Question 1 – Monica Elgawly.

Assignment #3

```
*================================================================================================
*Title: multnonr_monica_answers.do
*Author: Monica Elgawly
*Date Modified: 3/30/2018
*Assignment #3
*================================================================================================

cap log close
clear
set more off

*================================================================================================
*to create a log file for results
*================================================================================================

log using "/Users/monicaelgawly/Downloads/metrics/multnr_monica_answers.log", replace

*================================================================================================
*Question 1: *Open the data set on how tuition is valued/ how today values college education
*================================================================================================

use "/Users/monicaelgawly/Downloads/HTV.dta"

*================================================================================================
*a): Show the range of values that the edu variable takes in the sample (sum, detail)
*Create a dummy variable that equals 1 if men completed exactly 12 years of education and
equals 0 otherwise (gen)
*Show what percentage of men completed exactly 12 years of education (tab)
*Determine whether the men or their parents have higher levels of education (sum)
```



```

sum
ge highschedu=0
replace highschedu=1 if educ==12

tab highschedu
* From the tabulation, 41.74% of the men in 1991 had finished 12 years of education -- that
is
* 498 out of (695+498) 1193 total in the sample
*The sum function shows that the mothers' has a typical education of 12.17519 +/- 2.28
years while the
*fathers had only slightly higher with 12.44258 years with more variation +/- 3.27. The
mean for the men
*overall was 13.03 +/- 2.35 years. Therefore the men had a year on average more of
education over their
*parents.
*================================================================================================
*================================================================================================
*b): Estimate the regression model: educi = B0 + B1(mothereduci) + B2(fatheduci) + ui

reg educ motheduc fatheduc
reg educ motheduc fatheduc, r
reg educ motheduc fatheduc, vce(robust)

*How much of the variation in educ is explained by parents' education? Interpret the
coefficient on mothereduc.
*The amount of variation explained by these two regressors is found via R^2 which is approx
25%
*Here, B1 reflects the difference in educations for men whose mothers ranged from highly to
least traditionally educated via
*number of years. Since the coefficient [B1 = .31] > [B2 = .18], men born to more educated
mothers were likewise expected
*to be more educated on average and fathers' education [B2] didn't correlate with their
sons' education as much.
```

```
54  *================================================================================================
    ==========
55
56
57  *c) Add the variable abil (a measure of cognitive ability) to the model in part (b). Does
    ability help to
58  *explain variation in education, even after controlling for parents' education? Use at
    least one test statistic
59  *to justify your answer.
60
61  reg educ abil motheduc fatheduc
62  reg educ abil motheduc fatheduc, r
63
64  *Adding ability to the regression does help to explain variation based off the observations
    found in the
65  *decrease of the standard error of all the regressors and the increase in width of range
    whereby we
66  *attain more confidence in our observations. From the addition of abil, we see that B1, the
    coefficient
67  *for motheduc, increases numerically from .31 to .49. Eliminating abil introduces positive
    bias in the regression then.
68  *(LC #9) The large t statistic for abil tells us that we reject the null hypothesis at 95%
    confidence that
69  * H0 = abil = motheduc = fatheduc = 0. Disproving this hypothesis shows the validity of the
    need of all three
70  *regressors in the equation. The p-value being zero also indicates a rejection of the null
    (H0 = all are assumed zero).
71
72  *d) Compare the coefficient estimates on motheduc and fatheduc from the model in part b) to
    the model in part c).
73  *What do the differences tell us about omitted variable bias in the model in part b)? Be
    specific.
74
75  * There was omitted variable bias with regards to ability from the recognition of the value
    of
76  * R^2 increasing from .25 to .42. The correlation coefficients of motheduc and fatheduc go
```



```
    down
77  * when abil is added to the regression, revealing that ability is correlated with parents'
    abilities.
78  * The same way when children are given more capable parents --maybe providing more
    opportunities or
79  * support as there were resources previously passed down to them. So too the sooner that
    more educated
80  * parents can pass down that fountain of knowledge to their children. The coefficient then
    of ability
81  * reveals the men's educated results owing more to their own merit than to their parents'
    credit.
82  * Or, said another way, weighing in numerically the men's childhood environment growing up.
83
84  *e) Use the results from part (c) to test the hypothesis that the effect of mothers'
    education on
85  * men's education equals two times the effect of fathers' education on men's education.
86  * Write null and alternative hypotheseses.
87  * Evaluate the hypothesis using the F-statistic and test command.
88  * Rearrange the regression model and evaluate the hypothesis using a t-statistic.
89  * Do you arrive at the same conclusion using the F- and t- statistics?
90
91  *H0 = B1 = (2 * B2)
92  *H1 = B1 \= (2 * B2)
93  * to evaluate the hypothesis:
94
95  test motheduc fatheduc
96  * We reject the null that B1 = B2 = 0
97
98  test motheduc-(2*fatheduc)=0
99  *The F-test yields a p-value of .7731 so we fail to reject the null.
100
101 *We arrive at results that support one another. The t-tests reveal better reliability in
102 *the regression once ability is accounted for and the F-test reveals the
103 *relationship between the two parents' associated impacts on the men's education.
104
105
```

