ECON 717A: Problem Set 2

Alex von Hafften

March 6, 2022

1 Write-Up

Problem 0

I drop observations with `sample` equal to 3. This step dropped 2,490 observations.

Problem 1

I regressed earnings in 1978 on treatment with and without the covariate of age, age squared, education, indicators for black, Hispanic, married, and no degree, and earnings in 1974 and 1975. The treatment effect is \$886.30 without covariates and \$818.70 with covariates with both statistically significant at the 10 percent level. It is important to include covariate even in experimental data because we get a more precise estimate for the treatment effect.

	(1)	(2)
VARIABLES	re78	re78
treated	886.3*	818.7*
	(488.1)	(487.8)
age		-145.9
		(200.8)
age_2		2.799
		(3.246)
educ		206.8
		(165.5)
black		-1,461**
		(734.3)
hisp		100.5
		(958.6)
married		133.9
		(660.0)
nodegree		-405.9
		(752.1)
re74		0.0871
		(0.106)
re75		0.0840
		(0.119)
Constant	5,090***	5,649
	(277.4)	(3,757)
Observations	722	722
R-squared	0.005	0.045

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Problem 2

I drop observations with `sample` equal to 1 and `treated` equal to 1. This step dropped 297 observations.

Problem 3

I define `in_control` equal to one if `sample` equals one and zero otherwise. The probit estimation for the propensity scores are the coarse scores and the rich scores are below.

VARIABLES	(1) in_control	(2) in_control
age	0.253***	0.322***
	(0.0293)	(0.0316)
age_2	-0.00453***	-0.00548***
	(0.000493)	(0.000530)
educ	0.0169	0.0178
	(0.0181)	(0.0183)
black	1.990***	1.950***
	(0.0778)	(0.0796)
hisp	0.973***	0.978***
	(0.103)	(0.106)
married	-1.101***	-0.909***
	(0.0826)	(0.0869)
nodegree	1.133***	1.071***
	(0.100)	(0.104)
re74		-1.07e-06
		(8.60e-06)
re75		-5.76e-05***
		(9.56e-06)
Constant	-6.358***	-7.108***
	(0.483)	(0.509)
Observations	16,417	16,417
Comparison group obs. completely determined	727	1359
Control group obs. completely determined	0	0

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The completely determined observations are 727 and 1359 comparison group observations that have propensity scores with almost zero.

Problem 4

The table below shows descriptive statistics of `pscorea` and `pscoreb`.

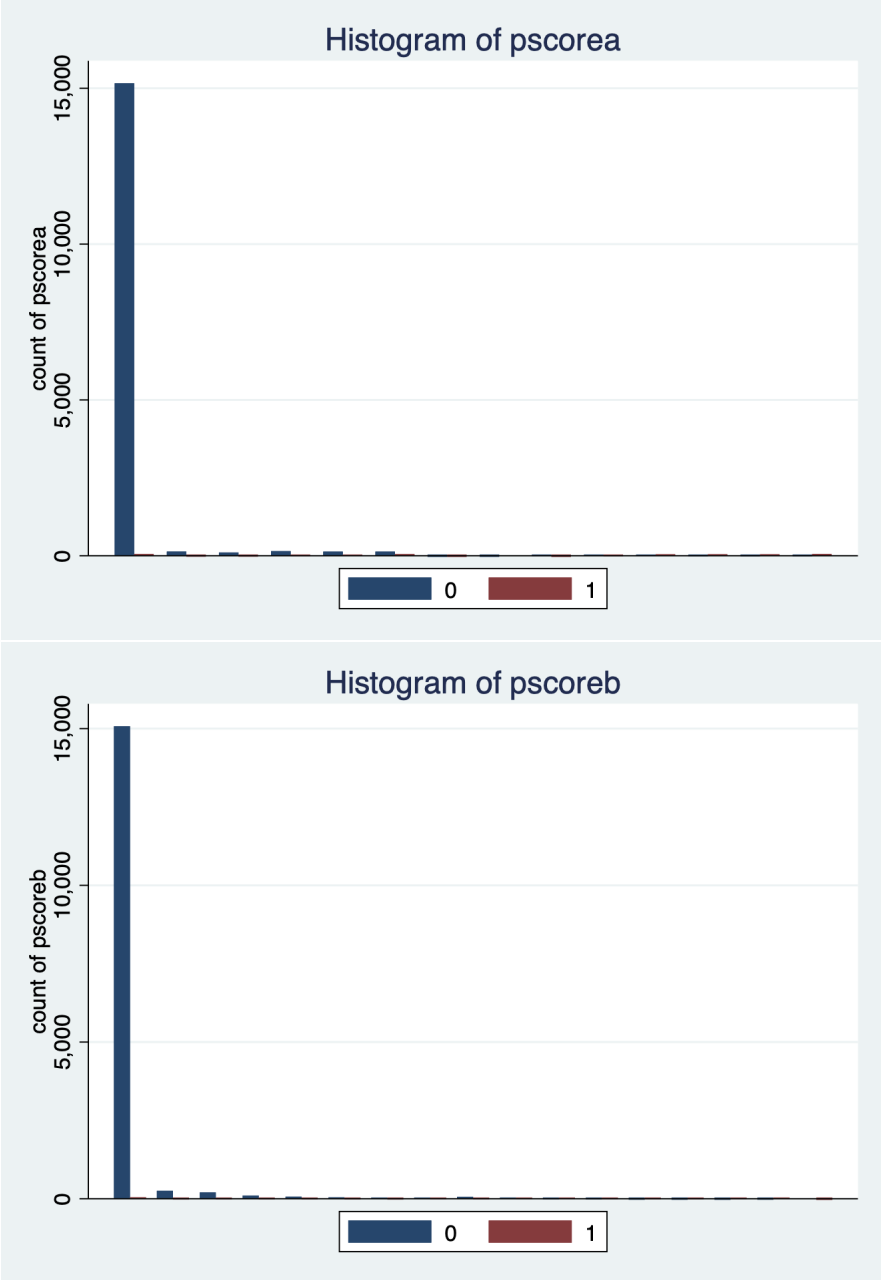
	Mean	SD	Min	Median	Max	N
0	0.02	0.07	0.00	0.00	0.69	15,992
1	0.39	0.23	0.00	0.47	0.69	425
Total	0.03	0.09	0.00	0.00	0.69	16,417

