. \*----------------------------------\*

. \* Modelo de diferencias \*

. \*----------------------------------\*

Algunos puntos antes de hacer el análisis:

El símbolo asterisco () reconoce una linea del do-file como comentarios

y cuando se corre el do-file las líneas precedidas por este símbolo no son

reconocidas como comandos.

Lo primero que debemos hacer es abrir la base de datos del modelo de diferencias.

Para simplificar los comandos definimos un vector X compuesto de las principales características del hogar.

global X "personas hombre orden\_n baja ocupado\_jefe educa\_jefe ingresos\_hogar\_jefe"

.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \*\*\*\* Talla para la edad\*\*\*\*

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

.

. \*--------------------------------------\*

. \*1. Estadísticas descriptivas

. \*--------------------------------------\*

.

.

. sum ha\_nchs, detail

Talla para la edad

-------------------------------------------------------------

Percentiles Smallest

1% -3.936628 -5.852884

5% -2.535352 -5.652266

10% -2.09676 -5.504454 Obs 4000

25% -1.532769 -5.456038 Sum of Wgt. 4000

50% -.7905374 Mean -.7288392

Largest Std. Dev. 1.270162

75% -.0348815 4.856632

90% .7795621 5.004343 Variance 1.613311

95% 1.317513 5.941913 Skewness .4067137

99% 3.087887 5.994383 Kurtosis 5.273376

Podemos ver algunas características de la variable ha\_nchs. Entre estas, el máximo, mínimo, desviación estándar y promedio.

. \*--------------------------------------\*

. \*2. Estadísticas descriptivas por grupo

. \*--------------------------------------\*

.

. sum ha\_nchs if D==1, detail

Talla para la edad

-------------------------------------------------------------

Percentiles Smallest

1% -3.805689 -5.852884

5% -2.442669 -5.456038

10% -2.026661 -5.141572 Obs 2000

25% -1.418647 -5.092871 Sum of Wgt. 2000

50% -.6427878 Mean -.6116743

Largest Std. Dev. 1.265438

75% .1056483 4.732728

90% .8861872 4.856632 Variance 1.601334

95% 1.424549 5.004343 Skewness .3645176

99% 3.135453 5.941913 Kurtosis 5.138183

.

. sum ha\_nchs if D==0, detail

Talla para la edad

-------------------------------------------------------------

Percentiles Smallest

1% -4.212201 -5.652266

5% -2.660601 -5.504454

10% -2.155744 -5.241397 Obs 2000

25% -1.636544 -5.192995 Sum of Wgt. 2000

50% -.9208255 Mean -.8460041

Largest Std. Dev. 1.264368

75% -.1738329 4.510983

90% .6514927 4.622804 Variance 1.598626

95% 1.216139 4.758943 Skewness .4590554

99% 2.929789 5.994383 Kurtosis 5.522551

.

. \*-------------------------------------------------------------------\*

. \*3.Regresión con "D" siendo la única variable explicativa \*

. \*-------------------------------------------------------------------\*

.

. 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

-------------+------------------------------ Adj R-squared = 0.0083

Total | 6451.63121 3999 1.61331113 Root MSE = 1.2649

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2343298 .0399998 5.86 0.000 .1559079 .3127516

\_cons | -.8460041 .0282841 -29.91 0.000 -.9014567 -.7905515

------------------------------------------------------------------------------

Mediante la estimación de MCO podemos ver una prueba de diferencia de medias entre los dos grupos (tratamiento y control), sin controles adicionales.

. \*------------------------------------------------\*

. \*4. Regresión adicionando variables explicativas

. \*------------------------------------------------\*

.

.

