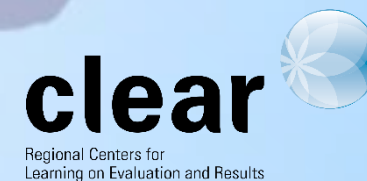
Total (centered) SS = 1230.139487
Total (uncentered) SS = 1243.678874
Residual SS = 1173.111503

ha_inchs	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
D	.1390646	.1537524	0.90	0.366	-.1622846	.4404138
personas	-.0117333	.0056077	-2.09	0.036	-.0227241	-.0007424
orden_n	-.053356	.029773	-1.79	0.073	-.1117099	.004998
ocupado_jefe	-.0275681	.0255749	-1.08	0.281	-.077694	.0225578
educa_jefe	.0053271	.0025021	2.13	0.033	.0004231	.010231
ingresos_hogar_jefe	.0001958	.0001113	1.76	0.079	-.0000224	.0004139
_cons	-.0348009	.0821434	-0.42	0.672	-.1957989	.1261971

Underidentification test (Anderson canon. corr. LM statistic): 50.443
Chi-sq(1) P-val = 0.0000

Weak identification test (Cragg-Donald Wald F statistic): 50.998
Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38
15% maximal IV size 8.96
20% maximal IV size 6.66

Resultados de la segunda etapa.
Variable dependiente: estatura según la edad.
El impacto del programa está dado por D.
Los niños participantes tienen una altura adicional de 0.13 desviaciones estándares por encima según la edad comparado con los niños que no participaron en el programa.



PRUEBAS DE RELEVANCIA Y EXOGENEIDAD PARA LOS INSTRUMENTOS



opbox\evaluacion de politicas publicas 2016\talleres de evaluacion de impacto\4 variables instrumentales\variables_instrumenta

Statistics User Window Help

Variable	Shea Partial R2	Partial R2	F(1, 3993)	P-value
D	0.0126	0.0126	51.00	0.0000

Underidentification tests
 Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
 Ha: matrix has rank=K1 (identified)
Anderson canon. corr. N*CCEV LM statistic Chi-sq(1)=50.44 P-val=0.0000
Cragg-Donald N*CDEV Wald statistic Chi-sq(1)=51.09 P-val=0.0000

Weak identification test
 Ho: equation is weakly identified
Cragg-Donald Wald F-statistic 51.00
 See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference
 Tests of joint significance of endogenous regressors B1 in main equation
 Ho: B1=0 and overidentifying restrictions are valid
Anderson-Rubin Wald test F(1,3993)=0.78 P-val=0.3760
Anderson-Rubin Wald test Chi-sq(1)=0.79 P-val=0.3755
Stock-Wright LM S statistic Chi-sq(1)=0.79 P-val=0.3756

Number of observations N = 4000
 Number of regressors K = 7
 Number of instruments L = 7
 Number of excluded instruments L1 = 1

opbox\evaluacion de politicas publicas 2016\talleres de evaluacion de impacto\4 variables instrumentales\variables_instrumenta

Statistics User Window Help

ha_nchs	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
D	.1390646	.1537524	0.90	0.366	-.1622846 .4404138
personas	-.0117333	.0056077	-2.09	0.036	-.0227241 -.0007424
orden_n	-.053356	.029773	-1.79	0.073	-.1117099 .004998
ocupado_jefe	-.0275681	.0255749	-1.08	0.281	-.077694 .0225578
educa_jefe	.0053271	.0025021	2.13	0.033	.0004231 .010231
ingresos_hogar_jefe	.0001958	.0001113	1.76	0.079	-.0000224 .0004139
_cons	-.0348009	.0821434	-0.42	0.672	-.1957989 .1261971

Underidentification test (Anderson canon. corr. LM statistic): 50.443
 Chi-sq(1) P-val = 0.0000

Weak identification test (Cragg-Donald Wald F statistic): 50.998
 Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38
 15% maximal IV size 8.96
 20% maximal IV size 6.66
 25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

