Problem Set 4

Question #1: LPM

1.1

```
reg train unem74 unem75 age educ black hisp married

      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  7,   437) =    1.43
       Model |  2.41922955     7  .345604222           Prob > F      =  0.1915
    Residual |  105.670658   437  .241809286           R-squared     =  0.0224
-------------+------------------------------           Adj R-squared =  0.0067
       Total |  108.089888   444  .243445693           Root MSE      =  .49174

------------------------------------------------------------------------------
       train |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      unem74 |     .02088   .0772939     0.27   0.787    -.1310341     .172794
      unem75 |  -.0955711   .0719021    -1.33   0.184     -.236888    .0457459
         age |   .0032057   .0034027     0.94   0.347     -.003482    .0098933
        educ |   .0120131   .0133419     0.90   0.368    -.0142092    .0382354
       black |  -.0816663   .0877325    -0.93   0.352    -.2540963    .0907637
        hisp |  -.2000168   .1169708    -1.71   0.088    -.4299122    .0298785
     married |   .0372887   .0644037     0.58   0.563    -.0892909    .1638683
       _cons |   .3380222   .1894451     1.78   0.075    -.0343147    .7103591
------------------------------------------------------------------------------

testparm unem74 unem75 age educ black hisp married

 ( 1)  unem74 = 0
 ( 2)  unem75 = 0
 ( 3)  age = 0
 ( 4)  educ = 0
 ( 5)  black = 0
 ( 6)  hisp = 0
 ( 7)  married = 0

       F(  7,   437) =    1.43
            Prob > F =    0.1915
```

Since we cannot reject H_0 that All $\beta_j = 0$, these explanatory variables are not statistically jointly significant at 5% level.

Calculate the percent of correctly predicted:

```
First, generate dummy variable refer to predicted binary result:
. gen Dytrain = 1 if ytrain >=.5
. replace Dytrain = 0 if ytrain <.5

. tab train Dytrain, cell

    =1 if |
 assigned |
   to job |       Dytrain
 training |         0          1 |     Total
-----------+----------------------+----------
         0 |       237         23 |       260
           |     53.26       5.17 |     58.43
-----------+----------------------+----------
         1 |       158         27 |       185
           |     35.51       6.07 |     41.57
-----------+----------------------+----------
     Total |       395         50 |       445
           |     88.76      11.24 |    100.00
```

```
. display 53.26+6.07
59.33
```

Percent correctly predicted is 59.33%.

1.2 Probit estimation

```
. probit train unem74 unem75 age educ black hisp married

Iteration 0:   log likelihood =     -302.1
Iteration 1:   log likelihood = -297.01499
Iteration 2:   log likelihood =  -297.0088
Iteration 3:   log likelihood =  -297.0088

Probit regression                                 Number of obs   =        445
                                                  LR chi2(7)      =      10.18
                                                  Prob > chi2     =     0.1785
Log likelihood =  -297.0088                       Pseudo R2       =     0.0169

------------------------------------------------------------------------------
       train |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      unem74 |   .0530256   .1992686     0.27   0.790    -.3375337    .4435849
      unem75 |  -.2477249     .18505    -1.34   0.181    -.6104163    .1149665
         age |   .0083443   .0087982     0.95   0.343    -.0088999    .0255886
        educ |   .0314431   .0343238     0.92   0.360    -.0358304    .0987165
       black |  -.2069299   .2249003    -0.92   0.358    -.6477264    .2338666
        hisp |  -.5397772   .3085029    -1.75   0.080    -1.144432    .0648773
     married |   .0966251   .1655823     0.58   0.560    -.2279101    .4211604
       _cons |  -.4241079   .4870267    -0.87   0.384    -1.378663    .5304469
------------------------------------------------------------------------------

. est sto ptrain

. probit train

Iteration 0:   log likelihood =     -302.1
Iteration 1:   log likelihood =     -302.1

Probit regression                                 Number of obs   =        445
                                                  LR chi2(0)      =      -0.00
                                                  Prob > chi2     =          .
Log likelihood =     -302.1                       Pseudo R2       =    -0.0000

------------------------------------------------------------------------------
       train |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.2128286   .0599043    -3.55   0.000    -.3302389   -.0954182
------------------------------------------------------------------------------

. est sto p0

. lrtest ptrain p0

Likelihood-ratio test                                 LR chi2(7)  =     10.18
(Assumption: p0 nested in ptrain)                     Prob > chi2 =    0.1785
```

