Practical Exercises Discussion
Propensity Score

2. CRUDE EFFECT

Perform a t-test to estimate the mean LDL among drinkers and non-drinkers

```
ttest LDL, by(drinkany)
```

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      no |    1675    146.3786    .9487574    38.82957    144.5178    148.2395
     yes |    1075    142.9931    1.100692     36.0886    140.8334    145.1529
---------+--------------------------------------------------------------------
combined |    2750    145.0552    .7210321    37.81124    143.6414     146.469
---------+--------------------------------------------------------------------
    diff |            3.385511    1.476521                .4903076    6.280714
------------------------------------------------------------------------------
    diff = mean(no) - mean(yes)                                   t =   2.2929
Ho: diff = 0                                     degrees of freedom =     2748

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9890         Pr(|T| > |t|) = 0.0219          Pr(T > t) = 0.0110
```

The mean is 3.4 mg/dL lower among drinkers than among non-drinkers (+0.5 to -6.3), p = 0.02. This is completely unadjusted so it could be susceptible to differences due to confounders. This is what we attempt to address using propensity scores.

4. OVERLAP

Overlap, requires that we have a good number of drinkers and non-drinkers at the levels of the propensity.

What does the table for overlap indicate?

```
tabulate drinkany group, col
```

any current alcohol consumption	5 quantiles of out					
n	1	2	3	4	5	Total
no	421	337	320	323	275	1,676
	76.41	61.05	58.08	58.51	50.00	60.81
yes	130	215	231	229	275	1,080
	23.59	38.95	41.92	41.49	50.00	39.19
Total	551	552	551	552	550	2,756
	100.00	100.00	100.00	100.00	100.00	100.00

This is very good overlap with an adequate sample size in each group. We note that the group with the most imbalance is the first quintile with 130 v. 421 which means even in the first quantile there is a good sample size in the smaller group.

5. DRINKING BY PROPENSITY INTERACTION

Perform the likelihood ratio test for interaction.

.lrtest A

```
Likelihood-ratio test                                 LR chi2(4)  =      4.45
(Assumption: . nested in A)                           Prob > chi2 =    0.3484
```

What is the null hypothesis of this test? Is the null hypothesis rejected? Does there appear to be an interaction?

The null hypothesis is that the adjusted mean difference in LDL between drinkers and non-drinkers is the same across the quintile of the propensity score. The null hypothesis is not rejected which tells us there is not convincing evidence that the effect of drinking across quintiles of the propensity scores.

This can be further seen by examining the mean LDL by drinking and propensity quintile and we do see a range of mean effects of LDL of -7.7 in the first quartile to +0.7 in the second quartile. It seems that a degree of subject matter knowledge would be required to have a firm opinion about whether this is a clinical meaningful degree of variation.

```
table drinkany group, c(mean LDL)

------------------------------------------------------------
any       |
current   |
alcohol   |
consumpti |                5 quantiles of out
on        |        1          2          3          4          5
----------+-------------------------------------------------
       no | 151.0268   143.6577   145.5313    144.705   145.7033
      yes | 143.3246   144.3502    142.927   138.6176    145.492
------------------------------------------------------------
```

6. QUINTILE ANALYSIS

What is the estimated average causal effect of drinking on LDL using the quintile?

. reg LDL i. drinkany i.group

```
      Source |       SS       df       MS              Number of obs =    2745
-------------+------------------------------           F(  5,  2739) =    2.87
       Model |  20460.8981     5  4092.17962           Prob > F      =  0.0137
    Residual |  3903721.59  2739  1425.23607           R-squared     =  0.0052
-------------+------------------------------           Adj R-squared =  0.0034
       Total |  3924182.49  2744  1430.09566           Root MSE      =  37.752

------------------------------------------------------------------------------
         LDL |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
        yes  |  -2.937299   1.499656    -1.96   0.050    -5.877869    .0032715
             |
       group |
          2  |   -4.83042   2.290844    -2.11   0.035    -9.322376   -.3384641
          3  |  -4.225933   2.294728    -1.84   0.066    -8.725505    .2736395
          4  |  -6.493949   2.294933    -2.83   0.005    -10.99392   -1.993974
          5  |  -2.832651   2.313238    -1.22   0.221    -7.368519    1.703217
             |
       _cons |   149.8964   1.651472    90.77   0.000     146.6582    153.1347
------------------------------------------------------------------------------
```

The mean difference between drinkers and non-drinkers is -2.9 with a 95% confidence interval from -5.9 to +0.0, p = 0.05

Use the margins command to estimate the causal mean LDL if the entire population was drinkers v. non-drinkers.