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis             StataCorp
                                       4905 Lakeway Drive
                                       College Station, Texas 77845 USA
                                       800-STATA-PC        http://www.stata.com
                                       979-696-4600        stata@stata.com
                                       979-696-4601 (fax)

Single-user Stata license expires 20 Aug 2018:
       Serial number:  301509284846
         Licensed to:  Monica Elgawly
                       Hunter College

Notes:
      1.  Unicode is supported; see help unicode_advice.

1 . log using "/Users/monicaelgawly/Downloads/metrics/metrics_work_3_30_18.smcl"
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/monicaelgawly/Downloads/metrics/metrics_work_3_30_18.smcl
  log type:  smcl
 opened on:  30 Mar 2018, 23:13:10

2 . clear

3 . log close
      name:  <unnamed>
       log:  /Users/monicaelgawly/Downloads/metrics/metrics_work_3_30_18.smcl
  log type:  smcl
 closed on:  30 Mar 2018, 23:37:19
-------------------------------------------------------------------------------
```



```
4 . log using "/Users/monicaelgawly/Downloads/metrics/multnr_monica_answers.log"
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/monicaelgawly/Downloads/metrics/multnr_monica_answers.log
  log type:  text
 opened on:  30 Mar 2018, 23:37:45

5 . use "/Users/monicaelgawly/Downloads/HTV.dta"

6 . sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |      1,193    13.23942    9.116401   1.023529   91.30922
        abil |      1,193    1.805191    2.169927  -5.631463   6.263742
        educ |      1,193    13.03437    2.346208          6         20
          ne |      1,193     .210394    .4077594          0          1
          nc |      1,193    .3730092    .4838073          0          1
-------------+---------------------------------------------------------
        west |      1,193    .1684828    .3744514          0          1
       south |      1,193     .248114    .4320995          0          1
       exper |      1,193    10.72842    3.105527          1         19
   motheduc |      1,193    12.17519    2.282187          0         20
   fatheduc |      1,193    12.44258      3.2721          0         20
-------------+---------------------------------------------------------
    brkhme14 |      1,193    .1659681    .3722081          0          1
        sibs |      1,193    2.955574    1.860181          0         13
       urban |      1,193    .8147527     .388661          0          1
        ne18 |      1,193    .2279966    .4197164          0          1
        nc18 |      1,193    .4115675    .4923239          0          1
-------------+---------------------------------------------------------
     south18 |      1,193    .2145851    .4107064          0          1
     west18  |      1,193    .1458508    .3531046          0          1
     urban18 |      1,193    .8751048    .3307387          0          1
      tuit17 |      1,193    8.537157    4.121537          0   18.17392
```

tuit18	1,193	8.557239	4.042644	0	18.17392
lwage	1,193	2.408684	.5960768	.0232569	4.514252
expersq	1,193	124.7351	66.80132	1	361
ctuit	1,193	.0200824	.8146919	-7.623878	9.173661

```
7 . gen highschedu=educ=12;
invalid syntax
r(198);

8 . gen highschedu=educ=12
invalid syntax
r(198);

9 . ge highschedu=0

10 . replace highschool=1 if educ==12
variable highschool not found
r(111);

11 . replace highschedu=1 if educ==12
(498 real changes made)

12 . tab highschedu
```

highschedu	Freq.	Percent	Cum.
0	695	58.26	58.26
1	498	41.74	100.00
Total	1,193	100.00	