. reg ha\_nchs D $X

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 8, 3991) = 26.95

Model | 330.694078 8 41.3367597 Prob > F = 0.0000

Residual | 6120.93713 3991 1.53368507 R-squared = 0.0513

-------------+------------------------------ Adj R-squared = 0.0494

Total | 6451.63121 3999 1.61331113 Root MSE = 1.2384

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2305058 .0391847 5.88 0.000 .1536818 .3073297

personas | -.0458019 .0101454 -4.51 0.000 -.0656926 -.0259112

hombre | .0057317 .0391963 0.15 0.884 -.0711149 .0825784

orden\_n | -.1667524 .0624656 -2.67 0.008 -.2892198 -.0442849

baja | -.4218382 .0391912 -10.76 0.000 -.4986749 -.3450016

ocupado\_jefe | -.0485502 .048766 -1.00 0.320 -.1441587 .0470584

educa\_jefe | .0190576 .0052908 3.60 0.000 .0086846 .0294306

ingresos\_h~e | .0005315 .0002394 2.22 0.026 .0000621 .0010009

\_cons | -.3399708 .1083648 -3.14 0.002 -.5524263 -.1275153

------------------------------------------------------------------------------

Dado que en realida la variable talla para la edad depende de factores adicionales como el ingreso del jefe de hogar, el orden de nacimiento, la educación del jefe de hogar y el número de personas en el hogar, entre otras, realizamos la prueba de diferencia de medias entre los dos grupos pero ahora controlando por estas variables.

. \*---------------------------------------------------------------------\*

. \*5. Estimador de diferencias con efectos heterogéneos \*

. \*---------------------------------------------------------------------\*

.

.

. reg ha\_nchs D $X D\_baja

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 9, 3990) = 24.65

Model | 339.775832 9 37.7528702 Prob > F = 0.0000

Residual | 6111.85538 3990 1.53179333 R-squared = 0.0527

-------------+------------------------------ Adj R-squared = 0.0505

Total | 6451.63121 3999 1.61331113 Root MSE = 1.2377

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .3258111 .0553676 5.88 0.000 .2172597 .4343625

personas | -.0462821 .0101411 -4.56 0.000 -.0661642 -.0263999

hombre | .0046569 .0391746 0.12 0.905 -.0721473 .081461

orden\_n | -.1663797 .0624272 -2.67 0.008 -.288772 -.0439874

baja | -.3264538 .0553952 -5.89 0.000 -.4350593 -.2178484

ocupado\_jefe | -.0491834 .0487366 -1.01 0.313 -.1447344 .0463675

educa\_jefe | .0191783 .0052878 3.63 0.000 .0088113 .0295454

ingresos\_h~e | .0005296 .0002393 2.21 0.027 .0000605 .0009987

D\_baja | -.1906753 .0783086 -2.43 0.015 -.344204 -.0371466

\_cons | -.3852582 .1098834 -3.51 0.000 -.6006911 -.1698253

------------------------------------------------------------------------------

Al incorporar la variable "D\_baja"="D"\*"baja" podemos encontrar que el tratamiento tiene efectos heterogeneos sobre los individuos dependiendo de la raza a la que pertenecen.

. \*--------------------------------------------------\*

. \*6. Prueba estadística "D-D\_baja=0".

. \*--------------------------------------------------\*

.

. reg ha\_nchs D $X D\_baja

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 9, 3990) = 24.65

Model | 339.775832 9 37.7528702 Prob > F = 0.0000

Residual | 6111.85538 3990 1.53179333 R-squared = 0.0527

-------------+------------------------------ Adj R-squared = 0.0505

Total | 6451.63121 3999 1.61331113 Root MSE = 1.2377

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .3258111 .0553676 5.88 0.000 .2172597 .4343625