Or by hand: LR = 2(lur – lr) = 2*(-297.01-(-302.1)) = 10.18

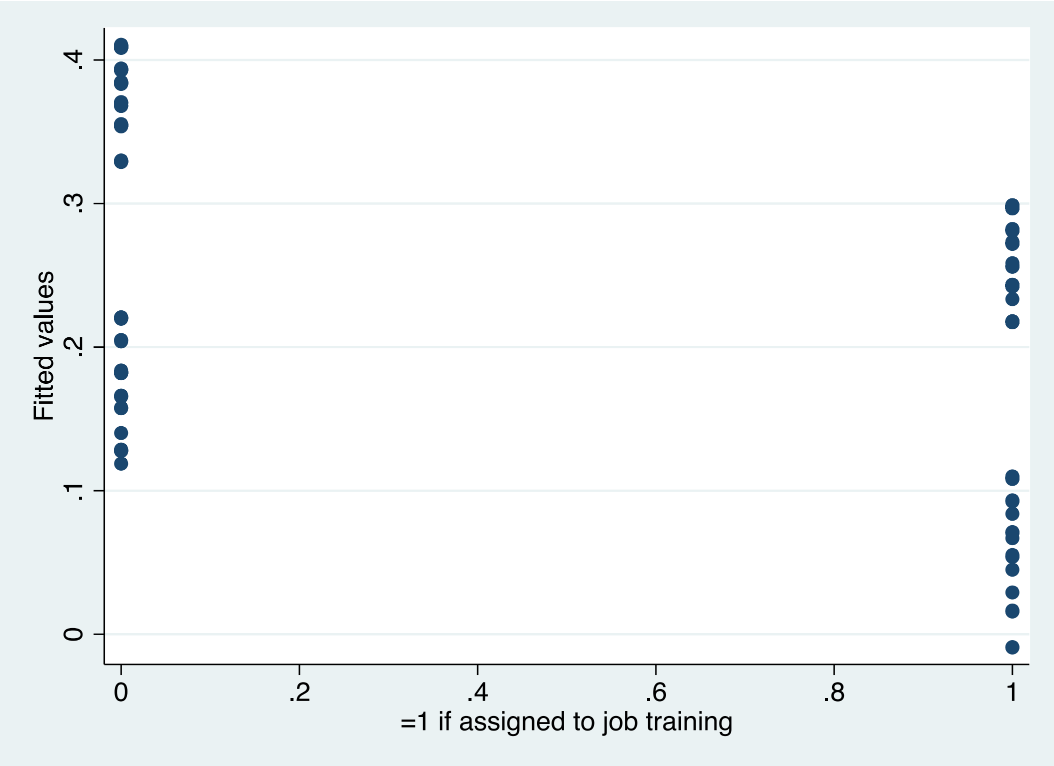
We fail to reject H_0, therefore all variables are not statistically jointly significant.

1.3 Participation in job training can be treated as exogenous for explaining 1978 unemployment status since it is not explained by other observed control variables (unem74,75, age, educ, black, hisp, married) due to its lack of statistical significance. Given these control variables, we can use participation in job training as an exogenous variable together with these control variables in explaining unem78.