```
13 . sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max

wage	1,193	13.23942	9.116401	1.023529	91.30922
abil	1,193	1.805191	2.169927	-5.631463	6.263742
educ	1,193	13.03437	2.346208	6	20
ne	1,193	.210394	.4077594	0	1
nc	1,193	.3730092	.4838073	0	1
west	1,193	.1684828	.3744514	0	1
south	1,193	.248114	.4320995	0	1
exper	1,193	10.72842	3.105527	1	19
motheduc	1,193	12.17519	2.282187	0	20
fatheduc	1,193	12.44258	3.2721	0	20
brkhme14	1,193	.1659681	.3722081	0	1
sibs	1,193	2.955574	1.860181	0	13
urban	1,193	.8147527	.388661	0	1
ne18	1,193	.2279966	.4197164	0	1
nc18	1,193	.4115675	.4923239	0	1
south18	1,193	.2145851	.4107064	0	1
west18	1,193	.1458508	.3531046	0	1
urban18	1,193	.8751048	.3307387	0	1
tuit17	1,193	8.537157	4.121537	0	18.17392
tuit18	1,193	8.557239	4.042644	0	18.17392
lwage	1,193	2.408684	.5960768	.0232569	4.514252
expersq	1,193	124.7351	66.80132	1	361
ctuit	1,193	.0200824	.8146919	-7.623878	9.173661
highschedu	1,193	.417435	.4933427	0	1

```
14 . reg educ motheduc fatheduc
```

Source	SS	df	MS	Number of obs	=	1,193
				F(2, 1190)	=	201.32
Model	1658.85617	2	829.428084	Prob > F	=	0.0000

Source	SS	df	MS		
Residual	4902.73478	1,190	4.11994519	R-squared	= 0.2528
				Adj R-squared	= 0.2516
Total	6561.59095	1,192	5.50469039	Root MSE	= 2.0298

educ	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
motheduc	.3133269	.0320423	9.78	0.000	.2504612	.3761925
fatheduc	.1848613	.0223485	8.27	0.000	.1410145	.2287081
_cons	6.919402	.3224073	21.46	0.000	6.286852	7.551952

15 . reg educ motheduc fatheduc, r

Linear regression

Number of obs	= 1,193
F(2, 1190)	= 166.79
Prob > F	= 0.0000
R-squared	= 0.2528
Root MSE	= 2.0298

educ	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
motheduc	.3133269	.0390176	8.03	0.000	.2367759	.3898778
fatheduc	.1848613	.0259798	7.12	0.000	.13389	.2358326
_cons	6.919402	.3610376	19.17	0.000	6.211061	7.627743

16 . reg educ motheduc fatheduc, vce(robust)

Linear regression

Number of obs	= 1,193
F(2, 1190)	= 166.79
Prob > F	= 0.0000
R-squared	= 0.2528

Root MSE	= 2.0298

educ	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
motheduc	.3133269	.0390176	8.03	0.000	.2367759	.3898778
fatheduc	.1848613	.0259798	7.12	0.000	.13389	.2358326
_cons	6.919402	.3610376	19.17	0.000	6.211061	7.627743

17 . reg educ abil motheduc fatheduc

Source	SS	df	MS		
Model	2781.42227	3	927.140756	Number of obs	= 1,193
				F(3, 1189)	= 291.62
Residual	3780.16868	1,189	3.179284	Prob > F	= 0.0000
				R-squared	= 0.4239
				Adj R-squared	= 0.4224
Total	6561.59095	1,192	5.50469039	Root MSE	= 1.7831

educ	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
abil	.4958059	.0263858	18.79	0.000	.4440379	.5475738
motheduc	.1937562	.028858	6.71	0.000	.137138	.2503745
fatheduc	.1093409	.0200393	5.46	0.000	.0700246	.1486572
_cons	8.419842	.2942611	28.61	0.000	7.842513	8.997171

