```
      name:  <unnamed>
       log:  C:\Users\XuQi\Desktop\第7章.smcl
  log type:  smcl
 opened on:  16 Jul 2024, 09:38:43

. do 第7章.do, nostop

. use "C:\Users\XuQi\Desktop\cfps2010.dta", clear

.
. *一元线性回归
. reg lninc college, vce(cluster provcd)

Linear regression                               Number of obs     =      4,137
                                                F(1, 24)          =     271.17
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1095
                                                Root MSE          =     1.1498

                                   (Std. err. adjusted for 25 clusters in provcd)
------------------------------------------------------------------------------
             |               Robust
       lninc | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     college |    .823612   .0500155    16.47   0.000     .7203851     .926839
       _cons |   9.353189   .1084703    86.23   0.000     9.129317    9.577061
------------------------------------------------------------------------------

.
. *倾向值细分
. tabulate fmedu, gen(fmedu)

  父母是否上过高中 |      Freq.     Percent        Cum.
------------------+-----------------------------------
                否 |      2,327       56.25       56.25
                是 |        898       21.71       77.96
              缺失 |        912       22.04      100.00
------------------+-----------------------------------
             Total |      4,137      100.00

. pscore college hukou age gender race sibling fmedu2 fmedu3, pscore(ps) blockid(strata) logit comsup


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


The treatment is college

  是否上大学 |      Freq.     Percent        Cum.
-------------+-----------------------------------
          否 |      2,494       60.29       60.29
          是 |      1,643       39.71      100.00
-------------+-----------------------------------
       Total |      4,137      100.00


Estimation of the propensity score

Iteration 0:  Log likelihood = -2779.3946
Iteration 1:  Log likelihood = -2437.5718
Iteration 2:  Log likelihood = -2433.2339
Iteration 3:  Log likelihood = -2433.2257
Iteration 4:  Log likelihood = -2433.2257

Logistic regression                             Number of obs     =      4137
                                                LR chi2(7)        =    692.34
                                                Prob > chi2       =    0.0000
Log likelihood = -2433.2257                     Pseudo R2         =    0.1245
```

college	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
hukou	.3694308	.0750791	4.92	0.000	.2222784	.5165832
age	-.0779021	.0046121	-16.89	0.000	-.0869417	-.0688625
gender	-.0055885	.0707311	-0.08	0.937	-.1442189	.1330418
race	-.1815893	.1555612	-1.17	0.243	-.4864837	.123305
sibling	.1551052	.1050747	1.48	0.140	-.0508374	.3610478
fmedu2	.8096408	.0889766	9.10	0.000	.6352498	.9840318
fmedu3	.1445431	.0935233	1.55	0.122	-.0387591	.3278453
_cons	2.403783	.2336584	10.29	0.000	1.945821	2.861745

Note: the common support option has been selected
The region of common support is [.11224518, .85697725]

Description of the estimated propensity score
in region of common support

Estimated propensity score

	Percentiles	Smallest		
1%	.1202454	.1122452		
5%	.1458332	.1122452		
10%	.1663307	.1122452	Obs	4,137
25%	.2160295	.1122452	Sum of wgt.	4,137
50%	.3705058		Mean	.3971477
		Largest	Std. dev.	.19613
75%	.5462371	.8367993		
90%	.6952771	.8368947	Variance	.038467
95%	.7616237	.8368947	Skewness	.4634977
99%	.8213217	.8569773	Kurtosis	2.12405

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

The final number of blocks is 8

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
****

Variable hukou is not balanced in block 2

Variable age is not balanced in block 2

Variable age is not balanced in block 3

Variable age is not balanced in block 5

Variable fmedu2 is not balanced in block 6

Variable hukou is not balanced in block 7

Variable sibling is not balanced in block 7

The balancing property is not satisfied

Try a different specification of the propensity score

Inferior of block of pscore	是否上大学 否	是	Total
0	**692**	**112**	**804**
.2	**370**	**86**	**456**
.25	**263**	**111**	**374**
.3	**374**	**258**	**632**
.4	**314**	**247**	**561**
.5	**252**	**307**	**559**
.6	**213**	**439**	**652**
.8	**16**	**83**	**99**
Total	**2,494**	**1,643**	**4,137**