personas | -.0462821 .0101411 -4.56 0.000 -.0661642 -.0263999

hombre | .0046569 .0391746 0.12 0.905 -.0721473 .081461

orden\_n | -.1663797 .0624272 -2.67 0.008 -.288772 -.0439874

baja | -.3264538 .0553952 -5.89 0.000 -.4350593 -.2178484

ocupado\_jefe | -.0491834 .0487366 -1.01 0.313 -.1447344 .0463675

educa\_jefe | .0191783 .0052878 3.63 0.000 .0088113 .0295454

ingresos\_h~e | .0005296 .0002393 2.21 0.027 .0000605 .0009987

D\_baja | -.1906753 .0783086 -2.43 0.015 -.344204 -.0371466

\_cons | -.3852582 .1098834 -3.51 0.000 -.6006911 -.1698253

------------------------------------------------------------------------------

. test D+D\_baja=0

( 1) D + D\_baja = 0

F( 1, 3990) = 5.95

Prob > F = 0.0147

Al correr la anterior regresión pudimos ver que el impacto del tratamiento disminuye si el individuo pertenece a la raza "baja". Sin embargo, al probar que "D+D\_baja" es diferente de cero, y al encontrar que "D" y "D\_baja" son significativas de manera individual, encontramos que pese a que el programa tiene un efecto reducido sobre la talla para la edad en los individuos de raza baja, este efecto es positivo.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \*\*\*\* Desnutrición Crónica\*\*\*\*

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. \*----------------------------\*

. \*1. Estadísticas descriptivas

. \*----------------------------\*

.

.

. sum desn\_cr

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

desn\_cr | 4000 .1175 .3220554 0 1

Por medio de este comando podemos verificar que

aproximadamente 11,75% de los individuos en la muestra sufren

de desnutrición crónica.

.

.

. \*----------------------\*

. \*2. Promedios por grupo

. \*----------------------\*

.

.

. sum desn\_cr if D==0

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

desn\_cr | 2000 .1315 .3380309 0 1

.

. sum desn\_cr if D==1

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

desn\_cr | 2000 .1035 .304687 0 1

Encontramos los datos para desnutrición crónica para el grupo de tratamiento y control.

. \*--------------------------------------------------------------------\*

. \*3. Regresión con "D" siendo la única variable explicativa

. \*--------------------------------------------------------------------\*

.

. dprobit desn\_cr D

Iteration 0: log likelihood = -1447.6566

Iteration 1: log likelihood = -1443.8703

Iteration 2: log likelihood = -1443.8683

Probit regression, reporting marginal effects Number of obs = 4000

LR chi2(1) = 7.58

Prob > chi2 = 0.0059

Log likelihood = -1443.8683 Pseudo R2 = 0.0026

------------------------------------------------------------------------------

desn\_cr | dF/dx Std. Err. z P>|z| x-bar [ 95% C.I. ]

---------+--------------------------------------------------------------------

D\*| -.028 .0101734 -2.75 0.006 .5 -.047939 -.008061

---------+--------------------------------------------------------------------

obs. P | .1175

pred. P | .1169067 (at x-bar)

------------------------------------------------------------------------------

(\*) 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

.

. logit desn\_cr D

Iteration 0: log likelihood = -1447.6566

Iteration 1: log likelihood = -1443.8784

Iteration 2: log likelihood = -1443.8683

Logistic regression Number of obs = 4000

LR chi2(1) = 7.58

Prob > chi2 = 0.0059

Log likelihood = -1443.8683 Pseudo R2 = 0.0026

------------------------------------------------------------------------------

desn\_cr | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | -.2711659 .0988255 -2.74 0.006 -.4648604 -.0774715

\_cons | -1.887761 .066166 -28.53 0.000 -2.017444 -1.758078

------------------------------------------------------------------------------

.

. mfx

Marginal effects after logit

y = Pr(desn\_cr) (predict)

= .11677369

------------------------------------------------------------------------------

variable | dy/dx Std. Err. z P>|z| [ 95% C.I. ] X

---------+--------------------------------------------------------------------

D\*| -.028 .01017 -2.75 0.006 -.047939 -.008061 .5

------------------------------------------------------------------------------

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

.

Al igual que con la talla para la edad, lo primero que hacemos es una prueba de diferencia de medias sin controlar por otras variables.

Dado que la variable dependiente es dicotómica, utilizamos un modelo probit y un modelo logit, respectivamente.

Es importante recordar que la diferencia entre estos dos modelos es la distribución que suponen para la procedencia de los datos. Mientras que el modelo probit supone una distribución normal, logit supone una logística.

El comando "dprobit" nos permite identificar los efectos marginales en probabilidad. "mfx" hace lo mismo para el modelo logit.

.