18 . reg educ abil motheduc fatheduc, r

Linear regression

Number of obs	= 1,193
F(3, 1189)	= 249.04
Prob > F	= 0.0000
R-squared	= 0.4239

Root MSE = 1.7831

educ	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
abil	.4958059	.0281263	17.63	0.000	.4406232	.5509885
motheduc	.1937562	.0344251	5.63	0.000	.1262154	.261297
fatheduc	.1093409	.0237096	4.61	0.000	.0628236	.1558581
_cons	8.419842	.3300946	25.51	0.000	7.772209	9.067474

```
19 . reg educ motheduc fatheduc abil, r

Linear regression                               Number of obs     =      1,193
                                                F(3, 1189)        =     249.04
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4239
                                                Root MSE          =     1.7831
```

educ	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
motheduc	.1937562	.0344251	5.63	0.000	.1262154	.261297
fatheduc	.1093409	.0237096	4.61	0.000	.0628236	.1558581
abil	.4958059	.0281263	17.63	0.000	.4406232	.5509885
_cons	8.419842	.3300946	25.51	0.000	7.772209	9.067474

```
20 . test motheduc fatheduc

 ( 1)  motheduc = 0
 ( 2)  fatheduc = 0

       F(  2,  1189) =   63.11
            Prob > F =    0.0000

21 . test motheduc-(2*fatheduc)=0

 ( 1)  motheduc - 2*fatheduc = 0

       F(  1,  1189) =    0.12
            Prob > F =    0.7331

22 .
```