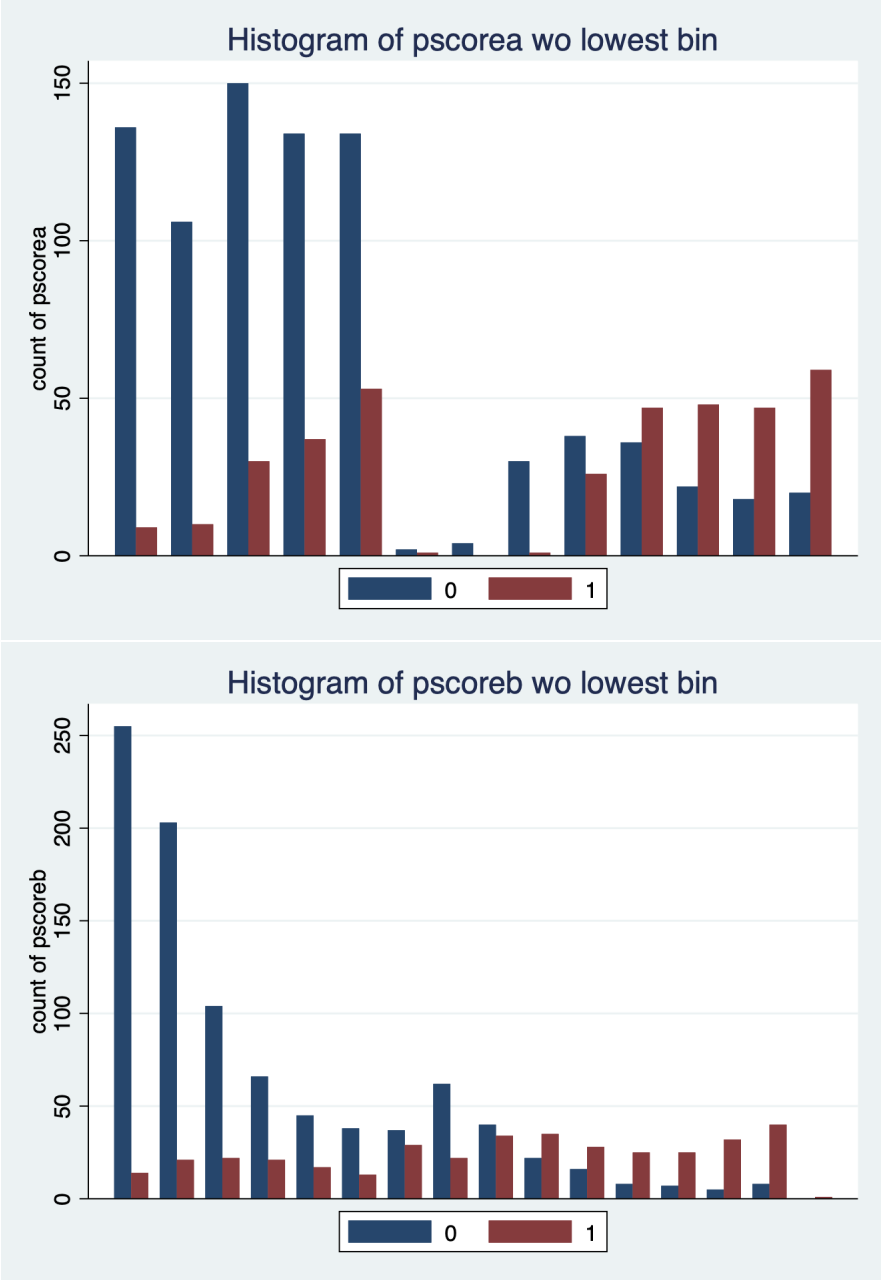
	Mean	SD	Min	Median	Max	N
0	0.02	0.06	0.00	0.00	0.79	15,992
1	0.42	0.25	0.00	0.46	0.80	425
Total	0.03	0.10	0.00	0.00	0.80	16,417

These descriptive statistics suggest that imposing the common support condition will not drop many observations. For `pscorea`, the minimum is zero for both groups and the maximum is 0.69 for both groups. This suggests that there are compared observations within the comparison group to the control group. Similarly for `pscoreb`, the minimum is zero for both groups and the maximum is around 0.8 for both groups. These descriptive statistics suggest that the CPS group is not very comparable to the experimental control group. That is, these descriptive statistics highlight the importance of matching because the control group has a much higher mean propensity score than the comparison group (i.e., 0.39 vs. 0.02 for `pscorea` and 0.42 vs. 0.02 for `proscoreb`).

Problem 5

Below are histograms (based on the provided code) for `pscorea` and `pscoreb` across the CPS comparison group (i.e., 0) and the experimental control group (i.e., 1) with and without the lowest bin (i.e., `pscorex` $\in (0.0, 0.05)$). Looking at the first two histograms, most of the comparison group observations have a very low estimated propensity score (confirming the descriptive statistics), so the comparison group is not very comparable to the control group. I added histograms without the lowest bin because many observations in the comparison group have estimated propensity scores of basically zero impairing our ability to evaluate the common support condition. Looking at the second two histograms, the common support condition seems to be well satisfied with comparison group observations at all propensity levels of the control group. These histograms highlight concerns about matching without replacement because the number of observations in the control group with high propensity scores is larger than the number of observations in the comparison group with large propensity scores (i.e. on the right of the graphs the red bars are taller than the blue bars). Matching without replacement would cause these the control group observations to be matched with comparison group observations with significantly lower propensity scores.


