Problem 5

In this problem our scientific question is does the average weight increases with age. And is the increase depend on gender. To address this specific question we need to look at 3 columns specifically, age, weight, gender.

```
. browse

. replace wt =. if wt > 80
(123 real changes made, 123 to missing)
```

the first display shoukd akways be scatter plot of the data. The ammount of information that can be seen in the scatter plot is enormous.

scatter plot of the data combined and seperately for males and females.

```
. graph wt age, xlab ylab ti(" Scatter Plot of the data") saving(g1, replace)

. * now we plot seperately.

. graph wt age if sex==1, xlab ylab ti(" Scatter Plot of the data for males") s
> aving(g2, replace)

. graph wt age if sex==2, xlab ylab ti(" Scatter Plot of the data for females")
>  saving(g3, replace)

. graph using g1 g2 g3
```

now we calculate the correrlation between age and weight

```
. corr wt age
(obs=877)

         |       wt      age
---------+------------------
      wt |   1.0000
     age |   0.8776   1.0000


. sort sex

. by sex: corr wt age

-> sex=        1  (obs=455)

         |       wt      age
---------+------------------
      wt |   1.0000
     age |   0.8867   1.0000


-> sex=        2  (obs=422)

         |       wt      age
---------+------------------
      wt |   1.0000
     age |   0.8696   1.0000
```

lets make some longitudional plots

```
. graph wt age, c(L) s(i) xlab ylab ti(" spegatti plots") saving(g1, replace)

. graph wt age if sex==1, c(L) s(i) xlab ylab ti(" spegatti plots for males") s
> aving(g2, replace)

. graph wt age if sex==2, c(L) s(i) xlab ylab ti(" spegatti plots for males") s
> aving(g3, replace)

. graph using g1 g2 g3

. regress wt age

  Source |       SS       df       MS              Number of obs =     877
---------+------------------------------           F(  1,   875) = 2930.92
   Model |  5840.07408     1  5840.07408           Prob > F      =  0.0000
Residual |   1743.4997   875  1.99257109           R-squared     =  0.7701
---------+------------------------------           Adj R-squared =  0.7698
   Total |  7583.57378   876   8.6570477           Root MSE      =  1.4116

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |    .139823   .0025827     54.138   0.000       .1347539     .144892
   _cons |   5.876167     .10909     53.865   0.000       5.662059    6.090276
------------------------------------------------------------------------------

. sort sex

. by sex: regress wt age

-> sex=        1
  Source |       SS       df       MS              Number of obs =     455
---------+------------------------------           F(  1,   453) = 1666.11
   Model |  3078.24828     1  3078.24828           Prob > F      =  0.0000
Residual |  836.948634   453  1.84756873           R-squared     =  0.7862
---------+------------------------------           Adj R-squared =  0.7858
   Total |  3915.19692   454  8.62378176           Root MSE      =  1.3593

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1418124   .0034743     40.818   0.000       .1349847      .14864
   _cons |   5.956472   .1483037     40.164   0.000       5.665023     6.24792
------------------------------------------------------------------------------

-> sex=        2
  Source |       SS       df       MS              Number of obs =     422
---------+------------------------------           F(  1,   420) = 1302.91
   Model |  2734.91802     1  2734.91802           Prob > F      =  0.0000
Residual |  881.614515   420  2.09908218           R-squared     =  0.7562
---------+------------------------------           Adj R-squared =  0.7556
   Total |  3616.53253   421  8.59033855           Root MSE      =  1.4488
```

```
------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1371646      .0038     36.096   0.000       .1296952     .144634
   _cons |   5.806323   .1586514     36.598   0.000       5.494473    6.118173
------------------------------------------------------------------------------


. * lets generate some variables
. summ age

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
     age |    1000      37.662   18.63188          0         76

. gen agesq = (age-37.662)^2

. gen agecu = (age-37.662)^3

. regress wt age agesq

  Source |       SS       df       MS              Number of obs =     877
---------+------------------------------           F(  2,   874) = 1546.43
   Model |  5912.72089     2  2956.36044           Prob > F      =  0.0000
Residual |   1670.8529   874    1.911731           R-squared     =  0.7797
---------+------------------------------           Adj R-squared =  0.7792
   Total |  7583.57378   876   8.6570477           Root MSE      =  1.3827

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1412892   .0025409     55.605   0.000       .1363021    .1462762
   agesq |  -.0008666   .0001406     -6.164   0.000      -.0011426   -.0005907
   _cons |   6.115743   .1137024     53.787   0.000       5.892581    6.338904
------------------------------------------------------------------------------

. by sex: regress wt age agesq

-> sex=        1
  Source |       SS       df       MS              Number of obs =     455
---------+------------------------------           F(  2,   452) =  890.46
   Model |  3122.66228     2  1561.33114           Prob > F      =  0.0000
Residual |  792.534637   452  1.75339522           R-squared     =  0.7976
---------+------------------------------           Adj R-squared =  0.7967
   Total |  3915.19692   454  8.62378176           Root MSE      =  1.3242

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1441405    .003416     42.195   0.000       .1374272    .1508537
   agesq |  -.0009042   .0001797     -5.033   0.000      -.0012573   -.0005511
   _cons |   6.171618   .1506662     40.962   0.000       5.875525    6.467711
------------------------------------------------------------------------------

-> sex=        2
  Source |       SS       df       MS              Number of obs =     422
---------+------------------------------           F(  2,   419) =  678.63
```

```
   Model |  2763.43203     2  1381.71601               Prob > F      =  0.0000
Residual |  853.100502   419  2.03603939               R-squared     =  0.7641
---------+------------------------------               Adj R-squared =  0.7630
   Total |  3616.53253   421  8.59033855               Root MSE      =  1.4269

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1378047   .0037464     36.783   0.000       .1304406    .1451688
   agesq |  -.0008225   .0002198     -3.742   0.000      -.0012546   -.0003905
   _cons |   6.065783   .1709423     35.484   0.000       5.729772    6.401794
------------------------------------------------------------------------------
```

