Evaluating the impact of HISP: Instrumental Variables

In the design of HISP, there are two rounds of data on two groups of households: one group that enrolled in the program, and the another that did not. As in the case of the enrolled and non-enrolled groups, **we realized that we cannot simply compare the average health expentidures of the two groups beacuase of selection bias.** As we have data for two periods for wach household in the sample, we can use those data to solve some of these challenges by comparing the change in health expenditures for the two groups.

Set up

Lauching 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

clear
set more off, perm

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

# load data
use "evaluation.dta", clear
```

Create(rename) variable for treatment effect evaluation

In [3]:
```
%%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 global list of regressors
global xs "age_hh age_sp educ_hh educ_sp female_hh indigenous hhsize dirtfloor bathroom land hospital_distar
```

Intrumental Variables from Text book

In [4]:
```
%%stata
describe locality_identifier household_identifier promotion_locality enrolled enrolled_rp
```

```
Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
locality_iden~r float   %9.0g                 Locality identifier
household_ide~r float   %9.0g                 Unique household identifier
promotion_loc~y float   %9.0g                 Household is located in promoted
                                                community (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)
```

In [5]:
```
%%stata
ttest health_expenditures if round==0, by(promotion_locality)
```

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |   4,831    17.23795    .0814034    5.657973    17.07836    17.39754
       1 |   5,082    17.18535    .0774503    5.521289    17.03352    17.33719
---------+--------------------------------------------------------------------
Combined |   9,913    17.21099    .0561257    5.588098    17.10097      17.321
---------+--------------------------------------------------------------------
    diff |            .0525987    .1122917               -.167516    .2727133
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   0.4684
H0: diff = 0                                     Degrees of freedom =     9911

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6802         Pr(|T| > |t|) = 0.6395          Pr(T > t) = 0.3198
```

In [7]:
```
%%stata
ttest health_expenditures if round==1, by(promotion_locality)
```

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |   4,831    18.84538    .1558312    10.83111    18.53988    19.15088
       1 |   5,083    14.97152     .175731    12.52877    14.62701    15.31603
---------+--------------------------------------------------------------------
Combined |   9,914    16.85922    .1194185    11.89039    16.62513     17.0933
---------+--------------------------------------------------------------------
    diff |             3.87386    .2357366                3.411769    4.335952
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  16.4330
H0: diff = 0                                     Degrees of freedom =     9912

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

In [6]:
```
%%stata
ttest enrolled_rp if round==1, by(promotion_locality)
```

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |   4,831    .0842476    .0039966    .2777875    .0764123    .0920828
       1 |   5,083    .4920323    .0070129    .4999857     .478284    .5057806
---------+--------------------------------------------------------------------
Combined |   9,914    .2933226    .0045728     .455308     .284359    .3022862
---------+--------------------------------------------------------------------
    diff |           -.4077847    .0081809               -.423821   -.3917484
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t = -49.8458
H0: diff = 0                                     Degrees of freedom =     9912

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

In [9]:
```
%%stata
ivreg health_expenditures (enrolled_rp = promotion_locality) if round==1, first
```

```
First-stage regressions
-----------------------

      Source |       SS           df       MS      Number of obs   =     9,914
-------------+----------------------------------   F(1, 9912)      =   2484.60
       Model |  411.879408         1  411.879408   Prob > F        =    0.0000
    Residual |  1643.13855     9,912  .165772654   R-squared       =    0.2004
-------------+----------------------------------   Adj R-squared   =    0.2003
       Total |  2055.01795     9,913  .207305352   Root MSE        =    .40715

------------------------------------------------------------------------------
 enrolled_rp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
promotion_~y |   .4077847   .0081809    49.85   0.000     .3917484     .423821
       _cons |   .0842476   .0058578    14.38   0.000      .072765    .0957301
------------------------------------------------------------------------------


Instrumental variables 2SLS regression

      Source |       SS           df       MS      Number of obs   =     9,914
-------------+----------------------------------   F(1, 9912)      =    337.77
       Model |  310737.314         1  310737.314   Prob > F        =    0.0000
    Residual |  1090776.36     9,912  110.046042   R-squared       =    0.2217
-------------+----------------------------------   Adj R-squared   =    0.2216
       Total |  1401513.68     9,913  141.381386   Root MSE        =     10.49

------------------------------------------------------------------------------
health_exp~s | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
 enrolled_rp |  -9.499769   .5168948   -18.38   0.000    -10.51299    -8.48655
       _cons |   19.64571   .1846287   106.41   0.000      19.2838    20.00762
------------------------------------------------------------------------------
Instrumented: enrolled_rp
 Instruments: promotion_locality
```

