```
* Propensity Score Matching in Stata
* Copyright 2013 by Ani Katchova

clear all
set more off

* Download and install Stata ado files for pscore
* net install st0026_2

use C:\Econometrics\Data\matching_earnings

* Define treatment, outcome, and independent variables
global treatment TREAT
global ylist RE78
global xlist AGE EDUC MARR
global breps 5

* For difference-in-differences, outcome is the differences in outcomes after and
before
* global ylist REDIFF

describe $treatment $ylist $xlist
summarize $treatment $ylist $xlist

bysort $treatment: summarize $ylist $xlist

* Regression with a dummy variable for treatment (t-test)
reg $ylist $treatment

* Regression with a dummy variable for treatment controlling for x
reg $ylist $treatment $xlist

* Propensity score matching with common support
pscore $treatment $xlist, pscore(myscore) blockid(myblock) comsup

* Matching methods

* Nearest neighbor matching
attnd $ylist $treatment $xlist, pscore(myscore) comsup boot reps($breps) dots

* Radius matching
attr $ylist $treatment $xlist, pscore(myscore) comsup boot reps($breps) dots
radius(0.1)

* Kernel Matching
attk $ylist $treatment $xlist, pscore(myscore) comsup boot reps($breps) dots

* Stratification Matching
atts $ylist $treatment $xlist, pscore(myscore) blockid(myblock) comsup boot
reps($breps) dots
```

```
. * Propensity Score Matching in Stata
. * Copyright 2013 by Ani Katchova
.
. clear all

. set more off

.
. * Download and install Stata ado files for pscore
. * net install st0026_2
.
. use C:\Econometrics\Data\matching_earnings

.
. * Define treatment, outcome, and independent variables
. global treatment TREAT

. global ylist RE78

. global xlist AGE EDUC MARR

. global breps 5

.
. * For difference-in-differences, outcome is the differences in outcomes after and
before
. * global ylist REDIFF
.
. describe $treatment $ylist $xlist

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
------------------
TREAT           float  %9.0g
RE78            float  %9.0g
AGE             float  %9.0g
EDUC            float  %9.0g
MARR            float  %9.0g

. summarize $treatment $ylist $xlist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       TREAT |      2675    .0691589    .2537716          0          1
        RE78 |      2675    20502.38    15632.52          0     121174
         AGE |      2675    34.22579    10.49984         17         55
        EDUC |      2675    11.99439    3.053556          0         17
        MARR |      2675    .8194393    .3847257          0          1

.
. bysort $treatment: summarize $ylist $xlist

------------------------------------------------------------------------------
------------------
```

```
-> TREAT = 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        RE78 |      2490    21553.92    15555.35          0     121174
         AGE |      2490     34.8506    10.44076         18         55
        EDUC |      2490    12.11687    3.082435          0         17
        MARR |      2490    .8662651    .3404357          0          1

-------------------------------------------------------------------------------
-------------------
-> TREAT = 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        RE78 |       185    6349.145    7867.405          0    60307.9
         AGE |       185    25.81622    7.155019         17         48
        EDUC |       185    10.34595     2.01065          4         16
        MARR |       185    .1891892    .3927217          0          1


.
. * Regression with a dummy variable for treatment (t-test)
. reg $ylist $treatment

      Source |       SS       df       MS              Number of obs =    2675
-------------+------------------------------           F(  1,  2673) =  173.41
       Model |  3.9811e+10     1  3.9811e+10           Prob > F      =  0.0000
    Residual |  6.1365e+11  2673   229573201           R-squared     =  0.0609
-------------+------------------------------           Adj R-squared =  0.0606
       Total |  6.5346e+11  2674   244375675           Root MSE      =   15152

------------------------------------------------------------------------------
        RE78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       TREAT |  -15204.78   1154.614   -13.17   0.000     -17468.8   -12940.75
       _cons |   21553.92   303.6414    70.98   0.000     20958.53    22149.32
------------------------------------------------------------------------------

.
. * Regression with a dummy variable for treatment controlling for x
. reg $ylist $treatment $xlist

      Source |       SS       df       MS              Number of obs =    2675
-------------+------------------------------           F(  4,  2670) =  171.99
       Model |  1.3388e+11     4  3.3470e+10           Prob > F      =  0.0000
    Residual |  5.1958e+11  2670   194600040           R-squared     =  0.2049
-------------+------------------------------           Adj R-squared =  0.2037
       Total |  6.5346e+11  2674   244375675           Root MSE      =   13950

------------------------------------------------------------------------------
        RE78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       TREAT |  -6901.379   1213.245    -5.69   0.000    -9280.375   -4522.384
         AGE |   176.6396   27.60029     6.40   0.000     122.5195    230.7597
```

```
        EDUC |   1900.046   91.77385    20.70   0.000     1720.091    2080.001
        MARR |   4937.016   800.0326     6.17   0.000     3368.269    6505.762
       _cons |  -11901.44   1703.792    -6.99   0.000    -15242.33   -8560.557
------------------------------------------------------------------------------

.
. * Propensity score matching with common support
. pscore $treatment $xlist, pscore(myscore) blockid(myblock) comsup



**************************************************
Algorithm to estimate the propensity score
**************************************************


The treatment is TREAT

      TREAT |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,490       93.08       93.08
          1 |        185        6.92      100.00
------------+-----------------------------------
      Total |      2,675      100.00


Estimation of the propensity score

Iteration 0:   log likelihood = -672.64954
Iteration 1:   log likelihood = -445.24322
Iteration 2:   log likelihood = -417.65611
Iteration 3:   log likelihood = -414.43392
Iteration 4:   log likelihood = -414.37065
Iteration 5:   log likelihood = -414.37062

Probit regression                                 Number of obs   =       2675
                                                  LR chi2(3)      =     516.56
                                                  Prob > chi2     =     0.0000
Log likelihood = -414.37062                       Pseudo R2       =     0.3840

------------------------------------------------------------------------------
       TREAT |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |  -.0479858   .0060942    -7.87   0.000    -.0599302   -.0360415
        EDUC |  -.1686459   .0192597    -8.76   0.000    -.2063942   -.1308977
        MARR |  -1.469566   .0995124   -14.77   0.000    -1.664607   -1.274526
       _cons |   2.729172   .3230556     8.45   0.000     2.095994    3.362349
------------------------------------------------------------------------------



Note: the common support option has been selected
The region of common support is [.000914, .80332316]
```