Histogram of pscorea
count of pscorea
0
5,000
10,000
15,000
0
1
Histogram of pscoreb
count of pscoreb
0
5,000
10,000
15,000
0
1


Histogram of pscorea wo lowest bin
count of pscorea
0
50
100
150
0
1
Histogram of pscoreb wo lowest bin
count of pscoreb
0
50
100
150
200
250
0
1

Problem 6

Using single nearest neighbor matching without replacement and common support condition, I estimate the non-experimental bias reported in the table below. The non-experimental bias is significant and negative indicating that even after matching observations in the control group that significantly lower earnings in 1978. The non-experimental bias is lower when matching based on the rich propensity scores. No observations are dropped when matching with coarse propensity scores. Seven observations are dropped when matching with fine propensity scores.

	Unmatched	ATT for `pscorea`	ATT for `pscoreb`
Difference	-9756.610000000001	-4439.07	-2340.78
SE	470.16	486.48	449.37

Problem 7

Using single nearest neighbor matching with replacement and common support condition, I estimate the non-experimental bias reported in the table below. The non-experimental bias is still significant and negative but it is lower than without replacement. As mentioned in problem 5, this reduction in the non-experimental bias is due the relatively fewer comparison group observations with high propensity scores, so that there's more non-experimental bias when matching without replacement.

	Unmatched	ATT for `pscorea`	ATT for `pscoreb`
Difference	-9756.610000000001	-3677.03	-1515.99
SE	470.16	934.5	707.62

Problem 8

The standardized differences are in the table below. Conditioning reduces the standardized bias for `re74` by 8 percent and for `re75` by 34 percent.

	`re74`	`re75`
Raw Data	1.26	1.41
Matched by `pscoreb` using Single Nearest Neighbor	1.16	0.98

Problem 9

Matching using the Gaussian kernel results in larger estimates of the non-experimental bias. As the bandwidth increases, the estimate of the non-experimental bias grows but the SE of the estimate decreases. This is probably due to putting positive weight on the mass of observations in the comparison group with propensity scores around zero.

	Bandwidth = 0.02	Bandwidth = 0.2	Bandwidth = 2.0
Difference	-2349.09	-7043.43	-9772.130000000001
SE	658.53	337.06	289.25

Problem 10

Using local linear matching, the estimates for the non-experimental bias is more-or-less in line with previous estimates for the two smaller bandwidths, but the estimate for the larger bandwidth is opposite sign.

	Bandwidth = 0.02	Bandwidth = 0.2	Bandwidth = 2.0
Difference	-1980.88	-2421.42	993.88
SE	707.62	707.62	707.62

Problem 11

In the table below is a regression with a dummy for `in_control`. The coefficient estimate is on the low side compared to the other estimates, but it is more-or-less in line with the previous results.

	(1)
VARIABLES	re78
in_control	-1,853***
	(343.2)
age	-235.5***
	(40.08)
age_2	1.858***
	(0.550)
educ	163.3***
	(28.53)
black	-832.4***
	(193.9)
hisp	-114.2
	(213.9)
married	199.4
	(150.8)
nodegree	296.6*
	(174.3)
re74	0.291***
	(0.0152)
re75	0.471***
	(0.0153)
Constant	7,757***
	(726.7)
Observations	16,417
R-squared	0.483

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Below is a table of summary statistics of the predicted values split by whether `in_control` equals one or zero.

	Mean	SD	Min	Median	Max	N
0	14,846.66	6,608.11	241.16	15,369.08	25,559.22	15,992
1	5,090.05	4,227.39	-1,476.13	3,915.22	31,078.45	425
Total	14,594.08	6,737.87	-1,476.13	15,111.14	31,078.45	16,417

Problem 12

In the table below is a regression the typical covariate but only on the comparison group.

VARIABLES	(1) re78
age	-252.0***
	(40.92)
age_2	2.041***
	(0.561)
educ	166.6***
	(28.79)
black	-773.9***
	(199.6)
hisp	-168.2
	(218.9)
married	244.1
	(153.1)
nodegree	330.7*
	(176.9)
re74	0.299***
	(0.0152)
re75	0.470***
	(0.0153)
Constant	7,908***
	(739.7)
Observations	15,992
R-squared	0.476

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Below is a table of summary statistics of the predicted values split by whether `in_control` equals one (out of sample) or zero (in sample). While the mean predicted value is different compared to problem 11, the difference is statistically insignificant.

	Mean	SD	Min	Median	Max	N
0	14,846.66	6,654.46	162.96	15,370.95	25,693.43	15,992
1	6,941.60	4,269.45	275.72	5,770.59	33,124.62	425
Total	14,642.01	6,721.77	162.96	15,110.59	33,124.62	16,417

Problem 13

The inverse probability weighting estimate for the treatment effect on the treated is -1420.38 without rescaling and -9756.61 with rescaling. In general, these estimates are in line with the previous estimates.