In [17]:
```
%%stata
ivreg health_expenditures (enrolled_rp = treatment_locality) if round==1, first
```

```
First-stage regressions
-----------------------

      Source |       SS           df       MS      Number of obs   =     9,914
-------------+----------------------------------   F(1, 9912)      =   7019.16
       Model |  851.950212         1  851.950212   Prob > F        =    0.0000
    Residual |  1203.06774     9,912  .121374873   R-squared       =    0.4146
-------------+----------------------------------   Adj R-squared   =    0.4145
       Total |  2055.01795     9,913  .207305352   Root MSE        =    .34839

------------------------------------------------------------------------------
 enrolled_rp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
treatment_~y |   .5862903   .0069979    83.78   0.000     .5725729    .6000077
       _cons |  -2.11e-15   .0049498    -0.00   1.000    -.0097026    .0097026
------------------------------------------------------------------------------


Instrumental variables 2SLS regression

      Source |       SS           df       MS      Number of obs   =     9,914
-------------+----------------------------------   F(1, 9912)      =    936.82
       Model |  325368.463         1  325368.463   Prob > F        =    0.0000
    Residual |  1076145.22     9,912  108.569937   R-squared       =    0.2322
-------------+----------------------------------   Adj R-squared   =    0.2321
       Total |  1401513.68     9,913  141.381386   Root MSE        =     10.42

------------------------------------------------------------------------------
health_exp~s | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
 enrolled_rp |  -10.92634   .3569831   -30.61   0.000     -11.6261   -10.22658
       _cons |   20.06416   .1480392   135.53   0.000     19.77397    20.35435
------------------------------------------------------------------------------
Instrumented: enrolled_rp
 Instruments: treatment_locality
```

In [18]:
```
%%stata
ivreg health_expenditures $xs (enrolled_rp = treatment_locality) if round==1, first
```

```
First-stage regressions
-----------------------

      Source |       SS           df       MS      Number of obs   =     9,914
-------------+----------------------------------   F(12, 9901)     =    818.72
       Model |  1023.53142        12  85.2942854   Prob > F        =    0.0000
    Residual |  1031.48653     9,901  .104180035   R-squared       =    0.4981
-------------+----------------------------------   Adj R-squared   =    0.4975
       Total |  2055.01795     9,913  .207305352   Root MSE        =    .32277

------------------------------------------------------------------------------
 enrolled_rp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      age_hh |  -.0035159   .0003963    -8.87   0.000    -.0042928    -.002739
      age_sp |  -.0021934   .0004545    -4.83   0.000    -.0030844   -.0013024
     educ_hh |  -.0074264   .0014646    -5.07   0.000    -.0102973   -.0045556
     educ_sp |  -.0037732   .0015998    -2.36   0.018    -.0069092   -.0006372
   female_hh |  -.0061348   .0118114    -0.52   0.603    -.0292875    .0170179
  indigenous |   .0482017    .007489     6.44   0.000     .0335218    .0628816
      hhsize |   .0350622   .0015355    22.83   0.000     .0320524     .038072
   dirtfloor |   .0984281   .0071018    13.86   0.000      .084507    .1123491
    bathroom |  -.0294194   .0068673    -4.28   0.000    -.0428807   -.0159581
        land |  -.0071562   .0010714    -6.68   0.000    -.0092563   -.0050561
hospital_d~e |   .0001851   .0000827     2.24   0.025     .0000231    .0003472
treatment_~y |   .5776975   .0064928    88.98   0.000     .5649703    .5904248
       _cons |   .0470624   .0218439     2.15   0.031      .004244    .0898809
------------------------------------------------------------------------------


Instrumental variables 2SLS regression

      Source |       SS           df       MS      Number of obs   =     9,914
-------------+----------------------------------   F(12, 9901)     =    477.27
       Model |  567367.314        12  47280.6095   Prob > F        =    0.0000
    Residual |  834146.365     9,901  84.2486986   R-squared       =    0.4048
-------------+----------------------------------   Adj R-squared   =    0.4041
       Total |  1401513.68     9,913  141.381386   Root MSE        =    9.1787

------------------------------------------------------------------------------
health_exp~s | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
 enrolled_rp |  -10.61031   .3196105   -33.20   0.000    -11.23682   -9.983812
      age_hh |   .0707064   .0113381     6.24   0.000     .0484815    .0929313
      age_sp |   -.015281   .0129483    -1.18   0.238    -.0406623    .0101002
     educ_hh |   .0338553   .0417059     0.81   0.417    -.0478967    .1156073
     educ_sp |  -.0498393   .0455227    -1.09   0.274     -.139073    .0393943
   female_hh |   1.024671   .3358944     3.05   0.002     .3662494    1.683092
  indigenous |  -2.294977   .2136229   -10.74   0.000    -2.713721   -1.876232
      hhsize |  -2.010351   .0450731   -44.60   0.000    -2.098704   -1.921999
   dirtfloor |  -1.999492   .2049496    -9.76   0.000    -2.401235   -1.597749
    bathroom |   .6589115   .1955502     3.37   0.001     .2755933     1.04223
        land |   .0895238   .0305617     2.93   0.003     .0296167     .149431
hospital_d~e |  -.0040322   .0023538    -1.71   0.087    -.0086462    .0005817
       _cons |    29.4564   .6235595    47.24   0.000      28.2341    30.67871
------------------------------------------------------------------------------
Instrumented: enrolled_rp
 Instruments: age_hh age_sp educ_hh educ_sp female_hh indigenous hhsize
              dirtfloor bathroom land hospital_distance treatment_locality
```