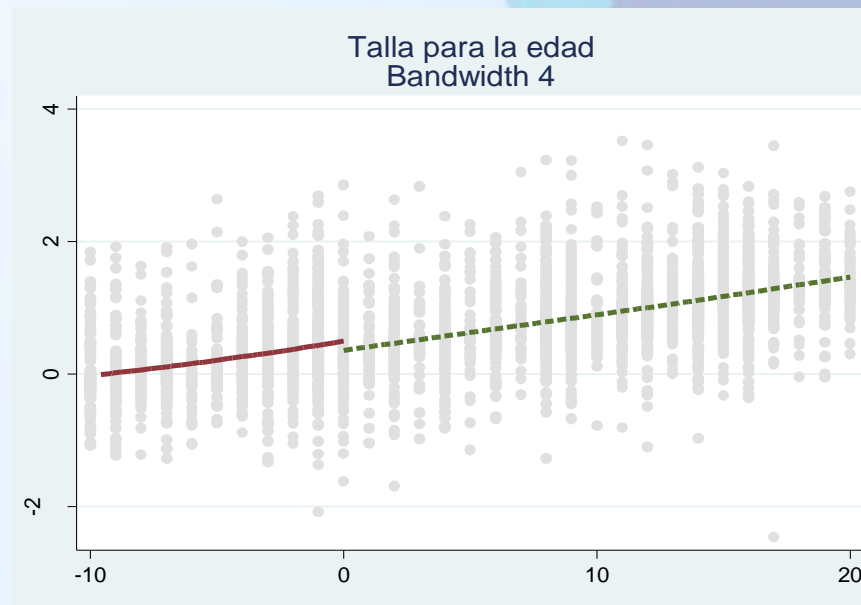
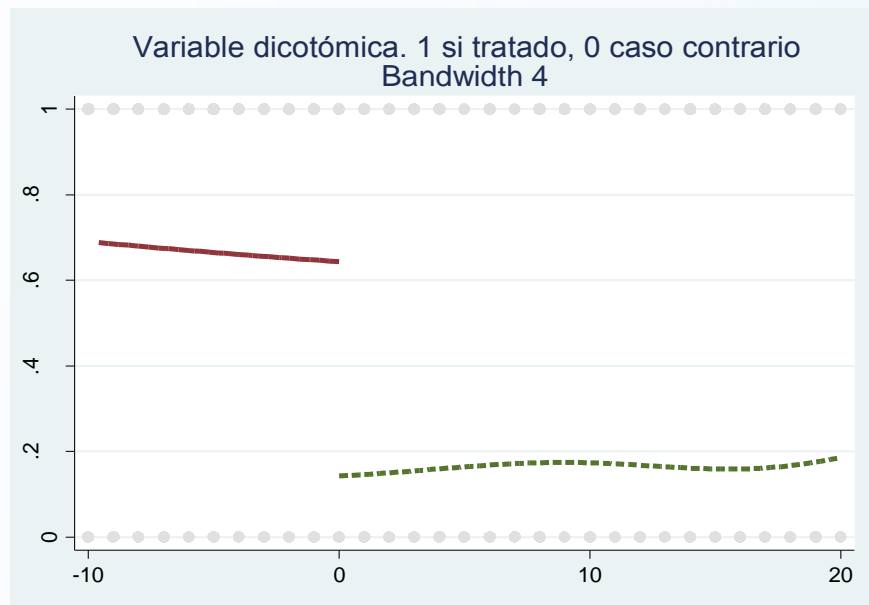
Sargan statistic (overidentification test of all instruments): 0.000
 (equation exactly identified)

Instrumented: D
 Included instruments: personas orden_n ocupado_jefe educa_jefe
 ingresos_hogar_jefe
 Excluded instruments: distancia

5. Regresión discontinua

Asumimos que el programa está disponible para hogares que tienen un Sisben menor a 10.

Le tenemos que especificar a Stata la variable de resultados, la variable de participación, la variable que mide nuestro umbral, el valor en donde observamos la discontinuidad y el peso que queremos darle a cada observación.



**Special Edition**

4905 Lakeway Drive
 College Station, Texas 77845 USA
 800-STATA-PC <http://www.stata.com>
 979-696-4600 stata@stata.com
 979-696-4601 (fax)

25-student Stata lab perpetual license:

Serial number: 401306201775

Licensed to: Rafael Garduno

Centro de Investigacion y Docencia Economicas

Notes:

1. (/v# option or -set maxvar-) 5000 maximum variables

```
. use "C:\Users\laura\Dropbox\evaluacion de politicas publicas 2016\talleres de evaluacion discontinua\regresion discontinua_base.dta"
```

```
rd_obs ha_nchs D sisben, bw(4) mbw(100) k(gau) z0(10)
```

Three variables specified; jump in treatment at Z=10 will be estimated. Local Wald Estimate is the ratio of jump in outcome to jump in treatment.

Assignment variable Z is sisben

Treatment variable X_T is D

Outcome variable y is ha_nchs

Command used: lpoly; Kernel used: gau
 Bandwidth: 4; Local Wald Estimate: .28356266



Comando para estimar la regresión discontinua.
 Variable de resultados: ha_nchs
 Variable de participación: D
 Umbral: sisben
 ventana: 4 puntos
 Pesos de kernel a las observaciones
 La discontinuidad se presenta en sisben=10

Los niños participantes tienen una altura de 0.28 desviaciones estándares adicionales por encima de la media comparado con los niños que no participaron en el programa.

clear

Regional Centers for
 Learning on Evaluation and Results