. \*-----------------------------------------------\*

. \*4. Regresión adicionando variables explicativas

. \*-----------------------------------------------\*

.

.

. dprobit desn\_cr D $X

Iteration 0: log likelihood = -1447.6566

Iteration 1: log likelihood = -1423.8892

Iteration 2: log likelihood = -1423.7972

Iteration 3: log likelihood = -1423.7972

Probit regression, reporting marginal effects Number of obs = 4000

LR chi2(8) = 47.72

Prob > chi2 = 0.0000

Log likelihood = -1423.7972 Pseudo R2 = 0.0165

------------------------------------------------------------------------------

desn\_cr | dF/dx Std. Err. z P>|z| x-bar [ 95% C.I. ]

---------+--------------------------------------------------------------------

D\*| -.0284681 .0100651 -2.82 0.005 .5 -.048195 -.008741

personas | .010929 .0024446 4.46 0.000 5.136 .006138 .01572

hombre\*| .0030032 .0100547 0.30 0.765 .51625 -.016704 .02271

orden\_n | .0116865 .0153682 0.76 0.447 1.09925 -.018435 .041808

baja\*| .0207448 .0100692 2.06 0.039 .5 .00101 .04048

ocupad~e\*| .0076168 .0123024 0.61 0.542 .79775 -.016495 .031729

educa\_~e | -.002926 .0013645 -2.14 0.032 6.40325 -.0056 -.000252

ingres~e | -.0001034 .0000644 -1.60 0.109 73.5289 -.00023 .000023

---------+--------------------------------------------------------------------

obs. P | .1175

pred. P | .113883 (at x-bar)

------------------------------------------------------------------------------

(\*) 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

. logit desn\_cr D $X

Iteration 0: log likelihood = -1447.6566

Iteration 1: log likelihood = -1424.5692

Iteration 2: log likelihood = -1423.949

Iteration 3: log likelihood = -1423.9486

Logistic regression Number of obs = 4000

LR chi2(8) = 47.42

Prob > chi2 = 0.0000

Log likelihood = -1423.9486 Pseudo R2 = 0.0164

------------------------------------------------------------------------------

desn\_cr | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | -.2681305 .0994012 -2.70 0.007 -.4629532 -.0733078

personas | .1050214 .0231686 4.53 0.000 .0596117 .150431

hombre | .0267168 .098932 0.27 0.787 -.1671864 .2206199

orden\_n | .1200175 .1458207 0.82 0.410 -.1657858 .4058208

baja | .2035689 .0992309 2.05 0.040 .00908 .3980579

ocupado\_jefe | .0746175 .1252646 0.60 0.551 -.1708966 .3201315

educa\_jefe | -.028599 .0135068 -2.12 0.034 -.0550719 -.0021262

ingresos\_h~e | -.0011171 .0006841 -1.63 0.103 -.0024579 .0002238

\_cons | -2.50544 .2643007 -9.48 0.000 -3.02346 -1.987421

------------------------------------------------------------------------------

.

. mfx

Marginal effects after logit

y = Pr(desn\_cr) (predict)

= .11321168

------------------------------------------------------------------------------

variable | dy/dx Std. Err. z P>|z| [ 95% C.I. ] X

---------+--------------------------------------------------------------------

D\*| -.0269509 .00996 -2.70 0.007 -.04648 -.007422 .5

personas | .0105436 .0023 4.58 0.000 .006028 .01506 5.136

hombre\*| .0026814 .00993 0.27 0.787 -.016772 .022135 .51625

orden\_n | .0120491 .01463 0.82 0.410 -.016634 .040732 1.09925

baja\*| .0204513 .00996 2.05 0.040 .000939 .039964 .5

ocupad~e\*| .0073639 .01215 0.61 0.544 -.016446 .031174 .79775

educa\_~e | -.0028712 .00135 -2.12 0.034 -.00552 -.000222 6.40325

ingres~e | -.0001121 .00007 -1.64 0.102 -.000246 .000022 73.5289

------------------------------------------------------------------------------

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

.

.