User: Monica Elgawly

Question 2 - Monica Elgawly.


```
1  *================================================================================================
2  *Title: mult2_answers_monica.do
3  *Author: Monica Elgawly
4  *Date Modified: 4/05/2018
5  *Assignment 3 Part 2
6  *================================================================================================
7
8  cap log close
9  clear
10 set more off
11
12 *================================================================================================
13
14 *================================================================================================
15 *2a) Use OLS to estimate the equation: log(wagei) = B0 + B1edu1 + B2experi + B3exper^2^i + ui
16 *Write the regression results in standard form (i.e. plug in the correct values for the
   parameters
17 *estimates and their standard errors): log(wagei) = B0 + B1edu1 + B2experi + B3exper^2^i
18
19 gen lwage = log(wage)
20 *already defined upon use of data file
21
22 gen exper2 = exper^2
23 regress lwage educ exper exper2
24
25 *================================================================================================
26
27 *================================================================================================
28 *b) Interpret B1 and test its statistical significance at the 1% level. What can we say
   about
29 *the relationship between education and log(wage)?
30
31 *B1 represents the coefficient of the multiplicative effect the number of years of
   education has
32 *on longterm wage outcomes. When we test B1 for statistical significance we take the
   coefficient
```



```
33 *subtract from its null hypothesized value of insignificance zero and divide by the
   standard error
34 *(.0903658 - 0)/.007468 to give 12.10. This value is greater than the critical value of
   2.58 which
35 *is the value used to test for 1% significance. |12.10| > 2.58 means B1 is significant at
   the
36 *1% level.
37
38 *The relationship between education and log(wage) is a log-linear relationship as found in
   lecture #15
39 *slide 6/19 where it is iterated that a one unit increase in education is expected to
   increase
40 *wages by (100 * .0903658) which is 9.04%.
41
42 *================================================================================================
43
44 *================================================================================================
45 *c) Test whether experience has a statistically significant effect on log(wage).
46
47 *To test whether experience has a statistically significant effect on log(wage), we look at
   the
48 *t statistic value provided by the regression. This could've been a simplier way of looking
   at
49 *the educational statistical significance as well, but I preferred to utilize breaking down
50 *the information for my own understanding purposes.
51
52 *The t for experience is 7.89 which is again greater than 2.58. Experience is statistically
   significiant
53 *at the 1% level.
54 *================================================================================================
55
56 *================================================================================================
57 *d) Is exper^2 statistically significant at the 5% level? What does this hypothesis test say
58 *about the relationship between log(wage) and experience?
59
60 *|-6.16| > 1.96 so exper^2 is statistically significant at the 5% level.
```

```
61  *This hypothesis test says that the relationship between experience and log(wage) is
62  *in fact nonlinear. The null hypothesis would have been B2 = 0. The alternative B2 \= 0.
63  *Since |-6.16| > 1.95 or even 2.58 we reject the null, understanding that the exper^2
    should not
64  *be ommitted from the regression.
65  *================================================================================================
66
67  *================================================================================================
68  *e) Use the approximation of the marginal effect of experience:
69  *%change in wagei = dlog(wagei)/dexperi = [B_2_ + 2B_3_exper_i_] * 100%
70  *to find the approx return to going from 5 to 6 years of experience. What is the approx
71  *return to going from 20 to 21 years of experience? Do you think experience exhibits
72  *diminishing marginal returns?
73
74  *change from 5 to 6 years of experience:
75  * = [.0410089 + 2(-.0007136 * 6)] - [.0410089 + 2(-.0007136 * 5)]
76  * = .0324457 - .0338729
77  * = -.0014272 * 100%
78  * = -.14272%
79
80  *This means that a one year change in experience from 5 to 6 years of work results in an
81  *overall negative percent change of -.14% from increasing years of experience's effect on
    wage.
82  *This is an example of diminishing marginal returns.
83
84  *change from 20 to 21 years of experience:
85  * = [.0410089 + 2(-.0007136 * 21)] - [.0410089 + 2(-.0007136 * 20)]
86  * = .0110377 - .0124649
87  * = -.0014272 * 100%
88  * = -.14272%
89
90  *This means that a one year change in experience from 20 to 21 years of work results in an
91  *overall negative percent change of -.14% from increasing years of experience's effect on
    wage.
92  *This again exhibits the regression is an example of diminishing marginal returns.
93  *================================================================================================
```



```
94
95  *================================================================================================
96  *f) Find the predicted difference in log(wage) between someone with 10 years of experience
97  *and someone with 20 years of experience. Is the difference statistically significant?
98  *Note: Do not use the approx in (e). We only use approx in (e) when change in x = 1.
99  *In this case, change in x = 10.)
100
101 *for 20 years:
102 * = [.0410089 + 2(-.0007136 * 20)]
103 * = .0124649
104 * = 1.25%
105
106 *for 10 years:
107 * = [.0410089 + 2(-.0007136 * 10)]
108 * = .0267369
109 * = 2.67%
110
111 *change from 10 to 20 years of experience:
112 * = [.0410089 + 2(-.0007136 * 20)] - [.0410089 + 2(-.0007136 * 10)]
113 * = .0124649 - .0267369
114 * = -.014272 * 100%
115 * = -1.4272%
116
117 *At 20 years of experience you get less of an increase in wage than from 10 years.
118
119 *H0 = change in 10 years of experience = 0
120 *H1 = change in 10 years of experience \= 0
121
122 *As this is a log-level model, a 10 unit increase in experience decreases wage by -1.43% on
    average.
123 *We know the difference is statistically significant from parts (c) and (d).
124
125 *================================================================================================
126
127 *================================================================================================
128 *g) At what value of exper does an additional year of experience reduce predicted log(wage)?
```

```
*How many people have experience beyond that turning point in the sample? Do you think the
*turning point is a problem or do you think the results make sense?

tab exper
sum exper

*Turning point for where experience has a diminishing/increasing return on salary:
*-b/2a = -(.0410089)/(2 * -.0007136) = 28.73381446

*[.0410089 + 2(-.0007136 * 5)] * 100 = 3.39%  presumes an increase in log(wage)
*[.0410089 + 2(-.0007136 * 35)] * 100 = -8.94% presumes a decrease in log(wage)

*Beyond 28.7 there are 121 observations from a total of 526. That is 23% of the sample.
*Less than a fourth of the sample reaches the turning point so the scatterplot would be
*left-skewed in its results.

*The turning point does not make sense because it assumes that loyalty to an occupation
only pays
*off after a dedication of nearly 30 years. The correlation between experience and wages
should
*make sense that the more experienced one is at a job there's a certain leveling effect
where their
*experience makes them better equiped to cope with the stressors of the occupation and be a
more
*productive employee. It doesn't typically take an employee 29 years of experience for that
*experience to show itself in wage returns. It does however suggest that if higher paying
jobs
*are offered that it is in the interest in the employee to leave and not give their
*loyalty to one position.

*================================================================================================

*================================================================================================
*h) What if you think that there should not be a turning point for experience, but you
still think
*there are diminishing marginal returns to experience. How might you change the model?
```



```
(Hint: You
*need to change the functional form of experience.)

*I think the coefficient for the value of exper^2 could be changed to a positive value.
*There is a ceiling to which overcharging based solely on experience is dangerous but I
would
*tend to think that with more experience, especially someone who doesn't tire of their
chosen
*field and is solely first and foremost an expert on it, there should be reference point on
whether
*the regression is representative of the population of a country or whether it refers to an
employee's
*wages within a firm. Because it could be the case that working for a firm over too-long a
period may lead
*an employee to grow stagnant and not warrant their increase in wage due to having reached
their heights
*at a particular firm. Maybe it should be a consideration for highly skilled employees as a
marker
*of recognizing that maybe they could bring more value to a firm by taking a more
free-lance, contract
*work in addition their current employment - build an online course to generate passive
income to
*make the highest use of their time in a changing economy.

*On the whole of the population I don't believe people prefer to be workaholics so there is
a ceiling to
*where the squared variable may very well be reflective on the population of a community as
a whole.
*However in terms of economic opportunity I would like to think that in term of someone's
potential
*for earnings the negative variable does not reflect that possibility.
*================================================================================================
```