Description of the estimated propensity score
in region of common support

Estimated propensity score

	Percentiles	Smallest		
1%	.0010694	.000914		
5%	.0014751	.000914		
10%	.0020168	.000914	Obs	2332
25%	.0048323	.0009183	Sum of Wgt.	2332
50%	.0177077		Mean	.0796709
		Largest	Std. Dev.	.144896
75%	.0534804	.7537369		
90%	.3188462	.7681876	Variance	.0209948
95%	.4279398	.7897533	Skewness	2.41066
99%	.645166	.8033232	Kurtosis	8.272405

****
Step 1: Identification of the optimal number of blocks
Use option detail if you want more detailed output
****

The final number of blocks is 10

This number of blocks ensures that the mean propensity score
is not different for treated and controls in each blocks

****
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
****

The balancing property is satisfied

This table shows the inferior bound, the number of treated
and the number of controls for each block

Inferior of block of pscore	TREAT 0	1	Total
.000914	671	1	672
.00625	316	6	322
.0125	379	8	387
.025	341	14	355

```
       .05 |       135          8 |       143
        .1 |       110         12 |       122
        .2 |       137         61 |       198
        .4 |        51         51 |       102
        .6 |         7         23 |        30
        .8 |         0          1 |         1
-----------+----------------------+----------
     Total |     2,147        185 |     2,332

Note: the common support option has been selected


****************************************
End of the algorithm to estimate the pscore
****************************************


.
. * Matching methods
.
. * Nearest neighbor matching
. attnd $ylist $treatment $xlist, pscore(myscore) comsup boot reps($breps) dots


 The program is searching the nearest neighbor of each treated unit.
 This operation may take a while.



ATT estimation with Nearest Neighbor Matching method
(random draw version)
Analytical standard errors

------------------------------------------------------------
n. treat.   n. contr.         ATT    Std. Err.          t
------------------------------------------------------------

      185         431   -6715.251     1200.953     -5.592

------------------------------------------------------------
Note: the numbers of treated and controls refer to actual
nearest neighbour matches




Bootstrapping of standard errors

command:      attnd RE78 TREAT AGE EDUC MARR , pscore(myscore) comsup
statistic:    attnd      = r(attnd)
.....


Bootstrap statistics                            Number of obs    =      2675
                                                Replications     =         5
```

```
------------------------------------------------------------------------------
Variable     |  Reps  Observed      Bias  Std. Err. [95% Conf. Interval]
-------------+----------------------------------------------------------------
       attnd |     5 -6715.251  329.3118  917.7546  -9263.347  -4167.156   (N)
             |                                      -7521.834  -5499.719   (P)
             |                                      -7521.834  -5499.719  (BC)
------------------------------------------------------------------------------
Note:  N   = normal
       P   = percentile
       BC  = bias-corrected



ATT estimation with Nearest Neighbor Matching method
(random draw version)
Bootstrapped standard errors

-------------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
-------------------------------------------------------------

      185         431   -6715.251     917.755      -7.317

-------------------------------------------------------------
Note: the numbers of treated and controls refer to actual
nearest neighbour matches

.
. * Radius matching
. attr $ylist $treatment $xlist, pscore(myscore) comsup boot reps($breps) dots
radius(0.1)


 The program is searching for matches of treated units within radius.
 This operation may take a while.



ATT estimation with the Radius Matching method
Analytical standard errors

-------------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
-------------------------------------------------------------

      185        2147   -1.33e+04     652.019     -20.324

-------------------------------------------------------------
Note: the numbers of treated and controls refer to actual
matches within radius
```

```
Bootstrapping of standard errors

command:      attr RE78 TREAT AGE EDUC MARR , pscore(myscore) comsup radius(.1)
statistic:    attr       = r(attr)
.....


Bootstrap statistics                              Number of obs    =      2675
                                                  Replications     =         5