Note: the common support option has been selected

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

```
.
. drop ps strata comsup
. gen hukouage=hukou*age
. pscore college hukou hukouage age age2 gender race sibling fmedu2 fmedu3, pscore(ps) blockid(strata) logit comsup
```

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

The treatment is college

是否上大学	Freq.	Percent	Cum.
否	**2,494**	**60.29**	**60.29**
是	**1,643**	**39.71**	**100.00**
Total	**4,137**	**100.00**	

Estimation of the propensity score

```
Iteration 0:  Log likelihood = -2779.3946
Iteration 1:  Log likelihood = -2429.5513
Iteration 2:  Log likelihood =  -2422.173
Iteration 3:  Log likelihood = -2422.1088
Iteration 4:  Log likelihood = -2422.1087

Logistic regression                               Number of obs   =       4137
                                                  LR chi2(9)      =     714.57
                                                  Prob > chi2     =     0.0000
Log likelihood = -2422.1087                       Pseudo R2       =     0.1285
```

college	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
hukou	**.7612414**	**.3489558**	**2.18**	**0.029**	**.0773006**	**1.445182**
hukouage	**-.010548**	**.0087546**	**-1.20**	**0.228**	**-.0277067**	**.0066107**
age	**.1150714**	**.0429697**	**2.68**	**0.007**	**.0308523**	**.1992905**
age2	**-.0024131**	**.000549**	**-4.40**	**0.000**	**-.0034892**	**-.001337**
gender	**-.0093303**	**.0708578**	**-0.13**	**0.895**	**-.148209**	**.1295483**
race	**-.1836443**	**.1558151**	**-1.18**	**0.239**	**-.4890363**	**.1217476**
sibling	**.2137055**	**.1100417**	**1.94**	**0.052**	**-.0019724**	**.4293833**
fmedu2	**.8039633**	**.0889304**	**9.04**	**0.000**	**.6296628**	**.9782638**
fmedu3	**.198138**	**.095117**	**2.08**	**0.037**	**.0117121**	**.3845639**
_cons	**-1.281155**	**.8326789**	**-1.54**	**0.124**	**-2.913175**	**.350866**

Note: the common support option has been selected
The region of common support is [.07981392, .83238747]

Description of the estimated propensity score
in region of common support

Estimated propensity score

	Percentiles	Smallest		
1%	**.0913784**	**.0798139**		
5%	**.1134165**	**.0798139**		
10%	**.1397529**	**.0798139**	Obs	**4,137**
25%	**.2179474**	**.0798139**	Sum of wgt.	**4,137**
50%	**.4006644**		Mean	**.3971477**
		Largest	Std. dev.	**.1977145**
75%	**.530137**	**.8223759**		
90%	**.6705877**	**.8264575**	Variance	**.039091**
95%	**.7619456**	**.8264575**	Skewness	**.2716646**
99%	**.8007715**	**.8323875**	Kurtosis	**2.080826**

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

The final number of blocks is 9

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	是否上大学 否	是	Total
.0798139	**775**	**121**	**896**
.2	**252**	**55**	**307**
.25	**215**	**85**	**300**
.3	**334**	**220**	**554**
.4	**480**	**393**	**873**
.5	**194**	**213**	**407**
.6	**168**	**285**	**453**
.7	**69**	**223**	**292**
.8	**7**	**48**	**55**
Total	**2,494**	**1,643**	**4,137**