Instrumental Variables

In [7]:
```
%%stata
drop if round==0
```

```
(9,913 observations deleted)
```

In [8]:
```
%%stata
regress health_expenditures enrolled $xs, robust
```

```
Linear regression                               Number of obs     =      9,914
                                                F(12, 9901)       =     496.87
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4118
                                                Root MSE          =     9.1246

------------------------------------------------------------------------------
             |               Robust
health_exp~s | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    enrolled |  -10.00504   .1758574   -56.89   0.000    -10.34976   -9.660324
      age_hh |   .0738534   .0134845     5.48   0.000     .0474209    .1002858
      age_sp |  -.0152703   .0153175    -1.00   0.319    -.0452958    .0147552
     educ_hh |   .0416857   .0419395     0.99   0.320    -.0405242    .1238956
     educ_sp |  -.0495566   .0442701    -1.12   0.263    -.1363349    .0372217
   female_hh |   1.010568    .425649     2.37   0.018     .1762096    1.844927
  indigenous |  -2.344442   .2121613   -11.05   0.000    -2.760321   -1.928562
      hhsize |   -2.02707   .0497833   -40.72   0.000    -2.124655   -1.929485
   dirtfloor |  -2.036285   .2092479    -9.73   0.000    -2.446453   -1.626116
    bathroom |   .6130987    .191856     3.20   0.001     .2370218    .9891756
        land |   .0962154   .0322123     2.99   0.003     .0330728    .1593581
hospital_d~e |  -.0037318   .0024996    -1.49   0.135    -.0086315    .0011678
       _cons |   29.27489   .6138351    47.69   0.000     28.07165    30.47813
------------------------------------------------------------------------------
```

In [9]:
```
%%stata
regress health_expenditures treatment_locality $xs, robust
```

```
Linear regression                               Number of obs     =      9,914
                                                F(12, 9901)       =     317.19
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3443
                                                Root MSE          =     9.6343

------------------------------------------------------------------------------
             |               Robust
health_exp~s | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
treatment_~y |  -6.129552   .1937445   -31.64   0.000    -6.509331   -5.749773
      age_hh |   .1080116   .0140847     7.67   0.000     .0804027    .1356206
      age_sp |    .007992   .0161433     0.50   0.621    -.0236521    .0396361
     educ_hh |   .1126522   .0439669     2.56   0.010     .0264682    .1988363
     educ_sp |  -.0098043   .0478215    -0.21   0.838    -.1035441    .0839356
   female_hh |   1.089763    .442747     2.46   0.014     .2218885    1.957637
  indigenous |  -2.806412   .2240041   -12.53   0.000    -3.245506   -2.367318
      hhsize |  -2.382372   .0520814   -45.74   0.000    -2.484462   -2.280282
   dirtfloor |  -3.043845   .2204035   -13.81   0.000     -3.47588   -2.611809
    bathroom |   .9710603   .2036544     4.77   0.000     .5718562    1.370264
        land |   .1654536   .0331034     5.00   0.000     .1005642    .2303429
hospital_d~e |  -.0059965    .002615    -2.29   0.022    -.0111225   -.0008705
       _cons |   28.95706   .6450464    44.89   0.000     27.69263    30.22148
------------------------------------------------------------------------------
```