2 Stata Log File

```
------------------------------------------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/analysis.smcl
  log type:  smcl
 opened on:   6 Mar 2022, 16:02:41

. 
. * ECON 717A: Applied Econometrics

. * Problem set 2

. * Professor: Jeff Smith

. * Alex von Hafften

. * Matching and weighting

. 
. * clear workspace

. clear

. 
. * install user defined functions (if needed)

. ssc install outreg2
checking outreg2 consistency and verifying not already installed...
all files already exist and are up to date.

. ssc install psmatch2
checking psmatch2 consistency and verifying not already installed...
all files already exist and are up to date.

. ssc install texdoc
checking texdoc consistency and verifying not already installed...
all files already exist and are up to date.
```

```
. ssc install stddiff
checking stddiff consistency and verifying not already installed...
all files already exist and are up to date.

.
. * change working directory

. cd "/Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/"
/Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2

.
. * open dataset

. use "Economics 717 Spring 2022 NSW Data.dta"

.
. ******************************************************************************

. * problem #0 - drop observation from the PSID

. ******************************************************************************

.
. drop if sample == 3
(2,490 observations deleted)

.
. ******************************************************************************

. * problem #1 - treatment effect from experimental data

. ******************************************************************************

.
. gen age_2 = age^2

.
. regress re78 treated, robust
```

```
Linear regression                               Number of obs     =        722
                                                F(1, 720)         =       3.30
                                                Prob > F          =     0.0698
                                                R-squared         =     0.0049
                                                Root MSE          =       6242

------------------------------------------------------------------------------
             |               Robust
        re78 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     treated |   886.3038   488.1385     1.82   0.070    -72.04111    1844.649
       _cons |   5090.048    277.426    18.35   0.000     4545.388    5634.709
------------------------------------------------------------------------------

. outreg2 using table_1, tex(frag) replace
table_1.tex
dir : seeout

.
. regress re78 treated age age_2 educ black hisp married nodegree re74 re75, robust

Linear regression                               Number of obs     =        722
                                                F(10, 711)        =       2.75
                                                Prob > F          =     0.0025
                                                R-squared         =     0.0454
                                                Root MSE          =     6152.3

------------------------------------------------------------------------------
             |               Robust
        re78 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     treated |   818.7003   487.8295     1.68   0.094    -139.0582    1776.459
         age |  -145.9217   200.7603    -0.73   0.468    -540.0755    248.2322
       age_2 |   2.799479   3.245728     0.86   0.389     -3.57288    9.171837
        educ |   206.8112   165.4509     1.25   0.212    -118.0195     531.642
       black |  -1461.261   734.3153    -1.99   0.047    -2902.947   -19.57563
        hisp |   100.4831   958.5652     0.10   0.917    -1781.474     1982.44
     married |   133.9094   660.0178     0.20   0.839    -1161.908    1429.726
    nodegree |   -405.909   752.0756    -0.54   0.590    -1882.464    1070.646
```

```
        re74 |   .0871344   .1061708     0.82   0.412    -.1213114    .2955802
        re75 |   .0839624   .1188638     0.71   0.480    -.1494036    .3173283
       _cons |    5648.81   3757.499     1.50   0.133     -1728.31    13025.93
------------------------------------------------------------------------------

. outreg2 using table_1, tex(frag) append
table_1.tex
dir : seeout

.
. ********************************************************************************

. * problem #2 - drop the experimental treatment group

. ********************************************************************************

.
. drop if treated == 1  & sample==1
(297 observations deleted)

.
. ********************************************************************************

. * problem #3 - estimate propensity scores

. ********************************************************************************

.
. gen in_control = (sample ==1)

.
. probit in_control age age_2 educ black hisp married nodegree

Iteration 0:   log likelihood = -1972.3937
Iteration 1:   log likelihood = -1152.9501
Iteration 2:   log likelihood = -946.08275
Iteration 3:   log likelihood = -919.36768
Iteration 4:   log likelihood = -917.89575
Iteration 5:   log likelihood =  -917.8945
```

```
Iteration 6:   log likelihood =  -917.8945

Probit regression                                       Number of obs =  16,417
                                                        LR chi2(7)    = 2109.00
                                                        Prob > chi2   =  0.0000
Log likelihood = -917.8945                              Pseudo R2     =  0.5346

------------------------------------------------------------------------------
  in_control | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .2532387   .0293232     8.64   0.000     .1957664    .3107111
       age_2 |  -.0045319   .0004929    -9.19   0.000    -.0054981   -.0035658
        educ |   .0168833   .0180737     0.93   0.350    -.0185405     .052307
       black |   1.989908   .0778422    25.56   0.000      1.83734    2.142476
        hisp |   .9732994   .1033626     9.42   0.000     .7707124    1.175886
     married |  -1.101057   .0826205   -13.33   0.000     -1.26299   -.9391236
    nodegree |   1.132688   .1004559    11.28   0.000     .9357984    1.329578
       _cons |     -6.358   .4834229   -13.15   0.000    -7.305492   -5.410509
------------------------------------------------------------------------------
Note: 727 failures and 0 successes completely determined.

. outreg2 using table_3, tex(frag) replace addstat(Comparison group obs. completely determined, e(N_cdf), Control group obs. completely de
table_3.tex
dir : seeout

. predict pscorea, pr

.
. probit in_control age age_2 educ black hisp married nodegree re74 re75

Iteration 0:   log likelihood = -1972.3937
Iteration 1:   log likelihood = -1134.7468
Iteration 2:   log likelihood = -901.47114
Iteration 3:   log likelihood =  -864.6975
Iteration 4:   log likelihood = -862.49864
Iteration 5:   log likelihood = -862.48901
