

Evaluation of Health Policy with HISP_Linear Regression

Description of the dataset for practical exercises

Impact Evaluation in Practice (second edition) and related technical companion uses a fictionalized case, the Health Insurance Subsidy Program (HISP) to illustrate many of the concepts and methods that are presented. While fictionalized, the HISP case is modelled after real-world examples of impact evaluations. One of the primary objectives of HISP is to reduce the burden of health-related out-of-pocket expenditures for low income households. <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464807794.pdf?sequence=2&isAllowed=true>

The HISP Dataset

The HISP dataset has the Stata format and is called "evaluation.dta". The dataset is at the level of household and round. This means that one observation (row) captures information for one household either in the baseline (round 0) or in the follow-up survey (round 1). In other words, every household appears on two rows: one for the baseline and one for the follow-up survey. Exercises 1-5 use only the cross-sectional follow-up survey, while Exercise 6 uses the full panel data.

Practical Exercise Topics:

- Exercise 1: Data Exploration
- Exercise 2: Regression (regress, teffects, areg)
- Exercise 3: Matching (cem, ebalance, probit, teffect nnmatch, teffect ipw/ipwra)
- Exercise 4: Instrument Variables (ivregress)
- Exercise 5: Regression Discontinuity Designs (rdrobust, rdplot)

Set up

Launching stata from the jupyter notebook

```
In [1]: %%capture
import stata_setup
import os
os.chdir('C:\Program Files\Stata17\utilities')
from pystata import config
config.init('mp');
```

Initial set up of log file and load data

```
In [2]: %%capture
%%stata
set more off, perm

# redirect to workplace
cd "C:\Users\USER\Desktop\Charlene\2022 Charlene at York\Evaluation of Health Policy\practical exercise"

# create log file
capture log close
log using "EHP_HISP_Charlene.log", replace

# Load data
use "evaluation.dta", clear

# Limit sample to the cross section of post-treatment data
drop if round==0
```

Data exploration

Conduct descriptive statistic on the HISP dataset

```
In [3]: %%stata
codebook, compact
codebook, notes
```

```
. codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
locality_i~rr	9914	200	73.93444	1	200	Locality identifier
household_~r	9914	9914	8038.721	2	15778	Unique household id...
treatment_~y	9914	2	.5003026	0	1	Household is locate...
promotion_~y	9914	2	.5127093	0	1	Household is locate...
eligible	9914	2	.5677829	0	1	Household eligible ...
enrolled	9914	2	.299072	0	1	HH enrolled in HISP...
enrolled_rp	9914	2	.2933226	0	1	Household enrolled ...
poverty_in~x	9914	1961	56.78873	20.47913	100	Poverty Index 1-100
round	9914	1	1	1	1	Survey round (0 = b...
health_exp~s	9914	9621	16.85922	0	116.8306	Out of pocket healt...
age_hh	9914	75	47.16584	15	88	Age of the head of ...
age_sp	9914	75	40.97952	14	88	Age of the spouse (...)
educ_hh	9914	19	2.83381	0	16	Education of the he...
educ_sp	9914	19	2.619629	0	17	Education of the sp...
female_hh	9914	2	.0990518	0	1	Head of the househo...
indigenous	9914	2	.3529352	0	1	Head of household s...
hsize	9914	13	5.178737	1	13	Number of household...
dirtfloor	9914	2	.6035909	0	1	Home has a dirt flo...
bathroom	9914	2	.6159976	0	1	Home with private b...
land	9914	24	2.079988	0	23	Number of hectares ...
hospital_d~e	9914	463	105.3254	9.465392	170.4576	Distance to closest...
hospital	5629	2	.0506307	0	1	HH member visited h...