```
margins drinkany
```

```
Predictive margins                                 Number of obs   =       2745
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
          no |   146.2183   .9292405   157.35   0.000     144.3962    148.0404
         yes |    143.281   1.163017   123.20   0.000     141.0005    145.5615
------------------------------------------------------------------------------
```

It is important to note (1)what quantities these numbers are meant to represent, (2) how they are calculated and (3)why they are unconfounded.

The mean value (143.281) which appears under drinking is an estimate of the mean LDL which would be predicted based on the regression model if everyone in the dataset was a drinker (but keep their confounder values and hence their propensity the same).

How would we calculate such a quantity? For people who were drinkers, we can use their values. For someone who is a non-drinker, we need to predict what their LDL would be if their confounders remained the same but they were a drinker. Confounders only exert their influence here through the propensity score - specifically, the propensity quintile. So, if a non-drinker was , say, in the first quintile of the propensity their estimated LDL if they drank would be 149.8964 (the first quintile is the reference group) -2.937299. This would be repeated for each non-drinker. The mean over all values in the data (actual values in the drinker and predicted values in non-drinkers) are what gives rise to the mean value for the drinkers. The mean for non-drinkers is defined analogously.

This strategy deals with confounding. Because the means are calculated over two populations with exactly the same distributions of potential confounders, confounding by these variables is eradicated. Here, the two identical populations are the full dataset (really one population compared under two conditions -- drinking and non-drinking).

Verify that the difference of these means corresponds to the coefficient in the above regression model.

The difference in means from the previous "margins" command is given by

```
margins, dydx(drinkany)
```

```
Average marginal effects                        Number of obs   =       2745
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 1.drinkany

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
        yes  |  -2.937299   1.499656    -1.96   0.050    -5.877869    .0032715
------------------------------------------------------------------------------
```

The mean difference between drinkers and non-drinkers is -2.9 with a 95% confidence interval from -5.9 to +0.0, p = 0.05. Note, this is exactly the difference of means of 146.2183 - 143.281 and exactly equal to the coefficient from the regression model on which the margins are based. There is something useful here. Usually, we can only talk about the adjusted mean difference (-2.94) and people have trouble judging this because they can see the means in the two groups. The means from the t-test are potentially confounded) However, margins gives us adjusted means and the difference in the adjusted means is identical to the adjusted difference between the groups.

6. 20-LEVEL ANALYSIS

Create a 20 level spilt of the propensities to rule out the possibility of residual confounding in the quintile analysis

```
xtile group20 = out, nq(20)

reg LDL i.drinkany i.group20
```

```
. reg LDL i.drinkany i.group20

      Source |       SS       df       MS              Number of obs =    2745
-------------+------------------------------           F( 20,  2724) =    2.05
       Model |  58237.2186    20  2911.86093           Prob > F      =  0.0039
    Residual |  3865945.27  2724  1419.21632           R-squared     =  0.0148
-------------+------------------------------           Adj R-squared =  0.0076
       Total |  3924182.49  2744  1430.09566           Root MSE      =  37.672
```

```
------------------------------------------------------------------------------
         LDL |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
        yes  |  -2.914729   1.502084    -1.94   0.052    -5.860069      .03061
             |
     group20 |
           2 |   5.645777   4.560364     1.24   0.216    -3.296345     14.5879
           3 |  -8.859952   4.565822    -1.94   0.052    -17.81278    .0928734
           4 |   1.091636   4.574203     0.24   0.811    -7.877623     10.0609
           5 |  -5.205367     4.5865    -1.13   0.257    -14.19874    3.788004
           6 |    -9.4825   4.569481    -2.08   0.038     -18.4425   -.5225004
           7 |  -2.816024   4.573241    -0.62   0.538     -11.7834    6.151347
           8 |  -3.964823   4.588219    -0.86   0.388    -12.96156    5.031919
           9 |  -7.204453   4.586245    -1.57   0.116    -16.19732    1.788417
          10 |  -7.339405   4.580946    -1.60   0.109    -16.32189    1.643074
          11 |  -1.108969   4.577644    -0.24   0.809    -10.08498    7.867037
          12 |   -3.44955   4.577644    -0.75   0.451    -12.42556    5.526456
          13 |  -9.140638   4.592145    -1.99   0.047    -18.14508   -.1361983
          14 |  -7.401126   4.583478    -1.61   0.106    -16.38857    1.586319
          15 |  -9.825125   4.575806    -2.15   0.032    -18.79753   -.8527237
          16 |  -1.804003   4.576712    -0.39   0.693    -10.77818    7.170176
          17 |   -3.92268    4.59361    -0.85   0.393    -12.92999    5.084633
          18 |   2.807445   4.598797     0.61   0.542    -6.210039    11.82493
          19 |  -3.364178   4.594766    -0.73   0.464    -12.37376    5.645401
          20 |   -9.06596   4.591376    -1.97   0.048    -18.06889   -.0630281
             |
       _cons |   150.4276   3.251559    46.26   0.000     144.0518    156.8034
------------------------------------------------------------------------------

margins drinkany

Predictive margins                                Number of obs   =       2745
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
          no |   146.2095   .9286607   157.44   0.000     144.3885    148.0304
         yes |   143.2947   1.163235   123.19   0.000     141.0138    145.5757
------------------------------------------------------------------------------
```

.How do these results compare to the quintile analysis

They are strikingly similar.

7. STANDARD ADJUSTED ANALYSIS

In a standard adjusted analysis we would do a linear regression on LDL with drinking adjusting for the potential confounders.