Iteration 6:   log likelihood = -862.48901

Probit regression                                       Number of obs =  16,417
```

```
                                                  LR chi2(9)    = 2219.81
                                                  Prob > chi2   =  0.0000
Log likelihood = -862.48901                       Pseudo R2     =  0.5627

------------------------------------------------------------------------------
  in_control | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .3224723    .031638    10.19   0.000     .2604629    .3844817
       age_2 |  -.0054825   .0005302   -10.34   0.000    -.0065216   -.0044433
        educ |   .0178156   .0182703     0.98   0.330    -.0179934    .0536247
       black |   1.950403   .0795723    24.51   0.000     1.794444    2.106362
        hisp |   .9775352   .1061721     9.21   0.000     .7694418    1.185629
     married |  -.9090639    .086934   -10.46   0.000    -1.079451   -.7386764
    nodegree |   1.071194   .1039852    10.30   0.000     .8673865    1.275001
        re74 |  -1.07e-06   8.60e-06    -0.12   0.901    -.0000179    .0000158
        re75 |  -.0000576   9.56e-06    -6.03   0.000    -.0000764   -.0000389
       _cons |  -7.108113   .5090983   -13.96   0.000    -8.105928   -6.110299
------------------------------------------------------------------------------
Note: 1359 failures and 0 successes completely determined.

. outreg2 using table_3, tex(frag) append addstat(Comparison group obs. completely determined, e(N_cdf), Control group obs. completely det
table_3.tex
dir : seeout

. predict pscoreb, pr

.
. ********************************************************************************

. * problem #4 - compare pscorea and pscoreb descriptive statistics

. ********************************************************************************

.
. est clear

. estpost tabstat pscorea, by(in_control) c(stat) stat(mean, sd, min, median, max, count)

Summary statistics: mean sd min p50 max count
```

```
     for variables: pscorea
  by categories of: in_control

  in_control |   e(mean)      e(sd)     e(min)     e(p50)     e(max)   e(count)
-------------+------------------------------------------------------------------
           0 |  .0164119   .0654307   9.79e-13   .0001069   .6872082      15992
           1 |  .3873009    .230529   .0000756   .4716014   .6872082        425
-------------+------------------------------------------------------------------
       Total |  .0260134   .0949319   9.79e-13   .0001121   .6872082      16417

. esttab, cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))  count(fmt(%13.0fc))") nonumber

--------------------------------------------------------------------------------------------------
                            Mean           SD          Min       Median          Max            N
--------------------------------------------------------------------------------------------------
0                           0.02         0.07         0.00         0.00         0.69       15,992
1                           0.39         0.23         0.00         0.47         0.69          425
Total                       0.03         0.09         0.00         0.00         0.69       16,417
--------------------------------------------------------------------------------------------------

. esttab using "table_4a.tex", replace cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))  c
(output written to table_4a.tex)

.
. est clear

. estpost tabstat pscoreb, by(in_control) c(stat) stat(mean, sd, min, median, max, count)

Summary statistics: mean sd min p50 max count
     for variables: pscoreb
  by categories of: in_control

  in_control |   e(mean)      e(sd)     e(min)     e(p50)     e(max)   e(count)
-------------+------------------------------------------------------------------
           0 |  .0154465   .0642919   2.08e-16   .0000677   .7886285      15992
           1 |  .4248305   .2459295   1.64e-06   .4634333   .8024473        425
-------------+------------------------------------------------------------------
       Total |  .0260445   .0990717   2.08e-16   .0000899   .8024473      16417
```

```
. esttab, cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))  count(fmt(%13.0fc))") nonumber

------------------------------------------------------------------------------------------------
                           Mean           SD          Min       Median          Max            N
------------------------------------------------------------------------------------------------
0                          0.02         0.06         0.00         0.00         0.79       15,992
1                          0.42         0.25         0.00         0.46         0.80          425
Total                      0.03         0.10         0.00         0.00         0.80       16,417
------------------------------------------------------------------------------------------------

. esttab using "table_4b.tex", replace cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))  c
(output written to table_4b.tex)

.
. ********************************************************************************

. * problem #5 - histograms

. ********************************************************************************

.
. * create bins based on pscorea and pscoreb

. egen binsa=cut(pscorea), at(0(.05)1) icodes

. egen binsb=cut(pscoreb), at(0(.05)1) icodes

.
. * histogram of pscorea

. graph bar (count) pscorea, over(in_control) over(binsa, label(nolab)) asyvars title("Histogram of pscorea")

. graph export figure_5a.png, replace
file /Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/figure_5a.png saved as PNG format

.
. * histogram of pscoreb

. graph bar (count) pscoreb, over(in_control) over(binsb, label(nolab)) asyvars title("Histogram of pscoreb")
```

```
. graph export figure_5b.png, replace
file /Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/figure_5b.png saved as PNG format

.
. * histogram of pscorea wo lowest bin

. graph bar (count) pscorea if binsa > 0, over(in_control) over(binsa, label(nolab)) asyvars title("Histogram of pscorea wo lowest bin")

. graph export figure_5a_2.png, replace
file /Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/figure_5a_2.png saved as PNG format

.
. * histogram of pscoreb wo lowest bin

. graph bar (count) pscoreb if binsb > 0, over(in_control) over(binsb, label(nolab)) asyvars title("Histogram of pscoreb wo lowest bin")

. graph export figure_5b_2.png, replace
file /Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/figure_5b_2.png saved as PNG format

.
. ******************************************************************************

. * problem #6 - Nearest neighbor wo replacement

. ******************************************************************************

.
. global table_number = 6

.
. est clear

. eststo: psmatch2 in_control, noreplacement outcome(re78) pscore(pscorea) neighbor(1) common