```
. codebook, notes
```

locality_identifier	Locality identifier
<hr/>	
<p>Type: Numeric (float)</p>	
Range: [1,200]	Units: 1
Unique values: 200	Missing :: 0/9,914
Mean: 73.9344	
Std. dev.: 55.0767	
Percentiles:	10% 25% 50% 75% 90%
	11 30 59 112 160
<hr/>	
household_identifier	Unique household identifier
<hr/>	
<p>Type: Numeric (float)</p>	
Range: [2,15778]	Units: 1
Unique values: 9,914	Missing :: 0/9,914
Mean: 8038.72	
Std. dev.: 4569.62	
Percentiles:	10% 25% 50% 75% 90%
	1672 3989 8037.5 12033 14384
<hr/>	
treatment_locality	Household is located in treatment community (0=no, 1=yes)
<hr/>	
<p>Type: Numeric (float)</p>	
Range: [0,1]	Units: 1
Unique values: 2	Missing :: 0/9,914
Tabulation: Freq. Value	
4,954	0
4,960	1
<hr/>	
promotion_locality	Household is located in promoted community (0=no, 1=yes)
<hr/>	
<p>Type: Numeric (float)</p>	
Range: [0,1]	Units: 1
Unique values: 2	Missing :: 0/9,914
Tabulation: Freq. Value	
4,831	0
5,083	1
<hr/>	
eligible	Household eligible to enroll in HISP (0=no, 1=yes)
<hr/>	

Type: Numeric (float)

Range: [0,1] Units: 1
Unique values: 2 Missing :: 0/9,914

Tabulation: Freq. Value
4,285 0
5,629 1

enrolled HH enrolled in HISP (0=no, 1=yes)

Type: Numeric (float)

Range: [0,1] Units: 1
Unique values: 2 Missing :: 0/9,914

Tabulation: Freq. Value
6,949 0
2,965 1

enrolled_rp Household enrolled in HISP under the random promotion scenario (0=no, 1=yes)

Type: Numeric (float)

Range: [0,1] Units: 1
Unique values: 2 Missing :: 0/9,914

Tabulation: Freq. Value
7,006 0
2,908 1

poverty_index Poverty Index 1-100

Type: Numeric (float)

Range: [20.479134,100] Units: 1.000e-06
Unique values: 1,961 Missing :: 0/9,914

Mean: 56.7887
Std. dev.: 10.6868

Percentiles: 10% 25% 50% 75% 90%
43.5085 49.6522 56.4077 62.9057 70.2473

round Survey round (0 = baseline; 1 = follow-up)

Type: Numeric (float)

Range: [1,1] Units: 1
Unique values: 1 Missing :: 0/9,914

Tabulation: Freq. Value
9,914 1

health_expenditures Out of pocket health expenditures (per person per year)

Type: Numeric (float)

Range: [0,116.83064] Units: 1.000e-10
Unique values: 9,621 Missing :: 0/9,914

Mean: 16.8592
Std. dev.: 11.8904

Percentiles: 10% 25% 50% 75% 90%
3.45648 9.95981 14.8696 21.5856 30.8137

age_hh Age of the head of the household (in years)

Type: Numeric (float)

Range: [15,88] Units: 1.000e-06
Unique values: 75 Missing :: 0/9,914

Mean: 47.1658
Std. dev.: 15.2898

Percentiles: 10% 25% 50% 75% 90%

age_sp Age of the spouse (in years)

Type: Numeric (float)

Range: [14,88] Unique values: 75 Units: 1 Missing .: 0/9,914

Mean: 40.9795
Std. dev.: 12.8395

Percentiles: 10% 25% 50% 75% 90%
25 31 41 48 59

educ_hh Education of the head of household (completed years of schooling)

Type: Numeric (float)

Range: [0,16] Unique values: 19 Units: 1.000e-07 Missing .: 0/9,914

Mean: 2.83381
Std. dev.: 2.75487

Percentiles: 10% 25% 50% 75% 90%
0 0 2 4 6

educ_sp Education of the spouse (completed years of schooling)

Type: Numeric (float)

Range: [0,17] Unique values: 19 Units: .1 Missing .: 0/9,914

Mean: 2.61963
Std. dev.: 2.54359

Percentiles: 10% 25% 50% 75% 90%
0 0 2 4 6

female_hh Head of the household is a woman (0=no, 1=yes)

Type: Numeric (float)

Range: [0,1] Unique values: 2 Units: 1 Missing .: 0/9,914

Tabulation: Freq. Value
8,932 0
982 1

indigenous Head of household speaks an indigenous language (0=no, 1=yes)

Type: Numeric (float)

Range: [0,1] Unique values: 2 Units: 1 Missing .: 0/9,914

Tabulation: Freq. Value
6,415 0
3,499 1

hhszie Number of household members (baseline)

Type: Numeric (float)

Range: [1,13] Unique values: 13 Units: 1 Missing .: 0/9,914

Mean: 5.17874
Std. dev.: 2.19519

Percentiles: 10% 25% 50% 75% 90%
2 4 5 7 8

dirtfloor Home has a dirt floor at baseline (0=no, 1=yes)

Type: Numeric (byte)

Range: [0,1] Units: 1
Unique values: 2 Missing .: 0/9,914

Tabulation: Freq. Value
3,930 0
5,984 1

bathroom Home with private bathroom at baseline (0=no, 1=yes)

Type: Numeric (byte)

Range: [0,1] Units: 1
Unique values: 2 Missing .: 0/9,914

Tabulation: Freq. Value
3,807 0
6,107 1

land Number of hectares of land owned by household at baseline

Type: Numeric (byte)

Range: [0,23] Units: 1
Unique values: 24 Missing .: 0/9,914

Mean: 2.07999
Std. dev.: 3.1334

Percentiles:	10%	25%	50%	75%	90%
	0	0	1	3	5

hospital_distance Distance to closest hospital

Type: Numeric (double)

Range: [9.4653916,170.45765] Units: 1.000e-08
Unique values: 463 Missing .: 0/9,914

Mean: 105.325
Std. dev.: 42.0655

Percentiles:	10%	25%	50%	75%	90%
	42.1686	72.8042	113.556	141.098	156.045

hospital HH member visited hospital in the past year (0=no, 1=yes)

Type: Numeric (float)

Range: [0,1] Units: 1
Unique values: 2 Missing .: 4,285/9,914

Tabulation: Freq. Value
5,344 0
285 1
4,285 .

.

In [4]:

