Homework #6

A. Shawn Bandy

March 25th, 2013

1 Lab Problems

L1 STATA log:

```
--------------------------------------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\cla-spa206.CAMPUS-DOMAIN\Downloads\lab6\lab6demo.log
  log type:  text
 opened on:  21 Mar 2013, 11:43:46

. /* L1. The dataset wage_guar.dta contains data on 1,289 individuals interviewed
> in March 1995 for the Current Population Survey (CPS) by the U.S. Census Bureau
> This question looks at the determinants of their wages. Create a do file
>  wage95cps.do and a log file wage95cpsdlog for this problem. Submit the log file
> with your answers.
> a. Create the following variables and then save data as wage95cps.dta
> lnwage : defined as the log(wage)
> exp_squared : defined as exper squared
> college : defined as having education of 16 years or above
> some_college : defined as having education of 12 or above but less than 16
> no_college : defined as having education of less than 12
> female_exp : interact the female and the experience variables
> lnexp : defined as the log(exper)
> annual_wage : defined as the annual wage if the person works full time at his/her
> hourly wage. So, 8 hours a day, 1289 days a year.*/
>
> gen lnwage = ln(wage);

. gen exp_squared = exper^2;

. gen college = 0;

. gen some_college = 0;

. gen nocollege = 0;

. replace college = 1 if education >= 16;
```

```
(349 real changes made)

. replace some_college = 1 if (education > 12 & education <16);
(104 real changes made)

. replace  nocollege = 1 if education <= 12;
(836 real changes made)

. gen female_exp  = female * exper;

. gen lnexp = ln(exper);
(14 missing values generated)

. gen annual_wage = wage * 8 * 300;

. /*b. Run the following regressions to look at the determinants of wages. After
> you estimate the first model (i), use the Breusch-Pagan test to test for
>  heteroskedasticity. If you find that the errors are heteroskedastic, use robust
>  standard errors to correct for heteroskedasticity by re-estimating that model and
> then go ahead and use robust standard errors for the rest of the models.
> i. Dependent variable: wage; explanatory variables are female nonwhite union education exper
> ii. Dependent variable: annual_wage; explanatory variables are female nonwhite union education exper
> iii. Dependent variable: lnwage; explanatory variables are female nonwhite union education exper
> iv. Dependent variable: lnwage; explanatory variables are female nonwhite union education lnexp
> v. Dependent variable: lnwage; explanatory variables are female nonwhite union education exper exp_square
> vi. Dependent variable: lnwage; explanatory variables are female nonwhite union college some_college expe
>
> regress wage female nonwhite union education exper;

      Source |       SS       df       MS              Number of obs =    1289
-------------+------------------------------           F(  5,  1283) =  122.61
       Model |  25967.2805     5  5193.45611           Prob > F      =  0.0000
    Residual |  54342.5442  1283  42.3558411           R-squared     =  0.3233
-------------+------------------------------           Adj R-squared =  0.3207
       Total |  80309.8247  1288  62.3523484           Root MSE      =  6.5081

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -3.074875   .3646162    -8.43   0.000    -3.790185   -2.359566
    nonwhite |  -1.565313   .5091875    -3.07   0.002    -2.564245   -.5663817
       union |   1.095976   .5060781     2.17   0.031     .1031443    2.088807
   education |   1.370301   .0659042    20.79   0.000     1.241009    1.499593
       exper |   .1666065   .0160476    10.38   0.000     .1351242    .1980889
       _cons |  -7.183338   1.015788    -7.07   0.000    -9.176126   -5.190551
------------------------------------------------------------------------------

. estat hettest;
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of wage

         chi2(1)      =   196.22
         Prob > chi2  =   0.0000

. regress wage female nonwhite union education exper,r;

Linear regression                                      Number of obs =    1289
                                                       F(  5,  1283) =  100.87
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3233
                                                       Root MSE      =  6.5081

------------------------------------------------------------------------------
             |               Robust
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -3.074875   .3642561    -8.44   0.000    -3.789478   -2.360272
    nonwhite |  -1.565313    .397626    -3.94   0.000    -2.345382   -.7852448
       union |   1.095976   .4258023     2.57   0.010     .2606306    1.931321
   education |   1.370301   .0834851    16.41   0.000     1.206519    1.534083
       exper |   .1666065   .0160487    10.38   0.000      .135122     .198091
       _cons |  -7.183338   1.090064    -6.59   0.000    -9.321841   -5.044835
------------------------------------------------------------------------------

. regress annual_wage female nonwhite union education exper exp_squared;

      Source |       SS       df       MS              Number of obs =    1289
-------------+------------------------------           F(  6,  1282) =  108.35
       Model |  1.5565e+11     6  2.5942e+10           Prob > F      =  0.0000
    Residual |  3.0693e+11  1282   239417084           R-squared     =  0.3365
-------------+------------------------------           Adj R-squared =  0.3334
       Total |  4.6258e+11  1288   359149529           Root MSE      =   15473

------------------------------------------------------------------------------
 annual_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -7222.464   867.4371    -8.33   0.000    -8924.216   -5520.712
    nonwhite |  -3686.584   1210.675    -3.05   0.002    -6061.705   -1311.463
       union |   2464.749    1203.65     2.05   0.041      103.408    4826.089
   education |   3176.989   158.2483    20.08   0.000     2866.535    3487.443
       exper |    1018.71   128.5913     7.92   0.000     766.4377    1270.983
 exp_squared |  -14.83991   2.944722    -5.04   0.000    -20.61691   -9.062904
       _cons |  -20205.68   2485.704    -8.13   0.000    -25082.18   -15329.19
------------------------------------------------------------------------------

. estat hettest;