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.    T-stat
---------------------------+--------------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT |  5090.0482   9529.11811  -4439.06991   486.482153    -9.12
```

```
---------------------------+--------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

           | psmatch2:
 psmatch2: |   Common
 Treatment |  support
assignment | On suppor |     Total
-----------+-----------+----------
 Untreated |    15,992 |    15,992
   Treated |       425 |       425
-----------+-----------+----------
     Total |    16,417 |    16,417
(est1 stored)

.
. global att_coarse = r(att)

. global att_coarse_se = r(seatt)

.
. esttab, se

----------------------------
                      (1)
                     re78
----------------------------
_treated          -9756.6***
                  (470.2)

_cons             14846.7***
                  (75.65)
----------------------------
N                   16417
----------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.
. global unm_coef = r(coefs)[1, 1]
```

```
. global unm_se = r(coefs)[1, 2]

.
. psmatch2 in_control, noreplacement outcome(re78) pscore(pscoreb) neighbor(1) common
------------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.    T-stat
-----------------------------+------------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   7408.17564  -2340.78116   449.369786    -5.21
-----------------------------+------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_fine = r(att)

. global att_fine_se = r(seatt)

.
. texdoc do table_maker

. global unm_coef = round($unm_coef, 0.01)

. global unm_se = round($unm_se, 0.01)

.
. global att_coarse = round($att_coarse, 0.01)

. global att_fine = round($att_fine, 0.01)
```

```
.
. global att_coarse_se = round($att_coarse_se, 0.01)

. global att_fine_se = round($att_fine_se, 0.01)

.
. global table_name = "table_$table_number.tex"

.
. local row1 = "Difference & $unm_coef & $att_coarse & $att_fine \\"

. local row2 = "SE & $unm_se & $att_coarse_se & $att_fine_se \\"

.
. texdoc init $table_name, replace
(texdoc output file is table_6.tex)

.
. texdoc append_snippet 2

.
. texdoc write ‘row1’

. texdoc write ‘row2’

.
. texdoc append_snippet 3

.
. texdoc close
(texdoc output written to table_6.tex)

.
end of do-file

.
. ********************************************************************************

. * problem #7 - nearest neighbor w replacement
```

```
. ********************************************************************************

.
. global table_number = 7

.
. est clear

. eststo: psmatch2 in_control, outcome(re78) pscore(pscorea) neighbor(1) common
------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
---------------------------+--------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT |  5090.0482   8767.08057  -3677.03237   934.497524    -3.93
---------------------------+--------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

           | psmatch2:
 psmatch2: |   Common
 Treatment |  support
assignment | On suppor |     Total
-----------+-----------+----------
 Untreated |    15,992 |    15,992
   Treated |       425 |       425
-----------+-----------+----------
     Total |    16,417 |    16,417
(est1 stored)

.
. global att_coarse = r(att)

. global att_coarse_se = r(seatt)

.
. esttab, se

----------------------------
                      (1)
```

```
                      re78
----------------------------
_treated           -9756.6***
                   (470.2)

_cons              14846.7***
                   (75.65)
----------------------------
N                    16417
----------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.
. global unm_coef = r(coefs)[1, 1]

. global unm_se = r(coefs)[1, 2]

.
. psmatch2 in_control, outcome(re78) pscore(pscoreb) neighbor(1) common
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
-----------------------------+------------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   6583.38855  -1515.99407   707.616814    -2.14
-----------------------------+------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_fine = r(att)
```

```
. global att_fine_se = r(seatt)

.
. texdoc do table_maker

. global unm_coef = round($unm_coef, 0.01)

. global unm_se = round($unm_se, 0.01)

.
. global att_coarse = round($att_coarse, 0.01)

. global att_fine = round($att_fine, 0.01)

.
. global att_coarse_se = round($att_coarse_se, 0.01)

. global att_fine_se = round($att_fine_se, 0.01)

.
. global table_name = "table_$table_number.tex"

.
. local row1 = "Difference & $unm_coef & $att_coarse & $att_fine \\"

. local row2 = "SE & $unm_se & $att_coarse_se & $att_fine_se \\"

.
. texdoc init $table_name, replace
(texdoc output file is table_7.tex)

.
. texdoc append_snippet 2

.
. texdoc write ‘row1’

. texdoc write ‘row2’
```

```
.
. texdoc append_snippet 3

.
. texdoc close
(texdoc output written to table_7.tex)

.
end of do-file

.
. ********************************************************************************

. * problem #8 - standardized differences

. ********************************************************************************

.
. * raw data

. stddiff re74 re75, by(in_control)

-------------------------------------------------------------------------------
             |      in_control=0       |      in_control=1       |
             | Mean or N     SD or (%) | Mean or N     SD or (%) |  Std Diff
-------------+-------------------------+-------------------------+------------
       re74 |     14017        9569.8 |      3672        6521.5 |    1.26323
       re75 |     13651        9270.4 |      3027        5201.2 |    1.41345
-------------------------------------------------------------------------------

.
. * rich pscore nearest neighbor - re74

. psmatch2 in_control, outcome(re74) pscore(pscoreb) neighbor(1) common
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.    T-stat
-----------------------------+------------------------------------------------------------
            re74  Unmatched | 3672.48515   14016.8007  -10344.3156   467.066411   -22.15
```

```
                        ATT |  3733.9861   4475.24661  -741.260515   638.936788    -1.16
----------------------------+-----------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. * rich pscore nearest neighbor - re74

. psmatch2 in_control, outcome(re75) pscore(pscoreb) neighbor(1) common
------------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
---------------------------+--------------------------------------------------------------
            re75  Unmatched | 3026.68274   13650.8032  -10624.1204   451.566731   -23.53
                        ATT | 3077.36881   3629.42503  -552.056218   562.743993    -0.98
---------------------------+--------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. ******************************************************************************

. * problem #9 - gaussian kernel matching
```

```