```
% stata
describe
summ
```

```
. describe
```

Contains data from evaluation.dta					
Observations:	9,914 <th data-cs="4" data-kind="parent">Household Roster (households with completed interview)</th> <th data-kind="ghost"></th> <th data-kind="ghost"></th> <th data-kind="ghost"></th>	Household Roster (households with completed interview)			
Variables:	22	9 Sep 2016 11:03 (.dta has notes)			

Variable name	Storage type	Display format	Value label	Variable label	
locality_iden~r	float	%9.0g	Locality identifier		
household_ide~r	float	%9.0g	Unique household identifier		
treatment_loc~y	float	%9.0g	Household is located in treatment community (0=no, 1=yes)		
promotion_loc~y	float	%9.0g	Household is located in promoted community (0=no, 1=yes)		
eligible	float	%9.0g	Household eligible to enroll in HISP (0=no, 1=yes)		
enrolled	float	%9.0g	HH enrolled in HISP (0=no, 1=yes)		
enrolled_rp	float	%9.0g	Household enrolled in HISP under the random promotion scenario (0=no, 1=yes)		
poverty_index	float	%9.0g	Poverty Index 1-100		
round	float	%9.0g	Survey round (0 = baseline; 1 = follow-up)		
health_expend~s	float	%9.0g	Out of pocket health expenditures (per person per year)		
age_hh	float	%9.0g	Age of the head of the household (in years)		
age_sp	float	%9.0g	Age of the spouse (in years)		
educ_hh	float	%9.0g	Education of the head of household (completed years of schooling)		
educ_sp	float	%9.0g	Education of the spouse (completed years of schooling)		
female_hh	float	%9.0g	Head of the household is a woman (0=no, 1=yes)		
indigenous	float	%9.0g	Head of household speaks an indigenous language (0=no, 1=yes)		
hhsize	float	%9.0g	Number of household members (baseline)		
dirtfloor	byte	%8.0g	Home has a dirt floor at baseline (0=no, 1=yes)		
bathroom	byte	%8.0g	Home with private bathroom at baseline (0=no, 1=yes)		
land	byte	%9.0g	Number of hectares of land owned by household at baseline		
hospital_dist~e	double	%11.0g	Distance to closest hospital		
hospital	float	%9.0g	HH member visited hospital in the past year (0=no, 1=yes)		

Sorted by:

Note: Dataset has changed since last saved.