1.4 LPM of unem78

```
. reg unem78 train unem74 unem75 age educ black hisp married

      Source |       SS       df       MS              Number of obs =     445
-------------+------------------------------           F(  8,   436) =    2.64
       Model |  4.38106947     8  .547633683           Prob > F      =  0.0078
    Residual |  90.4414024   436  .207434409           R-squared     =  0.0462
-------------+------------------------------           Adj R-squared =  0.0287
       Total |  94.8224719   444  .213564126           Root MSE      =  .45545

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |  -.1117028   .0443061    -2.52   0.012    -.1987828   -.0246228
      unem74 |   .0386926   .0715954     0.54   0.589    -.1020226    .1794077
      unem75 |   .0159613   .0667301     0.24   0.811    -.1151913    .1471139
         age |   .0000433   .0031548     0.01   0.989    -.0061571    .0062438
        educ |   .0001442   .0123687     0.01   0.991    -.0241655     .024454
       black |   .1888328   .0813382     2.32   0.021     .0289692    .3486964
        hisp |  -.0377011      .1087    -0.35   0.729    -.2513423    .1759401
     married |  -.0254373   .0596735    -0.43   0.670    -.1427208    .0918462
       _cons |   .1631823   .1761017     0.93   0.355    -.1829315    .5092961
------------------------------------------------------------------------------

. test train

 ( 1)  train = 0

       F(  1,   436) =    6.36
            Prob > F =    0.0121
```

The coefficient of train is -0.1117, meaning that participating in job training will reduce the probability of being unemployed in 1978 by 0.1117, or 11.17 percentage points, compared to those who did not participate in the job traning, holding other variables constant. This estimate is statistically significant at 5% level. See the graph below that a few predicted values are outside [0,1].



1.5 Probit estimation of unem78

It does not make sense to compare the probit coefficient on train with the coefficient obtained from the linear model in (1.4) although we can compare the sign of the coefficient. We can compare the results when we estimate average marginal effects since now we differentiate on predicted probability. From the result below, the average marginal effects of train on unem78 equal to -0.1132, meaning that participating in the job training reduce the probability of being unemployed in 1978 by 11.32 percentage points, ceteris paribus. This is very similar to the result from the linear model.

```
probit unem78 train unem74 unem75 age educ black hisp married

Iteration 0:   log likelihood = -274.73494
Iteration 1:   log likelihood =  -263.3816
Iteration 2:   log likelihood =  -263.3128
Iteration 3:   log likelihood = -263.31279

Probit regression                                 Number of obs   =        445
                                                  LR chi2(8)      =      22.84
                                                  Prob > chi2     =     0.0036
Log likelihood = -263.31279                       Pseudo R2       =     0.0416

------------------------------------------------------------------------------
      unem78 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |  -.3365897   .1316429    -2.56   0.011    -.5946051   -.0785744
      unem74 |    .106094   .2125598     0.50   0.618    -.3105155    .5227035
      unem75 |   .0636124   .1970995     0.32   0.747    -.3226956    .4499204
         age |   .0006757   .0091211     0.07   0.941    -.0172014    .0185529
        educ |  -.0018916   .0367938    -0.05   0.959    -.0740061    .0702229
       black |   .6336688   .2742692     2.31   0.021      .096111    1.171227
        hisp |  -.1649409   .3790471    -0.44   0.663    -.9078596    .5779777
     married |   -.077768   .1771557    -0.44   0.661    -.4249869    .2694509
       _cons |  -1.010331   .5380256    -1.88   0.060    -2.064842    .0441798
------------------------------------------------------------------------------
```

```
margins, dydx(train)

Average marginal effects                        Number of obs   =        445
Model VCE    : OIM

Expression   : Pr(unem78), predict()
dy/dx w.r.t. : train

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       train |  -.1131733   .0433255    -2.61   0.009    -.1980896   -.0282569
------------------------------------------------------------------------------
```

Question #2: Tobit model

2.1 There are about 27.92 percent of workers that have pension equal to zero. The range of pension for workers with nonzero pension benefits is between $7.28 to $2,880.27. The tobit model is appropriate for modeling pension since we have a nontrivial fraction of the dependent variable with zero value, which is considered as corner solution responses. If we use OLS, we could possibly obtain negative fitted values. Moreover, when the distribution of the dependent variable piles up at zero, we cannot have the dependent variable conditional normal distributed as in a linear model.

```
. gen Dpension = (pension>0 & pension<.)

. tab Dpension

   Dpension |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        172       27.92       27.92
          1 |        444       72.08      100.00
------------+-----------------------------------
      Total |        616      100.00


. sum pension if Dpension == 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     pension |        444    905.0439    550.3696       7.28    2880.27
```

2.2 Tobit model
For individual test, male have statistically significant $308.15 higher expected pension benefits compare to female at 1% significance level. On the other hand, whites does not have statistically significant higher expected pension benefits compare to nonwhites.