. ********************************************************************************

.
. global table_number = 9

.
. psmatch2 in_control, kernel outcome(re78) kerneltype(normal) pscore(pscoreb) bwidth(0.02) common
------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
---------------------------+--------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   7416.48625  -2349.09177   658.531426    -3.57
---------------------------+--------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_low = r(att)

. global att_low_se = r(seatt)

.
. psmatch2 in_control, kernel outcome(re78) kerneltype(normal) pscore(pscoreb) bwidth(0.2) common
------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
---------------------------+--------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   12110.8234  -7043.42897     337.0603   -20.90
---------------------------+--------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.
```

```
 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_med = r(att)

. global att_med_se = r(seatt)

.
. psmatch2 in_control, kernel outcome(re78) kerneltype(normal) pscore(pscoreb) bwidth(2.0) common
------------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
---------------------------+--------------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   14839.5273  -9772.13283   289.251216   -33.78
---------------------------+--------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_hi = r(att)

. global att_hi_se = r(seatt)

.
```

```
. texdoc do table_maker_2

.
. global att_low = round($att_low, 0.01)

. global att_med = round($att_med, 0.01)

. global att_hi = round($att_hi, 0.01)

.
. global att_low_se = round($att_low_se, 0.01)

. global att_med_se = round($att_med_se, 0.01)

. global att_hi_se = round($att_hi_se, 0.01)

.
. global table_name = "table_$table_number.tex"

.
. local row1 = "Difference & $att_low & $att_med & $att_hi \\"

. local row2 = "SE & $att_low_se & $att_med_se & $att_hi_se \\"

.
. texdoc init $table_name, replace
(texdoc output file is table_9.tex)

.
. texdoc append_snippet 2

.
. texdoc write ‘row1’

. texdoc write ‘row2’

.
. texdoc append_snippet 3
```

```
.
. texdoc close
(texdoc output written to table_9.tex)

.
end of do-file

.
. ******************************************************************************

. * problem #10 - local linear matching

. ******************************************************************************

.
. global table_number = 10

.
. psmatch2 in_control, llr outcome(re78) pscore(pscoreb) bwidth(0.02) common
------------------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.    T-stat
-----------------------------+------------------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   7048.27098   -1980.8765   707.616814    -2.80
-----------------------------+------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_low = r(att)
```

```
. global att_low_se = r(seatt)

.
. psmatch2 in_control, llr outcome(re78) pscore(pscoreb) bwidth(0.2) common
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
-----------------------------+------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   7488.81187  -2421.41739   707.616814    -3.42
-----------------------------+------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_med = r(att)

. global att_med_se = r(seatt)

.
. psmatch2 in_control, llr outcome(re78) pscore(pscoreb) bwidth(2.0) common
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference         S.E.   T-stat
-----------------------------+------------------------------------------------
            re78  Unmatched |  5090.0482   14846.6596  -9756.61142   470.155617   -20.75
                        ATT | 5067.39448   4073.51859   993.875888   707.616814     1.40
-----------------------------+------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
```

```
-----------+----------------------+----------
 Untreated |         0     15,992 |    15,992
   Treated |         7        418 |       425
-----------+----------------------+----------
     Total |         7     16,410 |    16,417

.
. global att_hi = r(att)

. global att_hi_se = r(seatt)

.
. texdoc do table_maker_2

.
. global att_low = round($att_low, 0.01)

. global att_med = round($att_med, 0.01)

. global att_hi = round($att_hi, 0.01)

.
. global att_low_se = round($att_low_se, 0.01)

. global att_med_se = round($att_med_se, 0.01)

. global att_hi_se = round($att_hi_se, 0.01)

.
. global table_name = "table_$table_number.tex"

.
. local row1 = "Difference & $att_low & $att_med & $att_hi \\"

. local row2 = "SE & $att_low_se & $att_med_se & $att_hi_se \\"

.
. texdoc init $table_name, replace
(texdoc output file is table_10.tex)
```

```
.
. texdoc append_snippet 2

.
. texdoc write ‘row1’

. texdoc write ‘row2’

.
. texdoc append_snippet 3

.
. texdoc close
(texdoc output written to table_10.tex)

.
end of do-file

.
. ******************************************************************************

. * problem #11 - linear regression of in_control

. ******************************************************************************

.
. regress re78 in_control age age_2 educ black hisp married nodegree re74 re75, robust

Linear regression                               Number of obs     =     16,417
                                                F(10, 16406)      =    1816.18
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4834
                                                Root MSE          =       6967

------------------------------------------------------------------------------
             |               Robust
        re78 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
```

```
  in_control |   -1853.39   343.2473    -5.40   0.000    -2526.192   -1180.588
         age |  -235.5257   40.07742    -5.88   0.000    -314.0818   -156.9696
       age_2 |   1.857809   .5503194     3.38   0.001     .7791228    2.936494
        educ |   163.3092   28.52916     5.72   0.000      107.389    219.2295
       black |  -832.3813     193.93    -4.29   0.000    -1212.505   -452.2574
        hisp |  -114.1536   213.9065    -0.53   0.594    -533.4335    305.1264
     married |   199.4396   150.7536     1.32   0.186    -96.05388    494.9331
   nodegree |   296.5755   174.3492     1.70   0.089    -45.16774    638.3188
        re74 |   .2912688   .0151515    19.22   0.000     .2615701    .3209675
        re75 |   .4710542   .0152652    30.86   0.000     .4411326    .5009757
       _cons |    7757.18   726.6778    10.67   0.000     6332.813    9181.548
------------------------------------------------------------------------------

. outreg2 using table_11a, tex(frag) replace
table_11a.tex
dir : seeout

.
. predict y_hat_11
(option xb assumed; fitted values)

.
. est clear

. estpost tabstat y_hat_11, by(in_control) c(stat) stat(mean, sd, min, median, max, count)

Summary statistics: mean sd min p50 max count
     for variables: y_hat_11
  by categories of: in_control