. regress wt age agesq agecu

```
  Source |       SS       df       MS              Number of obs =     877
---------+------------------------------           F(  3,   873) = 1030.17
   Model |  5913.22334     3  1971.07445           Prob > F      =  0.0000
Residual |  1670.35044   873   1.9133453           R-squared     =  0.7797
---------+------------------------------           Adj R-squared =  0.7790
   Total |  7583.57378   876   8.6570477           Root MSE      =  1.3832

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1437697   .0054675     26.295   0.000       .1330388    .1545007
   agesq |  -.0008593   .0001414     -6.079   0.000      -.0011368   -.0005819
   agecu |  -3.74e-06   7.30e-06     -0.512   0.608      -.0000181    .0000106
   _cons |   6.021586   .2160994     27.865   0.000       5.597451    6.445721
------------------------------------------------------------------------------
```

. by sex: regress wt age agesq agecu

```
-> sex=        1
  Source |       SS       df       MS              Number of obs =     455
---------+------------------------------           F(  3,   451) =  592.36
   Model |  3122.69977     3  1040.89992           Prob > F      =  0.0000
Residual |  792.497152   451  1.75719989           R-squared     =  0.7976
---------+------------------------------           Adj R-squared =  0.7962
   Total |  3915.19692   454  8.62378176           Root MSE      =  1.3256

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1432186    .007179     19.950   0.000         .12911    .1573271
   agesq |  -.0009084   .0001822     -4.987   0.000      -.0012664   -.0005504
   agecu |   1.34e-06   9.17e-06      0.146   0.884      -.0000167    .0000194
   _cons |   6.207017   .2854639     21.744   0.000       5.646012    6.768021
------------------------------------------------------------------------------

-> sex=        2
  Source |       SS       df       MS              Number of obs =     422
---------+------------------------------           F(  3,   418) =  453.25
   Model |  2766.17844     3   922.05948           Prob > F      =  0.0000
Residual |  850.354091   418  2.03433993           R-squared     =  0.7649
---------+------------------------------           Adj R-squared =  0.7632
```

```
   Total |  3616.53253   421  8.59033855               Root MSE      =  1.4263

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |    .146517   .0083814     17.481   0.000        .130042    .1629921
   agesq |  -.0008192   .0002197     -3.728   0.000      -.0012511   -.0003873
   agecu |  -.0000137   .0000118     -1.162   0.246      -.0000369    9.49e-06
   _cons |   5.741244   .3274362     17.534   0.000       5.097617     6.38487
------------------------------------------------------------------------------


. * calculate the autocorrelation matrix
. sort id age

. by id: generate num=_n

. autocor wt num id

          |    time1    time2    time3    time4    time5
----------+--------------------------------------------
    time1 |   1.0000
    time2 |   0.9821   1.0000
    time3 |   0.9799   0.9840   1.0000
    time4 |   0.9787   0.9757   0.9867   1.0000
    time5 |   0.9676   0.9660   0.9796   0.9831   1.0000

          acf
  1.  .9817402
  2.  .9756382
  3.  .9708902
  4.  .9676073
(Note: file acf.gph not found)

. autocor wt num id if sex==1

          |    time1    time2    time3    time4    time5
----------+--------------------------------------------
    time1 |   1.0000
    time2 |   0.9821   1.0000
    time3 |   0.9799   0.9840   1.0000
    time4 |   0.9787   0.9757   0.9867   1.0000
    time5 |   0.9676   0.9660   0.9796   0.9831   1.0000

          acf
  1.  .9817402
  2.  .9756382
  3.  .9708902
  4.  .9676073

. autocor wt num id if sex==2

          |    time1    time2    time3    time4    time5
----------+--------------------------------------------
    time1 |   1.0000
    time2 |   0.9821   1.0000
    time3 |   0.9799   0.9840   1.0000
```

```
    time4 |   0.9787   0.9757   0.9867   1.0000
    time5 |   0.9676   0.9660   0.9796   0.9831   1.0000

           acf
  1.  .9817402
  2.  .9756382
  3.  .9708902
  4.  .9676073

. * now based on these results we need to choose a model

. xtreg wt age agesq agecu, i(id) be wls

Between regression (regression on group means)  Number of obs      =       877
Group variable (i) : id                         Number of groups   =       197

R-sq:  within  = 0.7625                         Obs per group: min =         1
       between = 0.7800                                        avg =       4.5
       overall = 0.7783                                        max =         5