Note: the common support option has been selected

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

```
.
. atts lninc college, pscore(ps) blockid(strata) comsup


ATT estimation with the Stratification method
Analytical standard errors
```

n. treat.	n. contr.	ATT	Std. Err.	t
1643	**2494**	**0.691**	**0.043**	**15.984**

```
.
. *干预效应的异质性
. hte sm lninc college hukou hukouage age age2 gender race sibling fmedu2 fmedu3, logit comsup

                                                      Number of obs =     4137
```

lninc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TE by strata						
1	**1.127695**	**.1261502**	**8.94**	**0.000**	**.8804449**	**1.374945**
2	**1.022366**	**.1455883**	**7.02**	**0.000**	**.737018**	**1.307714**
3	**1.060989**	**.1569723**	**6.76**	**0.000**	**.7533288**	**1.368649**
4	**.7679974**	**.0852362**	**9.01**	**0.000**	**.6009374**	**.9350574**
5	**.5875918**	**.0806062**	**7.29**	**0.000**	**.4296066**	**.745577**
6	**.5250521**	**.112322**	**4.67**	**0.000**	**.304905**	**.7451992**
7	**.6410306**	**.0995247**	**6.44**	**0.000**	**.4459658**	**.8360953**
8	**.6611019**	**.1463207**	**4.52**	**0.000**	**.3743187**	**.9478852**
9	**.216667**	**.3767465**	**0.58**	**0.565**	**-.5217426**	**.9550766**
Linear trend						
_slope	**-.0853021**	**.0192192**	**-4.44**	**0.000**	**-.122971**	**-.0476333**
_cons	**1.141173**	**.0992395**	**11.50**	**0.000**	**.9466673**	**1.335679**

```
TE = treatment effect

.
. hte ms lninc college hukou hukouage age age2 gender race sibling fmedu2 fmedu3, logit common noscatter lpolyci
(running psmatch2 ...)

Logistic regression                                     Number of obs =  4,137
                                                        LR chi2(9)    = 714.57
                                                        Prob > chi2   = 0.0000
Log likelihood = -2422.1087                             Pseudo R2     = 0.1285
```

college	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
hukou	**.7612414**	**.3489558**	**2.18**	**0.029**	**.0773006**	**1.445182**
hukouage	**-.010548**	**.0087546**	**-1.20**	**0.228**	**-.0277067**	**.0066107**
age	**.1150714**	**.0429697**	**2.68**	**0.007**	**.0308523**	**.1992905**
age2	**-.0024131**	**.000549**	**-4.40**	**0.000**	**-.0034892**	**-.001337**
gender	**-.0093303**	**.0708578**	**-0.13**	**0.895**	**-.148209**	**.1295484**
race	**-.1836443**	**.1558151**	**-1.18**	**0.239**	**-.4890363**	**.1217476**
sibling	**.2137055**	**.1100417**	**1.94**	**0.052**	**-.0019724**	**.4293833**
fmedu2	**.8039633**	**.0889304**	**9.04**	**0.000**	**.6296628**	**.9782638**
fmedu3	**.198138**	**.095117**	**2.08**	**0.037**	**.0117121**	**.3845639**
_cons	**-1.281155**	**.8326789**	**-1.54**	**0.124**	**-2.913175**	**.350866**

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
lninc	Unmatched	**10.1768011**	**9.35318907**	**.823612033**	**.03653383**	**22.54**
	ATT	**10.1782293**	**9.42346546**	**.75476386**	**.0881435**	**8.56**
	ATU	**9.35318907**	**10.1910798**	**.837890778**	**.**	**.**
	ATE			**.804937755**	**.**	**.**