  in_control |   e(mean)      e(sd)     e(min)     e(p50)     e(max)   e(count)
-------------+------------------------------------------------------------------
           0 |  14846.66   6608.113   241.1631   15369.08   25559.22      15992
           1 |  5090.048   4227.391  -1476.131   3915.223   31078.45        425
-------------+------------------------------------------------------------------
       Total |  14594.08   6737.868  -1476.131   15111.14   31078.45      16417

. esttab, cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))  count(fmt(%13.0fc))") nonumber
```

```
------------------------------------------------------------------------------------------------
                            Mean           SD          Min       Median          Max            N
------------------------------------------------------------------------------------------------
0                     14,846.66     6,608.11       241.16    15,369.08    25,559.22       15,992
1                      5,090.05     4,227.39    -1,476.13     3,915.22    31,078.45          425
Total                 14,594.08     6,737.87    -1,476.13    15,111.14    31,078.45       16,417
------------------------------------------------------------------------------------------------

. esttab using "table_11b.tex", replace cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))
(output written to table_11b.tex)

.
. ********************************************************************************

. * problem #12 - linear regression of in_control out-of-sample

. ********************************************************************************

.
. regress re78 age age_2 educ black hisp married nodegree re74 re75 if in_control == 0, robust

Linear regression                               Number of obs     =     15,992
                                                F(9, 15982)       =    1896.13
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4758
                                                Root MSE          =       6987

------------------------------------------------------------------------------
             |               Robust
        re78 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |  -252.0485    40.9195    -6.16   0.000    -332.2553   -171.8417
       age_2 |   2.040805   .5606718     3.64   0.000     .9418257    3.139785
        educ |   166.5983    28.7942     5.79   0.000     110.1584    223.0382
       black |  -773.8793   199.5513    -3.88   0.000    -1165.022   -382.7364
        hisp |  -168.2284   218.9343    -0.77   0.442    -597.3642    260.9074
     married |   244.1442   153.1012     1.59   0.111     -55.9513    544.2398
    nodegree |   330.6767   176.8639     1.87   0.062    -15.99642    677.3497
        re74 |   .2988693    .015235    19.62   0.000     .2690069    .3287317
```

```
        re75 |   .4699589   .0153351    30.65   0.000     .4399004    .5000173
       _cons |   7908.362   739.6679    10.69   0.000      6458.53    9358.194
------------------------------------------------------------------------------

. outreg2 using table_12a, tex(frag) replace
table_12a.tex
dir : seeout

.
. predict y_hat_12
(option xb assumed; fitted values)

.
. est clear

. estpost tabstat y_hat_12, by(in_control) c(stat) stat(mean, sd, min, median, max, count)

Summary statistics: mean sd min p50 max count
     for variables: y_hat_12
  by categories of: in_control

  in_control |   e(mean)      e(sd)     e(min)     e(p50)     e(max)   e(count)
-------------+------------------------------------------------------------------
           0 |  14846.66    6654.46   162.9647   15370.95   25693.43      15992
           1 |  6941.597   4269.447   275.7228   5770.588   33124.62        425
-------------+------------------------------------------------------------------
       Total |  14642.01   6721.768   162.9647   15110.59   33124.62      16417

. esttab, cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))  count(fmt(%13.0fc))") nonumber

------------------------------------------------------------------------------------------------
                            Mean           SD          Min       Median          Max            N
------------------------------------------------------------------------------------------------
0                       14,846.66     6,654.46       162.96    15,370.95    25,693.43       15,992
1                        6,941.60     4,269.45       275.72     5,770.59    33,124.62          425
Total                   14,642.01     6,721.77       162.96    15,110.59    33,124.62       16,417
------------------------------------------------------------------------------------------------

. esttab using "table_12b.tex", replace cells("mean(fmt(%13.2fc)) sd(fmt(%13.2fc)) min(fmt(%13.2fc)) p50(fmt(%13.2fc)) max(fmt(%13.2fc))
```

```
(output written to table_12b.tex)

.
. ******************************************************************************

. * problem #13 - inverse probability weighting

. ******************************************************************************

.
. * get number of treated

. count if in_control == 1
  425

. scalar n_1 = r(N)

.
. * get number of untreated

. count if in_control == 0
  15,992

. scalar n_0 = r(N)

.
. * get unconditional probability of treatment

. scalar p_hat = n_1/(n_0 + n_1)

.
. * get y * d

. gen y_d = re78 * in_control

. summarize y_d

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
```

```
         y_d |     16,417    131.7701    1223.885          0   39483.53

. scalar y_d_sum = r(sum)

.
. * get second term wo rescaling

. gen wo_rescale = (1 - p_hat) / p_hat * pscoreb * re78 * (1 - in_control)/(1 - pscoreb)

. summarize wo_rescale

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  wo_rescale |     16,417    6341.886    45200.03          0    1596524

. scalar wo_rescale_sum = r(sum)

.
. * get second term w rescaling

. gen rescaling = pscoreb * (1-in_control)/(1-pscoreb)

. summarize rescaling

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   rescaling |     16,417    .0237833    .1427705          0   3.731006

. scalar rescaling = 1/n_0 * r(sum)

. gen w_rescale = (1/rescaling) * pscoreb * re78 * (1 - in_control)/(1-pscoreb)
(425 missing values generated)

. summarize w_rescale

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   w_rescale |     15,992    14846.66    9647.391          0   25564.67
```

```
. scalar w_rescale_sum = r(sum)

.
. * treatment effect on treated ipw w and wo rescaling

. display 1/n_1 * y_d_sum - 1/n_0 * wo_rescale_sum
-1420.3787

. display 1/n_1 * y_d_sum - 1/n_0 * w_rescale_sum
-9756.6114

.
.
. log close
      name:  <unnamed>
       log:  /Users/vonhafften/Documents/UW Madison/problem_sets/econ_717a/ps2/analysis.smcl
  log type:  smcl
 closed on:   6 Mar 2022, 16:03:56
-------------------------------------------------------------------------------------------------------------------------------
```