Problem Set 4
Nonlinear Specifications with Dummy Variables and Interactions

Continuing from last week's problem set, using the data set CPS78WNH.dta, es the relationship between wage rates and education and age different for women than for men? To estimate and test this, estimate a fully---interactive model by regressing the logarithm of the wage rate on years of education, age, and age squared, including a dummy variable indicating if the person is a women, and interactions of this dummy variable with education, age, and age-squared.

```
. rename female2 female

. generate log_wage = log(wage)

. generate age_squared = age^2

. generate f_educatn = female * educatn

. generate f_age = female * age

 generate f_age_squared = female * age_squared

. regress log_wage age age_squared educatn female f_education f_age f_age_sqaured

      Source |       SS           df       MS      Number of obs   =       457
-------------+----------------------------------   F(7, 449)       =     42.21
       Model |  43.7178168         7   6.2454024   Prob > F        =    0.0000
    Residual |  66.4317298       449  .147954855   R-squared       =    0.3969
-------------+----------------------------------   Adj R-squared   =    0.3875
       Total |  110.149547       456  .241556023   Root MSE        =    .38465

------------------------------------------------------------------------------
     log_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
          age |   .0620367   .0127396     4.87   0.000     .0370001    .0870733
  age_squared |  -.0005835    .000161    -3.62   0.000    -.0008999    -.000267
      educatn |   .0527068   .0085229     6.18   0.000     .0359571    .0694564
       female |   .2737215   .4085129     0.67   0.503    -.5291131    1.076556
  f_education |   .0103007   .0173132     0.59   0.552    -.0237244    .0443258
        f_age |   -.039502   .0207235    -1.91   0.057    -.0802291    .0012251
f_age_sqaured |    .000448   .0002637     1.70   0.090    -.0000703    .0009663
        _cons |  -.2131666   .2359878    -0.90   0.367    -.6769443    .2506112
------------------------------------------------------------------------------
```

1. Test the null hypothesis that the same model applies to both men and women versus the alternative hypothesis that men and women follow different models.

```
. regress log_wage age age_squared female educatn

      Source |       SS           df       MS      Number of obs   =       457
-------------+----------------------------------   F(4, 452)       =     72.06
       Model |  42.8918917         4  10.7229729   Prob > F        =    0.0000
    Residual |   67.257655       452  .148800122   R-squared       =    0.3894
-------------+----------------------------------   Adj R-squared   =    0.3840
       Total |  110.149547       456  .241556023   Root MSE        =    .38575

------------------------------------------------------------------------------
    log_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0465571    .010068     4.62   0.000     .0267712    .0663429
 age_squared |  -.0004066   .0001277    -3.18   0.002    -.0006577   -.0001556
      female |   -.356736   .0377467    -9.45   0.000    -.4309167   -.2825553
     educatn |   .0559797   .0074208     7.54   0.000     .0413961    .0705632
       _cons |   .0460618   .1920034     0.24   0.811    -.3312682    .4233918
------------------------------------------------------------------------------
```

The coefficient for female has a p value of 0.000 which means that null hypothesis that the binary variable ‘female’ has no effect on the model can be rejected at the 1% level.

2. Is the percentage increase in wage rates for an additional year of education the same for men and women? Test the appropriate null hypothesis.

```
. regress log_wage age age_squared educatn female f_educatn f_age f_ag
> e_squared

      Source |       SS           df       MS      Number of obs   =       457
-------------+----------------------------------   F(7, 449)       =     42.21
       Model |  43.7178168         7   6.2454024   Prob > F        =    0.0000
    Residual |  66.4317298       449  .147954855   R-squared       =    0.3969
-------------+----------------------------------   Adj R-squared   =    0.3875
       Total |  110.149547       456  .241556023   Root MSE        =    .38465

------------------------------------------------------------------------------
     log_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
          age |   .0620367   .0127396     4.87   0.000     .0370001    .0870733
  age_squared |  -.0005835    .000161    -3.62   0.000    -.0008999    -.000267
      educatn |   .0527068   .0085229     6.18   0.000     .0359571    .0694564
       female |   .2737215   .4085129     0.67   0.503    -.5291131    1.076556
  f_education |   .0103007   .0173132     0.59   0.552    -.0237244    .0443258
        f_age |   -.039502   .0207235    -1.91   0.057    -.0802291    .0012251
f_age_sqaured |    .000448   .0002637     1.70   0.090    -.0000703    .0009663
        _cons |  -.2131666   .2359878    -0.90   0.367    -.6769443    .2506112
------------------------------------------------------------------------------
```

As the p value for f_education is over the 0.05, we can accept the null hypothesis that log_wage rises with age by the same about for female and male.

3. For men and women with a 4-year college degree, graph their expected wage rate age profiles on the same graph, and briefly describe and compare the shapes of the profiles. The wage rate age profile has the logarithm of the wage rate on the vertical axis, and age on the horizontal axis.

(As stata would crash overtime I tried graphing both functions in one plot, I superimposed the two images)

```
generate yhat_male = _b[_cons] + _b[age] * age + _b[age_squared] *
age_squared + _b[educatn] * 16

generate yhat_female = _b[_cons] + _b[age] * age + _b[age_squared] *
age_squared + _b[educatn] * 16 + _b[f_age] * age + _b[f_age_squared] *
age_squared + _b[f_educatn] * 16
```