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**	**2,494**	**2,494**
Treated	**5**	**1,638**	**1,643**
Total	**5**	**4,132**	**4,137**

```
(compiling HTE graph ...)

. hte sd lninc college hukou hukouage age age2 gender race sibling fmedu2 fmedu3, logit comsup

.
. *倾向值加权（手动实现）
. qui logit college hukou hukou##c.age c.age##c.age gender race sibling fmedu2 fmedu3

. predict p
(option pr assumed; Pr(college))

.
. gen w_ate=1/p if college==1
(2,494 missing values generated)

. replace w_ate=1/(1-p) if college==0
(2,494 real changes made)

.
. gen w_att=1 if college==1
(2,494 missing values generated)

. replace w_att=p/(1-p) if college==0
(2,494 real changes made)

.
. gen w_atu=(1-p)/p if college==1
(2,494 missing values generated)

. replace w_atu=1 if college==0
(2,494 real changes made)

.
. reg lninc college [pw=w_ate], vce(cluster provcd)
(sum of wgt is 8,296.73905217648)

Linear regression                               Number of obs     =      4,137
                                                F(1, 24)          =     238.29
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1157
                                                Root MSE          =     1.1086

                                   (Std. err. adjusted for 25 clusters in provcd)
```

lninc	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
college	**.801988**	**.0519531**	**15.44**	**0.000**	**.694762**	**.9092139**
_cons	**9.405058**	**.1124885**	**83.61**	**0.000**	**9.172893**	**9.637223**

```
. reg lninc college [pw=w_att], vce(cluster provcd)
(sum of wgt is 3,285.49713142961)

Linear regression                               Number of obs     =      4,137
                                                F(1, 24)          =     184.39
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0890
                                                Root MSE          =     1.1085

                                   (Std. err. adjusted for 25 clusters in provcd)
```