In [10]:
```
%%stata
regress enrolled treatment_locality $xs, robust
```

```
Linear regression                               Number of obs     =      9,914
                                                F(12, 9901)       =     987.32
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5111
                                                Root MSE          =     .32036

------------------------------------------------------------------------------
             |               Robust
    enrolled | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
treatment_~y |   .5889406   .0064712    91.01   0.000     .5762557    .6016255
      age_hh |  -.0034289   .0004028    -8.51   0.000    -.0042184   -.0026393
      age_sp |  -.0023327   .0004669    -5.00   0.000    -.0032478   -.0014175
     educ_hh |  -.0070704   .0014737    -4.80   0.000    -.0099591   -.0041816
     educ_sp |  -.0040275   .0015657    -2.57   0.010    -.0070967   -.0009583
   female_hh |  -.0080185   .0120307    -0.67   0.505    -.0316011    .0155641
  indigenous |   .0463516   .0075163     6.17   0.000     .0316181    .0610851
      hhsize |   .0355089   .0015386    23.08   0.000     .0324928    .0385249
   dirtfloor |    .10117   .0073809    13.71   0.000      .086702    .1156381
    bathroom |  -.0358753    .006836    -5.25   0.000    -.0492753   -.0224753
        land |  -.0069355   .0011225    -6.18   0.000    -.0091359   -.0047351
hospital_d~e |   .0002364   .0000808     2.93   0.003     .0000781    .0003947
       _cons |   .0434316   .0211063     2.06   0.040     .0020591    .0848042
------------------------------------------------------------------------------
```

In [11]:
```
%%stata
ivregress 2sls health_expenditures $xs (enrolled=treatment_locality), vce(robust)
```

```
Instrumental variables 2SLS regression            Number of obs   =      9,914
                                                  Wald chi2(12)   =    3984.74
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.4116
                                                  Root MSE        =     9.1203

------------------------------------------------------------------------------
             |               Robust
health_exp~s | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    enrolled |  -10.40776   .3110995   -33.45   0.000     -11.0175   -9.798016
      age_hh |   .0723248   .0135174     5.35   0.000     .0458311    .0988184
      age_sp |   -.016286   .0152964    -1.06   0.287    -.0462664    .0136945
     educ_hh |   .0390656   .0417476     0.94   0.349    -.0427581    .1208893
     educ_sp |  -.0517214   .0442616    -1.17   0.243    -.1384726    .0350297
   female_hh |   1.006308   .4252107     2.37   0.018     .1729102    1.839706
  indigenous |  -2.323996    .213317   -10.89   0.000    -2.742089   -1.905902
      hhsize |  -2.012804      .0507   -39.70   0.000    -2.112174   -1.913434
   dirtfloor |  -1.990891   .2102298    -9.47   0.000    -2.402934   -1.578848
    bathroom |   .5976787   .1921413     3.11   0.002     .2210886    .9742687
        land |   .0932704   .0321708     2.90   0.004     .0302169    .1563239
hospital_d~e |   -.003536   .0024983    -1.42   0.157    -.0084326    .0013605
       _cons |   29.40908   .6149662    47.82   0.000     28.20377    30.61439
------------------------------------------------------------------------------
Instrumented: enrolled
 Instruments: age_hh age_sp educ_hh educ_sp female_hh indigenous hhsize
              dirtfloor bathroom land hospital_distance treatment_locality
```

In [12]:
```
%%stata
estat endogenous
```

```
  Tests of endogeneity
  H0: Variables are exogenous

  Robust score chi2(1)            =   2.1799  (p = 0.1398)
  Robust regression F(1,9900)     =  2.18031  (p = 0.1398)
```

In [13]:
```
%%stata
estat firststage
```

```
  First-stage regression summary statistics
  ------------------------------------------------------------------------------
               |            Adjusted      Partial       Robust
      Variable |   R-sq.       R-sq.        R-sq.     F(1,9901)   Prob > F
  -------------+----------------------------------------------------------------
      enrolled |  0.5111      0.5105       0.4576       8282.64    0.0000
  ------------------------------------------------------------------------------
```