```
. summ
```

Variable	Obs	Mean	Std. dev.	Min	Max
locality_i~r	9,914	73.93444	55.07667	1	200
household_~r	9,914	8038.721	4569.619	2	15778
treatment_~y	9,914	.5003026	.5000251	0	1
promotion_~y	9,914	.5127093	.4998637	0	1
eligible	9,914	.5677829	.4954092	0	1
enrolled	9,914	.299072	.4578745	0	1
enrolled_rp	9,914	.2933226	.455308	0	1
poverty_in~x	9,914	56.78873	10.68676	20.47913	100
round	9,914	1	0	1	1
health_exps~s	9,914	16.85922	11.89039	0	116.8306
age_hh	9,914	47.16584	15.28985	15	88
age_sp	9,914	40.97952	12.83947	14	88
educ_hh	9,914	2.83381	2.754866	0	16
educ_sp	9,914	2.619629	2.54359	0	17
female_hh	9,914	.0990518	.2987467	0	1
indigenous	9,914	.3529352	.4779069	0	1
hhsize	9,914	5.178737	2.195188	1	13
dirtfloor	9,914	.6035909	.4891759	0	1
bathroom	9,914	.6159976	.486383	0	1
land	9,914	2.079988	3.133397	0	23
hospital_d~e	9,914	105.3254	42.06549	9.465392	170.4576
hospital	5,629	.0506307	.2192618	0	1

Create(rename) variable for treatment effect evaluation

```
In [5]: %%capture
%%stata

# create generic variable (y)
clonevar y=health_expenditures
label var y "out of pocket health expenditure pc/pa"
clonevar d=enrolled
label var d "Treatment"

# create potential outcome y0 and y1
clonevar y0=y
replace y0=. if d==1
clonevar y1=y
replace y1=. if d==0

# summarise outcome of the treated and control group
bysort d:summm y y0 y1
tabstat y y0 y1, by(d)
```

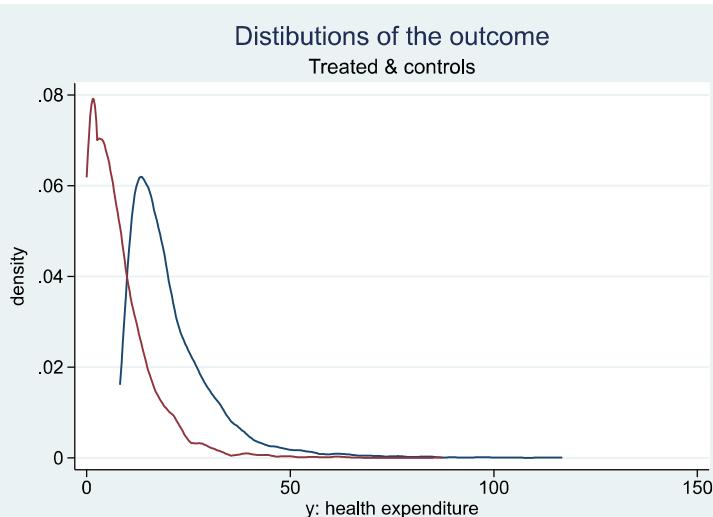
Produce two-way density plots for distribution of outcomes