For joint test, males and whites have statistically significant higher expected pension benefits than females and nonwhites.

```
. tobit pension exper age tenure educ depends married i.white i.male, ll(0)

Tobit regression                                Number of obs   =        616
                                                LR chi2(8)      =     184.70
                                                Prob > chi2     =     0.0000
Log likelihood = -3672.9635                     Pseudo R2       =     0.0245
```

```
------------------------------------------------------------------------------
     pension |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   5.203458    6.00928     0.87   0.387    -6.598007    17.00492
         age |  -4.638944   5.710741    -0.81   0.417    -15.85412    6.576228
      tenure |   36.02385   4.564348     7.89   0.000     27.06005    44.98765
        educ |   93.21262   10.89133     8.56   0.000     71.82343    114.6018
     depends |   35.28461   21.91691     1.61   0.108    -7.757432    78.32666
     married |   53.68858   71.73266     0.75   0.454    -87.18528    194.5624
       white |   144.0855   102.0753     1.41   0.159    -56.37738    344.5485
        male |   308.1505    69.8903     4.41   0.000     170.8948    445.4062
       _cons |  -1252.429   219.0692    -5.72   0.000    -1682.653   -822.2048
-------------+----------------------------------------------------------------
      /sigma |   677.7383   24.13815                      630.3341    725.1426
------------------------------------------------------------------------------
  Obs. summary:        172  left-censored observations at pension<=0
                       444     uncensored observations
                         0 right-censored observations

. test white male

 ( 1)  [model]white = 0
 ( 2)  [model]male = 0

       F(  2,   608) =   11.17
            Prob > F =    0.0000
```

2.3 The difference in expected pension benefits for a white male and a nonwhite female, both of whom are 35 years old, are single with no dependents, have 16 years of education, and have 10 years of experience is calculated below:

```
(suppose tenure=10)
margins male, at(age=35 married=0 depends=0 tenure=10 educ=16 exper=10 white=1)
predict(ystar(0,.))

Adjusted predictions                              Number of obs   =        616
Model VCE    : OIM

Expression   : E(pension*|pension>0), predict(ystar(0,.))
at           : exper           =          10
               age             =          35
               tenure          =          10
               educ            =          16
               depends         =           0
               married         =           0
               white           =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |
          0  |   696.9033    63.7512    10.93   0.000     571.9533    821.8534
          1  |   966.6001   84.67276    11.42   0.000     800.6445    1132.556
------------------------------------------------------------------------------

. margins male, at(age=35 married=0 depends=0 tenure=10 educ=16 exper=10 white=0)
predict(ystar(0,.))

Adjusted predictions                              Number of obs   =        616
Model VCE    : OIM

Expression   : E(pension*|pension>0), predict(ystar(0,.))
```

```
at           : exper          =          10
               age            =          35
               tenure         =          10
               educ           =          16
               depends        =           0
               married        =           0
               white          =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |
          0  |   582.2655   93.03065     6.26   0.000     399.9288    764.6023
          1  |   836.9647   118.9587     7.04   0.000       603.81    1070.119
------------------------------------------------------------------------------

display 966.60-582.27
384.33
```

Given other variables as stated above, a white male has $384.33 more pension benefits than a nonwhite female.

2.4 Add union to the tobit model

Since we reject H_0 that the coefficient of union is equal to zero at 1% significance level, being in union will increase *latent* pension benefits by $439.046, comparing to not being in the union, holding other variables constant.

```
. tobit pension exper age tenure educ depends married white male union, ll(0)

Tobit regression                                Number of obs     =        616
                                                LR chi2(9)        =     233.52
                                                Prob > chi2       =     0.0000
Log likelihood = -3648.5515                     Pseudo R2         =     0.0310

------------------------------------------------------------------------------
     pension |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   4.393523   5.830946     0.75   0.451    -7.057754     15.8448
         age |  -1.653532   5.555708    -0.30   0.766    -12.56427    9.257211
      tenure |   28.77837   4.504963     6.39   0.000     19.93116    37.62557
        educ |   106.8277   10.77274     9.92   0.000     85.67134    127.9841
     depends |   41.46623   21.21414     1.95   0.051    -.1957922    83.12824
     married |   19.74554   69.50047     0.28   0.776     -116.745    156.2361
       white |   159.2972   98.96747     1.61   0.108    -35.06298    353.6575
        male |   257.2457   68.02051     3.78   0.000     123.6615    390.8298
       union |    439.046   62.48832     7.03   0.000     316.3265    561.7656
       _cons |  -1571.506   218.5445    -7.19   0.000    -2000.701   -1142.311
-------------+----------------------------------------------------------------
      /sigma |   652.8974   23.16287                      607.4083    698.3865
------------------------------------------------------------------------------
  Obs. summary:        172  left-censored observations at pension<=0
                       444     uncensored observations
                         0 right-censored observations
```