------------------------------------------------------------------------------
Variable     |  Reps  Observed      Bias  Std. Err. [95% Conf. Interval]
-------------+----------------------------------------------------------------
        attr |     5 -13251.41  612.6666  1070.465   -16223.5  -10279.32   (N)
             |                                      -14296.53  -11596.59   (P)
             |                                      -14296.53     -11994  (BC)
------------------------------------------------------------------------------
Note:  N   = normal
       P   = percentile
       BC  = bias-corrected



ATT estimation with the Radius Matching method
Bootstrapped standard errors

-------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
-------------------------------------------------------

      185        2147   -1.33e+04    1070.465     -12.379

-------------------------------------------------------
Note: the numbers of treated and controls refer to actual
matches within radius

.
. * Kernel Matching
. attk $ylist $treatment $xlist, pscore(myscore) comsup boot reps($breps) dots


 The program is searching for matches of each treated unit.
 This operation may take a while.



ATT estimation with the Kernel Matching method

-------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
-------------------------------------------------------

      185        2147   -7009.315           .           .
```

```
-------------------------------------------------------
Note: Analytical standard errors cannot be computed. Use
the bootstrap option to get bootstrapped standard errors.




Bootstrapping of standard errors

command:      attk RE78 TREAT AGE EDUC MARR , pscore(myscore) comsup bwidth(.06)
statistic:    attk       = r(attk)
.....


Bootstrap statistics                              Number of obs    =      2675
                                                  Replications     =         5

------------------------------------------------------------------------------
Variable     |  Reps  Observed      Bias  Std. Err. [95% Conf. Interval]
-------------+----------------------------------------------------------------
        attk |     5 -7009.315  42.95186  115.5546  -7330.146  -6688.484   (N)
             |                                      -7144.013  -6845.836   (P)
             |                                      -7144.013  -6896.341  (BC)
------------------------------------------------------------------------------
Note:  N   = normal
       P   = percentile
       BC  = bias-corrected



ATT estimation with the Kernel Matching method
Bootstrapped standard errors

-------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
-------------------------------------------------------

      185        2147   -7009.315     115.555     -60.658

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


.
. * Stratification Matching
. atts $ylist $treatment $xlist, pscore(myscore) blockid(myblock) comsup boot
reps($breps) dots



ATT estimation with the Stratification method
Analytical standard errors

-------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
-------------------------------------------------------
```

```
      184        2148   -6497.606           .           .

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



Bootstrapping of standard errors

command:      atts RE78 TREAT AGE EDUC MARR , pscore(myscore) blockid(myblock) comsup
statistic:    atts       = r(atts)
.....

note: label truncated to 80 characters

Bootstrap statistics                              Number of obs    =      2675
                                                  Replications     =         5

------------------------------------------------------------------------------
Variable     |  Reps  Observed      Bias  Std. Err. [95% Conf. Interval]
-------------+----------------------------------------------------------------
        atts |     5 -6497.605  235.7813  484.0005  -7841.406  -5153.805   (N)
             |                                      -7072.771  -5824.891   (P)
             |                                      -7072.771  -5988.378  (BC)
------------------------------------------------------------------------------
Note:  N   = normal
       P   = percentile
       BC  = bias-corrected



ATT estimation with the Stratification method
Bootstrapped standard errors

------------------------------------------------------------
n. treat.   n. contr.         ATT   Std. Err.           t
------------------------------------------------------------

      184        2148   -6497.605     484.000     -13.425

------------------------------------------------------------
```