Econ 312: Problem Set #5

Thursday, May 14

Question.1 Use the data set CPS04.dat to answer following questions.

a. Run a regression of average hourly earning (AHE) on age (Age), gender (Female), and education (Bachelor). If Age increase from 25 to 26, how are earnings expected to change? If Age increase from 33 to 34, how are earnings expected to change?

```
. regress ahe age female bachelor

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  3,  7982) =  624.10
       Model |   116386.54     3  38795.5133           Prob > F      =  0.0000
    Residual |  496180.729  7982  62.1624566           R-squared     =  0.1900
-------------+------------------------------           Adj R-squared =  0.1897
       Total |  612567.269  7985  76.7147487           Root MSE      =  7.8843

------------------------------------------------------------------------------
         ahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .4392042   .0305286    14.39   0.000     .3793601    .4990482
      female |  -3.157864   .1803647   -17.51   0.000    -3.511426   -2.804302
    bachelor |    6.86515   .1783686    38.49   0.000     6.515501    7.214799
       _cons |   1.883798   .9202918     2.05   0.041     .0797852     3.68781
------------------------------------------------------------------------------
```

Answer:If Age increases from 25 to 26, earnings are predicted to increase by $0.439 per hour. If Age increases from 33 to 34, earnings are predicted to increase by $0.439 per hour. These values are the same because the regression is a linear function relating AHE and Age.

b. Run a regression of the logarithm average hourly earnings, $\ln(AHE)$,on Age, Female, and Bachelor. If Age increase from 25 to 26, how are earnings expected to change? If Age increase from 33 to 34, how are earnings expected to change? (Stata: generate lnahe=ln(ahe))

```
 generate lnahe=ln(ahe)
 regress lnahe age female bachelor

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  3,  7982) =  633.75
       Model |  397.245741     3  132.415247           Prob > F      =  0.0000
    Residual |  1667.74691  7982  .208938476           R-squared     =  0.1924
-------------+------------------------------           Adj R-squared =  0.1921
```

```
       Total |  2064.99265  7985  .258608974           Root MSE      =   .4571

------------------------------------------------------------------------------
       lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0244429   .0017699    13.81   0.000     .0209735    .0279124
      female |  -.1804636   .0104567   -17.26   0.000    -.2009616   -.1599657
    bachelor |   .4052749    .010341    39.19   0.000     .3850038     .425546
       _cons |   1.856457   .0533545    34.79   0.000     1.751868    1.961046
--------------------------------------------------------------------------
```

Answer:If Age increases from 25 to 26, ln(AHE) is predicted to increase by 0.024. This means that earnings are predicted to increase by 2.4%. If Age increases from 34 to 35, ln(AHE) is predicted to increase by 0.024. This means that earnings are predicted to increase by 2.4%. These values, in percentage terms, are the same because the regression is a linear function relating $ln(AHE)$ and Age.

c. Run a regression of the logarithm average hourly earnings, ln(AHE), on ln(Age), Female, and Bachelor. If Age increase from 25 to 26, how are earnings expected to change? If Age increase from 33 to 34, how are earnings expected to change?

```
. generate lnage=ln(age)

. regress lnahe lnage female bachelor

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  3,  7982) =  635.03
       Model |  397.892296     3  132.630765           Prob > F      =  0.0000
    Residual |  1667.10036  7982  .208857474           R-squared     =  0.1927
-------------+------------------------------           Adj R-squared =  0.1924
       Total |  2064.99265  7985  .258608974           Root MSE      =  .45701

------------------------------------------------------------------------------
       lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnage |   .7246973   .0520447    13.92   0.000     .6226762    .8267184
      female |  -.1802958    .010455   -17.24   0.000    -.2007903   -.1598013
    bachelor |   .4052329    .010339    39.19   0.000     .3849657       .4255
       _cons |   .1282838     .17662     0.73   0.468    -.2179376    .4745051
------------------------------------------------------------------------------
```

