**Homework 2 – Answer Key**

**Problem 1**

a. Suppose that I claim that the relationship between wage and tenure is not linear. What is the probability that I'm wrong?

```
. gen tenure2 = tenure*tenure

. reg lwage tenure tenure2

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  2,   932) =   19.13
       Model |  6.53391236     2  3.26695618           Prob > F      =  0.0000
    Residual |  159.122382   932  .170732169           R-squared     =  0.0394
-------------+------------------------------           Adj R-squared =  0.0374
       Total |  165.656294   934  .177362199           Root MSE      =   .4132

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      tenure |   .0343808   .0090926     3.78   0.000     .0165364    .0522252
     tenure2 |   -.001138   .0005219    -2.18   0.029    -.0021622   -.0001139
       _cons |   6.619126   .0323228   204.78   0.000     6.555692     6.68256
------------------------------------------------------------------------------
```

To test if the relationship between tenure and log(wage) is not linear we test if the coefficient on $tenure^2$ is different from zero. If the coefficient is statistically different from zero then the relationship is not linear. We find that the pvalue is 2.9%, which is enough evidence to reject a null hypothesis of non-significance at 95% confidence. The probability that we are falsely rejecting the null is 2.9%.

b. Please solve for the length of job tenure at which the returns to additional time at the same job are zero. Is this a maximum or minimum? Why?

To find the length of job tenure at which the additional time at the same job gives zero return we have to take the derivative of the function and set it equal to zero.

$$log(wage) = \beta_0 + \beta_{ten} tenure + \beta_{ten2} tenure^2 + u$$

$$log\widehat{(wage)} = 6.619 + 0.034 tenure - 0.001 tenure^2$$

Take derivative and set equal to zero:

$$\frac{dlog(wage)}{dtenure} = 0.034 - 0.002 tenure = 0$$

$$tenure = 17$$

To determine whether 17 years of tenure is a maximum or a minimum, we take the second derivative. If the second derivative is negative the function is concave and therefore we just found a maximum. If the second derivative is positive, the function is convex and we have just found a minimum.

Taking the second derivative:

$$\frac{d^2 log(wage)}{dtenure^2} = -0.002$$

Since the second derivative is negative, 17 years of tenure is a maximum.

## PROBLEM 2

a. Please write down the code used to construct the variable first using the variable brthord.

```
. gen first = 0

. replace first = 1 if brthord == 1

. replace first = . if brthord == .

** The third line is required in order to make people with missing values for birth order also
have missing values for first born.
```

b. Please construct and interpret a 90% confidence interval for $\beta_{first}$

```
. reg wage married urban south first, level(90)

      Source |       SS       df       MS              Number of obs =     852
-------------+------------------------------           F(  4,   847) =   21.35
       Model |  12713656.9     4  3178414.22           Prob > F      =  0.0000
    Residual |   126074248   847  148847.991           R-squared     =  0.0916
-------------+------------------------------           Adj R-squared =  0.0873
       Total |   138787905   851   163088.02           Root MSE      =  385.81

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   176.4662   45.34089     3.89   0.000     101.8054     251.127
       urban |   190.0187   29.56606     6.43   0.000     141.3336    238.7038
       south |  -100.3879    28.1076    -3.57   0.000    -146.6714   -54.10436
       first |   97.04306   26.95663     3.60   0.000     52.65481    141.4313
       _cons |   669.0287   51.60444    12.96   0.000     584.0541    754.0034
------------------------------------------------------------------------------
```

At 90% confidence, and holding other variables constant, someone who is first born earns between 52.66 and 141.43 more dollars per month than someone who is not first born on average.

c. Please construct and interpret a 99% confidence interval for $\beta_0$

```
. reg wage married urban south first, level(99)

      Source |       SS       df       MS              Number of obs =     852
-------------+------------------------------           F(  4,   847) =   21.35
       Model |  12713656.9     4  3178414.22           Prob > F      =  0.0000
    Residual |   126074248   847  148847.991           R-squared     =  0.0916
-------------+------------------------------           Adj R-squared =  0.0873
       Total |   138787905   851   163088.02           Root MSE      =  385.81

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [99% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   176.4662   45.34089     3.89   0.000     59.41202    293.5203
       urban |   190.0187   29.56606     6.43   0.000     113.6896    266.3478
       south |  -100.3879    28.1076    -3.57   0.000    -172.9518   -27.82399
       first |   97.04306   26.95663     3.60   0.000     27.45059    166.6355
       _cons |   669.0287   51.60444    12.96   0.000     535.8043    802.2531
------------------------------------------------------------------------------
```

At 99% confidence, someone who is not married, lives in a rural area, lives in the north and is not the first born earns between 535.80 and 802.25 dollars per month on average.

**PROBLEM 3**

a. Please construct and carefully interpret a 90% confidence interval for $\beta_{educ}$

```
. gen educ_first = educ*first

. reg lwage educ first educ_first, level(90)

      Source |       SS       df       MS              Number of obs =     852
-------------+------------------------------           F(  3,   848) =   35.33
       Model |  16.3115248     3  5.43717493           Prob > F      =  0.0000
    Residual |   130.50455   848  .153896875           R-squared     =  0.1111
-------------+------------------------------           Adj R-squared =  0.1080
       Total |  146.816075   851  .172521827           Root MSE      =   .3923

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0536367   .0082472     6.50   0.000     .0400564     .067217
       first |  -.1282655   .1721844    -0.74   0.457    -.4117934    .1552623
  educ_first |   .0134244   .0124802     1.08   0.282    -.0071261    .0339749
       _cons |   6.046634   .1104087    54.77   0.000     5.864829    6.228439
------------------------------------------------------------------------------
```

At 90% confidence, an additional year of education for someone who is not a first born on average has a return between 4.01% and 6.72% on monthly wage.

b. There are a variety of reasons why a first born child may receive more resources that later arriving children, and therefore benefit more from education. Please evaluate this hypothesis using the 95% confidence level and a two sided test. Please interpret your result, and interpret your coefficient estimate from your regression.

If first born receives additional returns to every additional year of education, then the coefficient $\beta_{EF}$ should be statistically different from zero. To test this, we use hypothesis testing at the 95% confidence level.

$H_0: \beta_{EF} = 0$
$H_A: \beta_{EF} \neq 0$

$$t_{stat} = \frac{.0134244 - 0}{.1104087} = 1.08$$

Tcritical = 1.96

At 95% confidence level we fail to reject that $\beta_{EF}$ is statistically different from zero, since the t-stat is less than the t-crit. This means that first borns don't get additional returns per each additional year of education.