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis             StataCorp
                                       4905 Lakeway Drive
                                       College Station, Texas 77845 USA
                                       800-STATA-PC        http://www.stata.com
                                       979-696-4600        stata@stata.com
                                       979-696-4601 (fax)

Single-user Stata license expires 20 Aug 2018:
       Serial number:  301509284846
         Licensed to:  Monica Elgawly
                       Hunter College

Notes:
      1.  Unicode is supported; see help unicode_advice.

1 . use "/Users/monicaelgawly/Downloads/wage1.dta"

2 . log using "/Applications/mult2_answers_monica.log"
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Applications/mult2_answers_monica.log
  log type:  text
 opened on:   5 Apr 2018, 21:33:13

3 . ge lwage = log(wage)
variable lwage already defined
r(110);

4 . twoway scatter lwage age
variable age not found
```



```
r(111);

5 . gen exper2 = exper^2

6 . regress lwage educ exper exper2
```

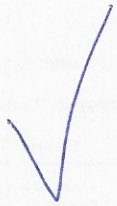
Source	SS	df	MS	Number of obs	=	526
				F(3, 522)	=	74.67
Model	44.5393713	3	14.8464571	Prob > F	=	0.0000
Residual	103.79038	522	.198832146	R-squared	=	0.3003
				Adj R-squared	=	0.2963
Total	148.329751	525	.28253286	Root MSE	=	.44591

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0903658	.007468	12.10	0.000	.0756948	.1050368
exper	.0410089	.0051965	7.89	0.000	.0308002	.0512175
exper2	-.0007136	.0001158	-6.16	0.000	-.000941	-.0004861
_cons	.1279975	.1059323	1.21	0.227	-.0801085	.3361035

```
7 . range exper
invalid syntax
r(198);

8 . detail exper
command detail is unrecognized
r(199);

9 . tab exper

       years |
   potential |
```

experience	Freq.	Percent	Cum.
1	27	5.13	5.13
2	35	6.65	11.79
3	35	6.65	18.44
4	15	2.85	21.29
5	24	4.56	25.86
6	18	3.42	29.28
7	19	3.61	32.89
8	15	2.85	35.74
9	18	3.42	39.16
10	14	2.66	41.83
11	21	3.99	45.82
12	6	1.14	46.96
13	16	3.04	50.00
14	20	3.80	53.80
15	15	2.85	56.65
16	8	1.52	58.17
17	9	1.71	59.89
18	9	1.71	61.60
19	11	2.09	63.69
20	6	1.14	64.83
21	8	1.52	66.35
22	12	2.28	68.63
23	12	2.28	70.91
24	6	1.14	72.05
25	6	1.14	73.19
26	11	2.09	75.29
27	3	0.57	75.86
28	6	1.14	77.00
29	10	1.90	78.90
30	3	0.57	79.47
31	10	1.90	81.37
32	1	0.19	81.56

33	8	1.52	83.08
34	8	1.52	84.60
35	9	1.71	86.31
36	10	1.90	88.21
37	6	1.14	89.35
38	5	0.95	90.30
39	6	1.14	91.44
40	5	0.95	92.40
41	5	0.95	93.35
42	5	0.95	94.30
43	4	0.76	95.06
44	4	0.76	95.82
45	6	1.14	96.96
46	1	0.19	97.15
47	6	1.14	98.29
48	3	0.57	98.86
49	4	0.76	99.62
50	1	0.19	99.81
51	1	0.19	100.00
Total	526	100.00	