Answer: If Age increases from 25 to 26, then ln(Age) has increased by $\ln(26) - \ln(25) = 0.0392$ (or 3.92%). The predicted increase in $ln(AHE)$ is $0.725 \times (.0392) = 0.0284$. This means that earnings are predicted to increase by 2.8%. If Age increases from 34 to 35, then ln(Age) has increased by $\ln(35) - \ln(34) = .0290$ (or 2.90%). The predicted increase in ln(AHE) is $0.725 \times (0.0290) = 0.0210$. This means that earnings are predicted to increase by 2.10%.

d. Run a regression of the logarithm average hourly earning, $\ln(AHE)$ on, Age, Age^2, Female, and Bachelor. If Age increase from 25 to 26, how are earnings expected to change? If Age increase from 33 to 34, how are earnings expected to change?

```
. generate agesqr=age*age

. regress lnahe age agesqr female bachelor

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  4,  7981) =  477.96
       Model |  399.069759     4  99.7674398           Prob > F      =  0.0000
    Residual |   1665.9229  7981   .20873611           R-squared     =  0.1933
-------------+------------------------------           Adj R-squared =  0.1929
       Total |  2064.99265  7985  .258608974           Root MSE      =  .45688

------------------------------------------------------------------------------
       lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1470452   .0415124     3.54   0.000     .0656701    .2284203
      agesqr |  -.0020706   .0007004    -2.96   0.003    -.0034436   -.0006975
      female |  -.1797868   .0104542   -17.20   0.000    -.2002797   -.1592938
    bachelor |   .4050769   .0103362    39.19   0.000     .3848152    .4253386
       _cons |   .0587333   .6104789     0.10   0.923    -1.137965    1.255431
------------------------------------------------------------------------------
```

Anwser: When Age increases from 25 to 26, the predicted change in ln(AHE) is $(0.147 \times 26 - 0.0021 \times 26^2) - (0.147 \times 25 - 0.0021 \times 25^2) = 0.0399$. This means that earnings are predicted to increase by 3.99%. When Age increases from 34 to 35, the predicted change in ln(AHE) is $(0.147 \times 35 - 0.0021 \times 35^2) - (0.147 \times 34 - 0.0021 \times 34^2) = 0.0063$. This means that earnings are predicted to increase by 0.63%.

e. Do you prefer the regression in (c) to the regression in (b)? Explain.

Answer: The regressions differ in their choice of one of the regressors. They can be compared on the basis of the $\bar{R}^2$. The regression in (3) has a (marginally) higher $\bar{R}^2$ so it is preferred.

f. Do you prefer the regression in (d) to the regression in (b)? Explain.

Answer: The regression in (4) adds the variable Age^2 to regression (2). The coefficient on Age^2 is statistically significant ($t = -2.96$), and this suggests that the addition of Age^2 is important. Thus, (4) is preferred to (2).

g. Do you prefer the regression in (d) to the regression in (c)? Explain.

The regressions differ in their choice of one of the regressors. They can be compared on the basis of the $\bar{R}^2$. The regression in (4) has a (marginally) higher $\bar{R}^2$ so it is preferred.

h. Run a regression of ln(AHE), on Age, Age^2, Female, Bachelor, and the interaction term $Female \times Bachelor$. What odes the coefficient on the interaction term measure? Alexis is a 30-year-old female with a bachelor's degree. What does the regression predict for her value of ln(AHE)? Jane is a 30 year old female with a high school degree? What does the regression predict for her value of ln(AHE)? What is the difference between Alexis' and Jane's earning?

```
. gen lnahe=log(ahe)

. gen agesqr=age*age

. gen fbachelor=female*bachelor

. regress lnahe age agesqr female bachelor fbachelor

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  5,  7980) =  384.61
       Model |  400.993272     5  80.1986543           Prob > F      =  0.0000
    Residual |  1663.99938  7980  .208521226           R-squared     =  0.1942
-------------+------------------------------           Adj R-squared =  0.1937
       Total |  2064.99265  7985  .258608974           Root MSE      =  .45664

------------------------------------------------------------------------------
       lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1461799    .041492     3.52   0.000     .0648448     .227515
      agesqr |  -.0020516   .0007001    -2.93   0.003     -.003424   -.0006792
      female |  -.2099559   .0144169   -14.56   0.000    -.2382168   -.1816951
    bachelor |   .3781866   .0136057    27.80   0.000     .3515158    .4048573
   fbachelor |   .0635898    .020937     3.04   0.002     .0225478    .1046319
       _cons |   .0784239    .610199     0.13   0.898    -1.117726    1.274573
------------------------------------------------------------------------------
```