```
reg LDL i.drinkany smoking nonwhite age BMI
```

```
      Source |       SS       df       MS              Number of obs =    2745
-------------+------------------------------           F(  5,  2739) =    5.97
       Model |  42279.1877     5  8455.83753           Prob > F      =  0.0000
    Residual |   3881903.3  2739  1417.27028           R-squared     =  0.0108
-------------+------------------------------           Adj R-squared =  0.0090
       Total |  3924182.49  2744  1430.09566           Root MSE      =  37.647

------------------------------------------------------------------------------
         LDL |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
        yes  |  -2.722354   1.498854    -1.82   0.069    -5.661352    .2166445
     smoking |   4.750738   2.210391     2.15   0.032     .4165362     9.08494
    nonwhite |   5.219436   2.323673     2.25   0.025     .6631079    9.775764
         age |  -.1897166   .1130776    -1.68   0.094    -.4114427    .0320095
         BMI |   .3591038   .1341047     2.68   0.007     .0961472    .6220605
       _cons |   147.3153   9.256449    15.91   0.000      129.165    165.4656
------------------------------------------------------------------------------
```

What is the estimated effect of drinking on LDL in the standard regression?

It is a difference of -2.7 mg/dL with a 95% confidence interval (-5.6 mg/dL to +0.2 mg/dL) with p-value 0.069.

What are the estimated mean LDL for drinkers and non-drinkers? How to these results compare to the propensity analysis?

```
margins drinkany
```

```
Predictive margins                                 Number of obs   =       2745
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
          no |   146.1342   .9274793   157.56   0.000     144.3156    147.9528
         yes |   143.4119   1.161385   123.48   0.000     141.1346    145.6891
------------------------------------------------------------------------------
```

They are very similar even through the p-values are significant in one case and not the other.

8. HIGH DIMENSIONAL PROPENSITY MODEL

Fit a very rich propensity model which includes splines and interactions.

```
mkspline agesp = age, nknots(5) cubic
mkspline bmisp = BMI, nknots(5) cubic

logit drinkany smoking nonwhite smoking##c.agesp*  smoking##c.b-
misp* nonwhite##c.agesp* nonwhite##c.bmisp*
```
(the above needs to be written on one line)

```
predict outsp, p

xtile groupsp = outsp, nq(5)
```

Assess the overlap of the quintile groups “groupsp” based on the richer model and estimate the average causal effect based on it. Have your conclusions changed?

```
     any |
   current |
   alcohol |
consumptio |                5 quantiles of outsp
         n |         1          2          3          4          5 |     Total
-----------+-------------------------------------------------------+----------
        no |       419        363        318        302        274 |     1,676
           |     76.04      65.76      57.71      54.71      49.82 |     60.81
-----------+-------------------------------------------------------+----------
       yes |       132        189        233        250        276 |     1,080
           |     23.96      34.24      42.29      45.29      50.18 |     39.19
-----------+-------------------------------------------------------+----------
     Total |       551        552        551        552        550 |     2,756
           |    100.00     100.00     100.00     100.00     100.00 |    100.00
```

Note, there was an error in the lab. The command for the quintile should have been “xtile groupsp = outsp, nq(5)”. Note, the correction in red. Overall, the overlap as well the as major result is quite similar to the less rich propensity model

```
 reg LDL i.drinkany i.groupsp

      Source |       SS       df       MS              Number of obs =    2745
-------------+------------------------------           F(  5,  2739) =    3.31
       Model |  23582.7607     5  4716.55214           Prob > F      =  0.0055
    Residual |  3900599.73  2739  1424.09629           R-squared     =  0.0060
-------------+------------------------------           Adj R-squared =  0.0042
       Total |  3924182.49  2744  1430.09566           Root MSE      =  37.737

------------------------------------------------------------------------------
         LDL |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
        yes  |  -2.962454   1.502953    -1.97   0.049     -5.90949   -.0154188
             |
```

```
     groupsp |
           2 |  -5.716826   2.282945    -2.50   0.012    -10.19329   -1.240358
           3 |   -6.96432   2.295057    -3.03   0.002    -11.46454   -2.464103
           4 |  -4.611981   2.299049    -2.01   0.045    -9.120026   -.1039359
           5 |  -2.834642   2.315005    -1.22   0.221    -7.373975     1.70469
             |
       _cons |   150.2571   1.651605    90.98   0.000     147.0186    153.4956
------------------------------------------------------------------------------


 margins drinkany

Predictive margins                                Number of obs   =       2745
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    drinkany |
          no |   146.2282   .9298323   157.26   0.000     144.4049    148.0514
         yes |   143.2657   1.164414   123.04   0.000     140.9825    145.5489
------------------------------------------------------------------------------
```