```
In [6]: %%stata

tw kdensity y0, || kdensity y1, ///
title(Distributions of the outcome) subtitle(Treated & controls) ///
xtitle(y: health expenditure) ytitle(density) ylabel( , angle(horizontal)) ///
legend(off) saving(hist, replace)
graph export pdf_y.png, replace

.
tw kdensity y0, || kdensity y1, ///
> title(Distributions of the outcome) subtitle(Treated & controls) ///
> xtitle(y: health expenditure) ytitle(density) ylabel( , angle(horizontal)) //
> /
> legend(off) saving(hist, replace)
file hist.gph saved

.
graph export pdf_y.png, replace
file pdf_y.png written in PNG format
```



The figure shows the density plots of the outcome for the treated and controls.

Linear Regression

Create global list of regressors

```
In [7]: %%capture
%%stata

global xs "age_hh age_sp educ_hh educ_sp female_hh indigenous hhsiz dirtfloor bathroom land hospital_distance"
```

Method 1: Command teffects

teffects

- teffect allows us to compute treatment effects directly using regression adjustment(ra).
- teffects: Treatment-effects estimation for observational data (POMs, ATEs, ATETs(ATT))

Regression adjustment to est. ATT (ATET) (Average Treatment Effect of Treated)

The average OPP is 9.8 dollar less when all the family who participate in the HISP program if they do not do so.

The average of 17.68 dollar would have occurred if none of those whom participate in the HISPM participate had not do so.

In [8]:

```
% stata  
teffects ra (y $xs) (d), atet  
. teffects ra (y $xs) (d), atet  
  
Iteration 0: EE criterion = 3.215e-26  
Iteration 1: EE criterion = 1.349e-30  
  
Treatment-effects estimation Number of obs = 9,914  
Estimator : regression adjustment  
Outcome model : linear  
Treatment model: none  
  
-----  
y | Robust  
   Coefficient std. err.      z     P>|z| [95% conf. interval]  
-----  
ATET  
d  
(1 vs 0) | -9.837815 .169695 -57.97 0.000 -10.17041 -9.505219  
-----  
P0mean  
d  
0 | 17.67799 .1313196 134.62 0.000 17.42061 17.93538  
-----  
.  
.
```

Regression adjustment to est. ATE (Average Treatment Effect)

The average OOP if all participates in HISP program would be 10 dollar less than the average of 19.80 if none of the family had participate.

Note here that from the given calculation "**ATET=-9.83 ATE=-10.41**", we know that the impact will be bigger if we conduct treatment to all population than just treating the one in the treatment group.

In [9]:

```
% stata  
teffects ra (y $xs) (d), ate  
. teffects ra (y $xs) (d), ate  
  
Iteration 0: EE criterion = 3.215e-26  
Iteration 1: EE criterion = 1.511e-30  
  
Treatment-effects estimation Number of obs = 9,914  
Estimator : regression adjustment  
Outcome model : linear  
Treatment model: none  
  
-----  
y | Robust  
   Coefficient std. err.      z     P>|z| [95% conf. interval]  
-----  
ATE  
d  
(1 vs 0) | -10.41073 .2153258 -48.35 0.000 -10.83276 -9.988702  
-----  
P0mean  
d  
0 | 19.80143 .1160521 170.63 0.000 19.57397 20.02889  
-----  
.  
.
```

Regression adjustment to the means of y0 and y1

The given command allows us to see the average treatment effect(just to see where the base benchmark is)

In [10]:

```
% stata  
teffects ra (y $xs) (d), pomeans
```

```

. teffects ra (y $xs) (d), pomeans

Iteration 0: EE criterion = 3.215e-26
Iteration 1: EE criterion = 1.510e-30

Treatment-effects estimation                               Number of obs      =    9,914
Estimator       : regression adjustment
Outcome model   : linear
Treatment model: none

-----+
y | Coefficient  Robust std. err.      z     P>|z| [95% conf. interval]
-----+
POmeans
d |
0 | 19.80143   .1160521   170.63  0.000   19.57397   20.02889
1 | 9.390699   .1926345    48.75  0.000   9.013142   9.768256
-----+

```

.