Answer: The coefficient on the interaction term Female Bachelor shows the "extra effect" of Bachelor on ln(AHE) for women relative the effect for men. Predicted values of ln(AHE):

Alexis: $0.146 \times 30 + 0.0021 \times 30^2 + 0.180 \times 1 + 0.405 \times 1 + 0.064 \times 1 + 0.078 = 4.504$
Jane: $0.146 \times 30 + 0.0021 \times 30^2 + 0.180 \times 1 + 0.405 \times 0 + 0.064 \times 0 + 0.078 = 4.063$

Difference in ln(AHE): Alexis- Jane= $4.504 - 4.063 = 0.441$

j. Is the effect of Age on earnings different for male than for female? Specify and estimate a regression that you can use to answer this question.

```
. gen fage=female*age

. gen fagesqr=female*agesqr

. regress lnahe age agesqr female bachelor fbachelor fage fagesqr

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  7,  7978) =  276.11
       Model |  402.709946     7  57.5299923           Prob > F      =  0.0000
    Residual |  1662.28271  7978  .208358324           R-squared     =  0.1950
-------------+------------------------------           Adj R-squared =  0.1943
       Total |  2064.99265  7985  .258608974           Root MSE      =  .45646

------------------------------------------------------------------------------
       lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .190451   .0542687     3.51   0.000     .0840703    .2968318
      agesqr |  -.0027297   .0009145    -2.98   0.003    -.0045223   -.000937
      female |   1.358419   1.237642     1.10   0.272    -1.067683    3.784522
    bachelor |   .3771552   .0136059    27.72   0.000     .3504842    .4038262
   fbachelor |    .063097    .020932     3.01   0.003     .0220647    .1041292
        fage |  -.0971479   .0842279    -1.15   0.249    -.2622566    .0679608
     fagesqr |   .0014796   .0014219     1.04   0.298    -.0013077    .0042669
       _cons |   -.632741   .7991521    -0.79   0.429    -2.199288     .933806
------------------------------------------------------------------------------
. test  fage fagesqr

 ( 1)  fage = 0
 ( 2)  fagesqr = 0

       F(  2,  7978) =    4.12
            Prob > F =    0.0163
```

Answer: This regression is includes two additional regressors: the interactions of Female and the age variables, Age and Age^2. The F-statistic testing the restriction that the coefficients on these interaction terms is equal to zero is F=4.12 with a p-value of 0.02. This implies that there is statistically significant evidence (at the 5% level) that there is a different effect of Age on $\ln(AHE)$ for men and women.

k. Is the effect of Age on earnings different for high school graduates than college graduates? Specify and estimate a regression that you can use to answer this question.

```
. gen bage=bachelor*age

. gen bagesqr=bachelor*agesqr
```

```
. regress lnahe age agesqr female bachelor fbachelor bage bagesqr

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  7,  7978) =  277.23
       Model |  404.022454     7  57.7174934           Prob > F      =  0.0000
    Residual |   1660.9702  7978  .208193808           R-squared     =  0.1957
-------------+------------------------------           Adj R-squared =  0.1949
       Total |  2064.99265  7985  .258608974           Root MSE      =  .45628

------------------------------------------------------------------------------
       lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1165546   .0564857     2.06   0.039     .0058278    .2272814
      agesqr |  -.0016535   .0009531    -1.73   0.083    -.0035219    .0002148
      female |  -.2094241   .0144098   -14.53   0.000    -.2376711   -.1811771
    bachelor |  -.7694888   1.223104    -0.63   0.529    -3.167092    1.628114
   fbachelor |   .0664672   .0209366     3.17   0.002      .025426    .1075084
        bage |   .0644098   .0831686     0.77   0.439    -.0986224     .227442
     bagesqr |  -.0008619   .0014033    -0.61   0.539    -.0036128     .001889
       _cons |   .6039909   .8306348     0.73   0.467     -1.02427    2.232252
------------------------------------------------------------------------------

. test  bage bagesqr

 ( 1)  bage = 0
 ( 2)  bagesqr = 0

       F(  2,  7978) =    7.27
            Prob > F =    0.0007
```