If adding the adjustment factor, we have average marginal effects = 325.11

```
margins, dydx(union) predict(ystar(0,.))

Average marginal effects                        Number of obs     =        616
Model VCE    : OIM
```

```
Expression   : E(pension*|pension>0), predict(ystar(0,.))
dy/dx w.r.t. : union

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       union |   325.1068   45.20549     7.19   0.000     236.5057     413.708
------------------------------------------------------------------------------
```

3.1 MA(1): $y_t = u_t + \alpha u_{t-1}$

$$\begin{aligned} Cov(y_t, y_{t-1}) &= E[y_t, y_{t-1}] = E[(u_t + \alpha u_{t-1})(u_{t-1} + \alpha u_{t-2})] \\ &= E[u_t u_{t-1} + \alpha u_{t-1}^2 + \alpha u_t u_{t-2} + \alpha^2 u_{t-1} u_{t-2}] \\ &= E[u_t u_{t-1}] + \alpha\sigma_u^2 + \alpha E[u_t u_{t-2}] + \alpha^2 E[u_{t-1} u_{t-2}], \end{aligned}$$

with no serial correlation: $Cov(u_t, u_{t+s}) = 0$, for t ≠ s

$\therefore Cov(y_t, y_{t-1}) = \alpha\sigma_u^2$

3.2 $E(y_t) = E(z + e_t) = E(z) + E(e_t) = 0$; we have constant mean.

$Var(y_t) = Var(z + e_t) = Var(z) + Var(e_t) + 2Cov(z, e_t)$, we have e_t is uncorrelated with z. So, $Var(y_t) = \sigma_z^2 + \sigma_u^2$; variance does not depend on time.

$$\begin{aligned} Cov(y_t, y_{t+s}) &= E[(z + e_t)(z + e_{t+1})] \\ &= E[z^2] + E[ze_t] + E[ze_{t+1}] + E[e_t e_{t-1}] = \sigma_u^2 \end{aligned}$$; also does not time-varying

Since we have constant means, variance, and covariance, we can say that this sequence is stationary.

3.3 $y_{t+1} = \alpha + \delta_1 z_t + u_{t+1}$ (1)

$u_t = \rho u_{t-1} + e_t$ (2)

Rearrange and move back 2 periods to get $u_{t-1} = y_{t-1} - \alpha - \delta_1 z_{t-2}$ (3)

Substitute (3) into (2), $u_t = \rho(y_{t-1} - \alpha - \delta_1 z_{t-2}) + e_t$

$u_{t+1} = \rho(y_t - \alpha - \delta_1 z_{t-1}) + e_{t+1}$ (4)

Substitute (4) into (1), $y_{t+1} = \alpha + \delta_1 z_t + \rho(y_t - \alpha - \delta_1 z_{t-1}) + e_{t+1}$

$$E[y_{t+1}|I_t] = E[\alpha + \delta_1 z_t + \rho(y_t - \alpha - \delta_1 z_{t-1}) + e_{t+1}|I_t]$$

$$E[y_{t+1}|I_t] = \alpha + \delta_1 z_t + \rho(y_t - \alpha - \delta_1 z_{t-1}) + E[e_{t+1}|I_t],\ E[e_{t+1}|I_t]=0$$

$$E[y_{t+1}|I_t] = (1-\rho)\alpha + \rho y_t + \delta_1 z_t - \rho\delta_1 z_{t-1}$$