lninc	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
college	**.6929841**	**.0510332**	**13.58**	**0.000**	**.5876567**	**.7983114**
_cons	**9.483817**	**.1180324**	**80.35**	**0.000**	**9.24021**	**9.727424**

```
. reg lninc college [pw=w_atu], vce(cluster provcd)
(sum of wgt is 5,011.24193204939)

Linear regression                               Number of obs     =      4,137
                                                F(1, 24)          =     237.54
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1348
                                                Root MSE          =      1.107

                                   (Std. err. adjusted for 25 clusters in provcd)
------------------------------------------------------------------------------
             |               Robust
      lninc  | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     college |   .8735978   .0566819    15.41   0.000     .7566121    .9905836
       _cons |   9.353189   .1084703    86.23   0.000     9.129317    9.577061
------------------------------------------------------------------------------

.
. *倾向值加权（teffects ipw）
. teffects ipw (lninc) (college hukou hukou##c.age c.age##c.age gender race sibling fmedu2 fmedu3)
note: 1.hukou omitted because of collinearity.
note: age omitted because of collinearity.
Iteration 0:  EE criterion =  3.679e-21
Iteration 1:  EE criterion =  5.271e-31

Treatment-effects estimation                    Number of obs     =      4,137
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
------------------------------------------------------------------------------
             |               Robust
      lninc  | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
     college |
 (是 vs 否)  |   .801988   .0379528    21.13   0.000     .7276019    .8763741
-------------+----------------------------------------------------------------
POmean       |
     college |
          否 |   9.405058   .0271151   346.86   0.000     9.351913    9.458203
------------------------------------------------------------------------------

. teffects ipw (lninc) (college hukou hukou##c.age c.age##c.age gender race sibling fmedu2 fmedu3), atet
note: 1.hukou omitted because of collinearity.
note: age omitted because of collinearity.

Iteration 0:  EE criterion =  3.679e-21
Iteration 1:  EE criterion =  1.176e-31

Treatment-effects estimation                    Number of obs     =      4,137
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
------------------------------------------------------------------------------
             |               Robust
      lninc  | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
     college |
 (是 vs 否)  |   .6929841   .0417058    16.62   0.000     .6112421     .774726
-------------+----------------------------------------------------------------
POmean       |
     college |
          否 |   9.483817   .0358907   264.24   0.000     9.413472    9.554162
------------------------------------------------------------------------------
```

```
.
. *平衡性检验
. reg age college [pw=w_ate], vce(cluster provcd)
(sum of wgt is 8,296.73905217648)

Linear regression                               Number of obs     =      4,137
                                                F(1, 24)          =       0.08
                                                Prob > F          =     0.7796
                                                R-squared         =     0.0001
                                                Root MSE          =     9.1311

                                  (Std. err. adjusted for 25 clusters in provcd)
------------------------------------------------------------------------------
             |               Robust
         age |      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     college |   .1454856   .5140792     0.28   0.780    -.9155217    1.206493
       _cons |   39.65485   .4101642    96.68   0.000     38.80831    40.50138
------------------------------------------------------------------------------

.
. qui teffects ipw (lninc) (college hukou hukou##c.age c.age##c.age gender race sibling fmedu2 fmedu3)

. tebalance summarize

Covariate balance summary

                      Raw              Weighted
Number of obs =       4,137             4,137.0
Treated obs   =       1,643             2,074.4
Control obs   =       2,494             2,062.6

------------------------------------------------------------------------------
                     |  Standardized differences      Variance ratio
                     |        Raw       Weighted        Raw       Weighted
---------------------+--------------------------------------------------------
              hukou  |   .2397965        .01079    1.082133       1.004144
                age  |  -.7815923     .0159334     .816613       1.034943
                     |
             hukou#  |
                age  |
      城镇户口       |   .0759021     .0149513    .7676175       1.022716
                     |
             age#    |
                age  |  -.7903714     .0196792    .7091451       1.050748
                     |
             gender  |  -.0820793     .0237794    1.025374       .9914028
               race  |  -.0477992    -.0021336    1.211041       1.008475
            sibling  |   .3643434     .0024262    2.011046       1.004868
             fmedu2  |   .5419858    -.0040888    2.039276       .9943991
             fmedu3  |  -.2528053    -.0025766    .6928676       .9964834
------------------------------------------------------------------------------

. tebalance density age

. tebalance overid, nolog

Overidentification test for covariate balance
H0: Covariates are balanced

         chi2(10)     =  24.7885
         Prob > chi2  =   0.0058

. tebalance overid, bconly nolog

Overidentification test for covariate balance
H0: Covariates are balanced

         chi2(8)      =  15.0045
         Prob > chi2  =   0.0591
```

```
.
. *双重稳健估计(ipwra)
. teffects ipwra (lninc hukou age gender race sibling fmedu2 fmedu3) ///
>                (college hukou hukou##c.age c.age##c.age gender race sibling fmedu2 fmedu3)
note: 1.hukou omitted because of collinearity.
note: age omitted because of collinearity.

Iteration 0:  EE criterion =  3.679e-21
Iteration 1:  EE criterion =  8.189e-31

Treatment-effects estimation                    Number of obs     =      4,137
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: logit
```

lninc	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
college (是 vs 否)	**.7980164**	**.0377585**	**21.13**	**0.000**	**.724011**	**.8720217**
POmean						
college 否	**9.404692**	**.0271286**	**346.67**	**0.000**	**9.351521**	**9.457863**

```
. teffects aipw (lninc hukou age gender race sibling fmedu2 fmedu3) ///
>               (college hukou hukou##c.age c.age##c.age gender race sibling fmedu2 fmedu3)
note: 1.hukou omitted because of collinearity.
note: age omitted because of collinearity.

Iteration 0:  EE criterion =  3.679e-21
Iteration 1:  EE criterion =  4.434e-31

Treatment-effects estimation                    Number of obs     =      4,137
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit
```

lninc	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
college (是 vs 否)	**.7972419**	**.0378972**	**21.04**	**0.000**	**.7229646**	**.8715191**
POmean						
college 否	**9.405056**	**.027155**	**346.35**	**0.000**	**9.351833**	**9.458279**

```
.
.
end of do-file

. log close
      name:  <unnamed>
       log:  C:\Users\XuQi\Desktop\第7章.smcl
  log type:  smcl
 closed on:  16 Jul 2024, 09:39:42
```