                                                F(3,193)           =    228.03
sd(u_i + avg(e_i.))=  1.338589                  Prob > F           =    0.0000

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1541216   .0155246      9.928   0.000       .1235018    .1847413
   agesq |   -.001004   .0003538     -2.838   0.005      -.0017019   -.0003062
   agecu |  -.0000186   .0000229     -0.814   0.417      -.0000638    .0000265
   _cons |   5.687868   .5998503      9.482   0.000       4.504764    6.870971
------------------------------------------------------------------------------


. xtreg wt age agesq agecu if sex==1, i(id) be wls

Between regression (regression on group means)  Number of obs      =       455
Group variable (i) : id                         Number of groups   =       104

R-sq:  within  = 0.7825                         Obs per group: min =         1
       between = 0.7987                                        avg =       4.4
       overall = 0.7970                                        max =         5

                                                F(3,100)           =    132.26
sd(u_i + avg(e_i.))=  1.288552                  Prob > F           =    0.0000



------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1503547   .0199631      7.532   0.000       .1107486    .1899609
   agesq |  -.0009854   .0004462     -2.208   0.030      -.0018707   -.0001001
   agecu |  -7.62e-06   .0000281     -0.272   0.786      -.0000633    .0000481
   _cons |   5.968342   .7778646      7.673   0.000       4.425081    7.511603
------------------------------------------------------------------------------
```

```
. xtreg wt age agesq agecu if sex==2, i(id) be wls

Between regression (regression on group means)  Number of obs      =       422
Group variable (i) : id                         Number of groups   =        93

R-sq:  within  = 0.6802                         Obs per group: min =         1
       between = 0.7662                                        avg =       4.5
       overall = 0.7592                                        max =         5

                                                F(3,89)            =     97.21
sd(u_i + avg(e_i.))=  1.389755                  Prob > F           =    0.0000

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1673639   .0253768      6.595   0.000       .1169409     .217787
   agesq |  -.0011024   .0005757     -1.915   0.059      -.0022463    .0000415
   agecu |  -.0000478   .0000398     -1.202   0.233      -.0001269    .0000312
   _cons |   5.065209   .9611655      5.270   0.000       3.155393    6.975024
------------------------------------------------------------------------------


. xtreg wt age agesq agecu if sex==1, i(id)  mle

Fitting constant-only model:
Iteration 0:   log likelihood = -1256.1461
Iteration 1:   log likelihood = -972.12896
Iteration 2:   log likelihood = -852.39403
Iteration 3:   log likelihood =  -813.7943
Iteration 4:   log likelihood = -806.62301
Iteration 5:   log likelihood =  -806.1854
Iteration 6:   log likelihood = -806.18226

Fitting full model:
Iteration 0:   log likelihood = -450.68453
Iteration 1:   log likelihood =  -443.5394
Iteration 2:   log likelihood = -442.76917
Iteration 3:   log likelihood = -442.76283

Random-effects ML regression                    Number of obs      =       455
Group variable (i) : id                         Number of groups   =       104

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       4.4
                                                               max =         5

                                                LR chi2(3)         =    726.84
Log likelihood  = -442.76283                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1268818   .0049508     25.629   0.000       .1171784    .1365852
   agesq |  -.0007514   .0001011     -7.434   0.000      -.0009495   -.0005533
   agecu |   .0000109   4.02e-06      2.708   0.007       3.01e-06    .0000188
   _cons |   6.759406   .2307678     29.291   0.000       6.307109    7.211703
---------+--------------------------------------------------------------------
```

```
/sigma_u |   1.292586   .0931499     13.876   0.000       1.110016    1.475157
/sigma_e |    .419181   .0158727     26.409   0.000        .388071    .4502909
---------+------------------------------------------------------------------
     rho |   .9048397    .014172                          .8739979     .929728
------------------------------------------------------------------------------
Likelihood ratio test of sigma_u=0:  chi2(1) =   658.18   Prob > chi2 = 0.0000
```