Ahora hacemos la diferencia de medias pero controlando por más variables. Entre estas la educación del jefe del hogar, el orden de nacimiento del individuo, los ingresos del jefe del hogar, género del individuo, raza del individuo etc.

.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

. \*\*\*\* Verificación de aleatorización \*\*\*\*

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

.

. dprobit D $X

Iteration 0: log likelihood = -2772.5887

Iteration 1: log likelihood = -2770.2976

Iteration 2: log likelihood = -2770.2976

Probit regression, reporting marginal effects Number of obs = 4000

LR chi2(7) = 4.58

Prob > chi2 = 0.7108

Log likelihood = -2770.2976 Pseudo R2 = 0.0008

------------------------------------------------------------------------------

D | dF/dx Std. Err. z P>|z| x-bar [ 95% C.I. ]

---------+--------------------------------------------------------------------

personas | -.0012586 .0040959 -0.31 0.759 5.136 -.009286 .006769

hombre\*| .0115823 .0158278 0.73 0.464 .51625 -.01944 .042604

orden\_n | .0288062 .0252374 1.14 0.254 1.09925 -.020658 .078271

baja\*| -.0011442 .0158283 -0.07 0.942 .5 -.032167 .029879

ocupad~e\*| .0097187 .0196932 0.49 0.622 .79775 -.028879 .048317

educa\_~e | .0030452 .0021364 1.43 0.154 6.40325 -.001142 .007232

ingres~e | -.0000347 .0000968 -0.36 0.720 73.5289 -.000224 .000155

---------+--------------------------------------------------------------------

obs. P | .5

pred. P | .5000003 (at x-bar)

------------------------------------------------------------------------------

(\*) 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

.

. logit D $X

Iteration 0: log likelihood = -2772.5887

Iteration 1: log likelihood = -2770.2968

Iteration 2: log likelihood = -2770.2968

Logistic regression Number of obs = 4000

LR chi2(7) = 4.58

Prob > chi2 = 0.7106

Log likelihood = -2770.2968 Pseudo R2 = 0.0008

------------------------------------------------------------------------------

D | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0050448 .0163964 -0.31 0.758 -.0371812 .0270917

hombre | .0463558 .0633327 0.73 0.464 -.0777741 .1704856

orden\_n | .1153557 .1010767 1.14 0.254 -.082751 .3134625

baja | -.0045836 .0633282 -0.07 0.942 -.1287045 .1195373

ocupado\_jefe | .0389278 .0788043 0.49 0.621 -.1155257 .1933814

educa\_jefe | .0121869 .0085499 1.43 0.154 -.0045706 .0289444

ingresos\_h~e | -.000139 .0003871 -0.36 0.719 -.0008978 .0006197

\_cons | -.2213942 .1729717 -1.28 0.201 -.5604126 .1176242

------------------------------------------------------------------------------

.

. mfx

Marginal effects after logit

y = Pr(D) (predict)

= .50000183

------------------------------------------------------------------------------

variable | dy/dx Std. Err. z P>|z| [ 95% C.I. ] X

---------+--------------------------------------------------------------------

personas | -.0012612 .0041 -0.31 0.758 -.009295 .006773 5.136

hombre\*| .0115884 .01583 0.73 0.464 -.01944 .042617 .51625

orden\_n | .0288389 .02527 1.14 0.254 -.020688 .078366 1.09925

baja\*| -.0011459 .01583 -0.07 0.942 -.032176 .029884 .5

ocupad~e\*| .0097313 .0197 0.49 0.621 -.028875 .048337 .79775

educa\_~e | .0030467 .00214 1.43 0.154 -.001143 .007236 6.40325

ingres~e | -.0000348 .0001 -0.36 0.719 -.000224 .000155 73.5289

------------------------------------------------------------------------------

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

En este último paso verificamos que en realidad el tratamiento no se asigna de acuerdo a características observables de los individuos. Ninguna de las variables es significativa en ninguno de los dos modelos y ningún modelo es significativo. Esto nos garantiza la aleatorización en la asignación del tratamiento.