```
10 . sum exper
```

Variable	Obs	Mean	Std. Dev.	Min	Max
exper	526	17.01711	13.57216	1	51

11 . log close

Assignment 3: Multiple and Nonlinear Regression Models

ECO 321

DUE: Thursday, April 12 in class

Instructions: You need to write a Stata do-file to answer these questions. Your do-file must produce a log file that contains your Stata output. You must submit your log file along with your answers to the assignment. Assignments that do not contain log files will be marked down.

1. In this question, we are interested in the determinants of years of education. Use the data set HTV.dta for this question. The data set includes information on education, ability, parents' education, and several other variables for men in 1991.

a) Load the data set in Stata (via the `use` command). First, use the `sum, detail` command to show the range of values that the education variable takes in the sample. Second, use the `gen` command to create a dummy variable that equals 1 if men completed exactly 12 years of education and equals 0 otherwise. Third, use the `tab` command to show what percentage of men completed exactly 12 years of education. Fourth, use the `sum` command to determine whether the men or their parents have higher levels of education.

b) Use the `reg` command to estimate the regression model:

$$educ_i = \beta_0 + \beta_1 motheduc_i + \beta_2 fatheduc_i + u_i$$

Make sure to use heteroskedasticity-robust standard errors. How much of the variation in *educ* is explained by parents' education? Interpret the coefficient on *motheduc*.

c) Add the variable *abil* (a measure of cognitive ability) to the model in part (b). Does ability help to explain variation in education, even after controlling for parents' education? Use at least one test statistic to justify your answer.

d) Compare the coefficient estimates on *motheduc* and *fatheduc* from the model in part (b) to the model in part (c). What do the differences tell us about omitted variable bias in the model in part (b)? Be specific.

e) Use the results from part (c) to test the hypothesis that the effect of mother's education on men's education equals two times the effect of father's education on men's education. First, write down the null hypothesis and the alternative hypothesis. Second, evaluate the hypothesis using the F-statistic and the `test` command. Third, rearrange the regression model and evaluate the hypothesis using a t-statistic. Do you arrive at the same conclusion using the F- and t-statistics?

2. Next we are interested in the determinants of wages. Use the data set wage1.dta to answer this question.

a) Use OLS to estimate the equation:

$$log(wage_i) = \beta_0 + \beta_1 edu_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

Write the regression results in standard form (i.e., plug in the correct values for the parameters estimates and their standard errors):

$$\widehat{log(wage_i)} = \underset{(SE(\hat{\beta}_0))}{\hat{\beta}_0} + \underset{(SE(\hat{\beta}_1))}{\hat{\beta}_1} educ_i + \underset{(SE(\hat{\beta}_2))}{\hat{\beta}_2} exper_i + \underset{(SE(\hat{\beta}_3))}{\hat{\beta}_3} exper_i^2$$

b) Interpret β_1 and test its statistical significance at the 1% level. What can we say about the relationship between education and $log(wage)$?

c) Test whether experience has a statistically significant effect on $log(wage)$.

d) Is $exper^2$ statistically significant at the 5% level? What does this hypothesis test say about the relationship between $log(wage)$ and experience?

e) Use the approximation of the marginal effect of experience:

$$\%\Delta\hat{wage}_i = \frac{dlog(wage_i)}{dexper_i} = (\hat{\beta}_2 + 2\hat{\beta}_3 exper_i) \times 100\%$$

to find the approximate return to going from 5 to 6 years of experience. What is the approximate return to going from 20 to 21 years of experience? Do you think experience exhibits diminishing marginal returns?

f) Find the predicted difference in $log(wage)$ between someone with 10 years of experience and someone with 20 years of experience. Is the difference statistically significant? (Note: Do not use the approximation from part (e). We only use the approximation in part (e) when $\Delta X = 1$. In this case, $\Delta X = 10$.)

g) At what value of *exper* does an additional year of experience reduce predicted $log(wage)$? How many people have experience beyond that turning point in the sample? Do you think the turning point is a problem or do you think the results make sense?

h) What if you think that there should not be a turning point for experience, but you still think there are diminishing marginal returns to experience. How might you change the model? (Hint: You need to change the functional form of experience.)