. xtreg wt age agesq agecu if sex==2, i(id) mle

```
Fitting constant-only model:
Iteration 0:   log likelihood =  -1104.067
Iteration 1:   log likelihood = -870.86552
Iteration 2:   log likelihood = -775.51694
Iteration 3:   log likelihood = -746.67906
Iteration 4:   log likelihood = -742.01616
Iteration 5:   log likelihood = -741.80733
Iteration 6:   log likelihood =  -741.8066

Fitting full model:
Iteration 0:   log likelihood = -424.17079
Iteration 1:   log likelihood = -417.93449
Iteration 2:   log likelihood = -405.97953
Iteration 3:   log likelihood = -405.02798
Iteration 4:   log likelihood = -405.01632

Random-effects ML regression                    Number of obs      =       422
Group variable (i) : id                         Number of groups   =        93

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       4.5
                                                               max =         5

                                                LR chi2(3)         =    673.58
Log likelihood  = -405.01632                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      wt |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1271716   .0053874     23.605   0.000       .1166124    .1377308
   agesq |  -.0004511   .0001022     -4.412   0.000      -.0006515   -.0002507
   agecu |   8.02e-06   4.63e-06      1.731   0.083      -1.06e-06    .0000171
   _cons |   6.305291   .2513334     25.087   0.000       5.812687    6.797896
---------+--------------------------------------------------------------------
/sigma_u |   1.381519   .1038861     13.298   0.000       1.177906    1.585132
/sigma_e |   .4098276   .0159908     25.629   0.000       .3784862    .4411691
---------+--------------------------------------------------------------------
     rho |   .9191166   .0126742                          .8912946    .9411928
------------------------------------------------------------------------------
Likelihood ratio test of sigma_u=0:  chi2(1) =   683.22   Prob > chi2 = 0.0000
```