This regression includes two additional regressors that are interactions of Bachelor and the age variables, Age and Age^2. The F-statistic testing the restriction that the coefficients on these interaction terms is zero is 7.15 with a p-value of 0.00. This implies that there is statistically significant evidence (at the 1% level) that there is a different effect of Age on $\ln(AHE)$ for high school and college graduates.

I. After running all of these regression (and any others that you want to run), summarize the effect of age on earnings for young workers.

```
. regress lnahe age agesqr female bachelor fbachelor fage fagesqr bage bagesqr

      Source |       SS       df       MS              Number of obs =    7986
-------------+------------------------------           F(  9,  7976) =  217.18
       Model |  406.449755     9  45.1610838           Prob > F      =  0.0000
    Residual |   1658.5429  7976  .207941688           R-squared     =  0.1968
-------------+------------------------------           Adj R-squared =  0.1959
```

```
      Total |  2064.99265  7985  .258608974           Root MSE      =  .45601

------------------------------------------------------------------------------
      lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age |   .1598879   .0638322     2.50   0.012     .0347601    .2850156
     agesqr |  -.0023137   .0010771    -2.15   0.032     -.004425   -.0002023
     female |   1.763571   1.250497     1.41   0.158    -.6877305    4.214873
   bachelor |  -1.186432   1.236376    -0.96   0.337    -3.610052    1.237188
  fbachelor |   .0663196   .0209272     3.17   0.002     .0252968    .1073425
       fage |  -.1228694    .085071    -1.44   0.149    -.2896308    .0438919
    fagesqr |   .0018832   .0014356     1.31   0.190    -.0009309    .0046973
       bage |   .0912362    .084036     1.09   0.278    -.0734963    .2559687
    bagesqr |  -.0012901   .0014174    -0.91   0.363    -.0040685    .0014884
      _cons |  -.0953266   .9387221    -0.10   0.919    -1.935467    1.744814
------------------------------------------------------------------------------
```

Answer: This regression includes Age and Age^2 and interactions terms involving Female and Bachelor.

The estimated regressions suggest that earnings increase as workers age from 25-35, the range of age studied in this sample. There is evidence that the quadratic term Age2 belongs in the regression. Curvature in the regression functions in particularly important for men. Gender and education are significant predictors of earnings, and there are statistically significant interaction effects between age and gender and age and education. The table below summarizes the regressions predictions for increases in earnings as a person ages from 25 to 32 and 32 to 35

Gender, Education	Predicted	ln(AHE)	at Age	(% per	year)
	25	32	35	25 to 32	32 to 35
Females, High School	2.32	2.41	2.44	1.2%	0.8%
Males, High School	2.46	2.65	2.67	2.8%	0.5%
Females, BA	2.68	2.89	2.93	3.0%	1.3%
Males, BA	2.74	3.06	3.09	4.6%	1.0%

Earnings for those with a college education are higher than those with a high school degree, and earnings of the college educated increase more rapidly early in their careers (age 25-32). Earnings for men are higher than those of women, and earnings of men increase more rapidly early in their careers (age 25-32). For all categories of workers (men/women, high school/college) earnings increase more rapidly from age 25-32 than from 32-35.