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of annual_wage

         chi2(1)      =   193.41
         Prob > chi2  =   0.0000

. regress annual_wage female nonwhite union education exper exp_squared,r;

Linear regression                                      Number of obs =    1289
                                                       F(  6,  1282) =   94.50
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3365
                                                       Root MSE      =   15473

------------------------------------------------------------------------------
             |               Robust
 annual_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -7222.464   870.2926    -8.30   0.000    -8929.818    -5515.11
    nonwhite |  -3686.584   946.2218    -3.90   0.000    -5542.897   -1830.271
       union |   2464.749    1008.42     2.44   0.015     486.4139    4443.084
   education |   3176.989   200.1992    15.87   0.000     2784.235    3569.743
       exper |    1018.71   121.3314     8.40   0.000     780.6804     1256.74
 exp_squared |  -14.83991   3.004312    -4.94   0.000    -20.73382   -8.945999
       _cons |  -20205.68   2684.516    -7.53   0.000    -25472.21   -14939.16
------------------------------------------------------------------------------

. regress lnwage  nonwhite union education exper;

      Source |       SS       df       MS              Number of obs =    1289
-------------+------------------------------           F(  4,  1284) =  138.22
       Model |  133.287008     4   33.321752           Prob > F      =  0.0000
    Residual |  309.544069  1284  .241077935           R-squared     =  0.3010
-------------+------------------------------           Adj R-squared =  0.2988
       Total |  442.831077  1288  .343812948           Root MSE      =    .491

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nonwhite |   -.150007   .0383718    -3.91   0.000    -.2252854   -.0747286
       union |   .2111055   .0380276     5.55   0.000     .1365025    .2857086
   education |   .1011594     .00497    20.35   0.000     .0914091    .1109096
       exper |   .0128887   .0012106    10.65   0.000     .0105137    .0152637
       _cons |   .7598444   .0749283    10.14   0.000     .6128491    .9068397
------------------------------------------------------------------------------

. estat hettest;
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =    10.81
         Prob > chi2  =   0.0010

. regress lnwage  nonwhite union education exper,r;

Linear regression                                      Number of obs =    1289
                                                       F(  4,  1284) =  159.80
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3010
                                                       Root MSE      =    .491

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nonwhite |   -.150007   .0351519    -4.27   0.000    -.2189684   -.0810456
       union |   .2111055   .0315819     6.68   0.000     .1491477    .2730634
   education |   .1011594   .0053516    18.90   0.000     .0906606    .1116581
       exper |   .0128887   .0012716    10.14   0.000      .010394    .0153835
       _cons |   .7598444   .0753165    10.09   0.000     .6120874    .9076013
------------------------------------------------------------------------------

. regress lnwage female nonwhite education lnexp;

      Source |       SS       df       MS              Number of obs =    1275
-------------+------------------------------           F(  4,  1270) =  176.68
       Model |  157.047841     4  39.2619603           Prob > F      =  0.0000
    Residual |  282.215421  1270  .222216867           R-squared     =  0.3575
-------------+------------------------------           Adj R-squared =  0.3555
       Total |  439.263262  1274  .344790629           Root MSE      =   .4714

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2533954   .0264672    -9.57   0.000    -.3053197   -.2014711
    nonwhite |  -.1200509   .0368087    -3.26   0.001    -.1922635   -.0478383
   education |   .0978298   .0047354    20.66   0.000     .0885398    .1071198
       lnexp |   .2111839   .0154079    13.71   0.000     .1809561    .2414117
       _cons |   .6409581   .0819226     7.82   0.000     .4802395    .8016766
------------------------------------------------------------------------------

. estat hettest;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =    19.61
         Prob > chi2  =   0.0000

. regress lnwage female nonwhite education lnexp,r;

Linear regression                                      Number of obs =    1275
                                                       F(  4,  1270) =  192.33
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3575
                                                       Root MSE      =   .4714

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2533954   .0265531    -9.54   0.000    -.3054881   -.2013027
    nonwhite |  -.1200509   .0352988    -3.40   0.001    -.1893012   -.0508006
   education |   .0978298   .0050452    19.39   0.000      .087932    .1077276
       lnexp |   .2111839   .0137559    15.35   0.000     .1841972    .2381706
       _cons |   .6409581   .0760409     8.43   0.000     .4917784    .7901378
------------------------------------------------------------------------------

. regress lnwage female nonwhite union education exper exp_squared;

      Source |       SS       df       MS              Number of obs =    1289
-------------+------------------------------           F(  6,  1282) =  125.77
       Model |   164.07853     6  27.3464217           Prob > F      =  0.0000
    Residual |  278.752547  1282  .217435684           R-squared     =  0.3705
-------------+------------------------------           Adj R-squared =  0.3676
       Total |  442.831077  1288  .343812948           Root MSE      =   .4663

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.242462   .0261412    -9.28   0.000    -.2937463   -.1911777
    nonwhite |  -.1305487   .0364851    -3.58   0.000    -.2021258   -.0589717
       union |   .1731559   .0362734     4.77   0.000     .1019941    .2443176
   education |   .0951149    .004769    19.94   0.000      .085759    .1044708
       exper |   .0390986   .0038752    10.09   0.000     .0314961    .0467012
 exp_squared |  -.0006316   .0000887    -7.12   0.000    -.0008057   -.0004575
       _cons |   .7792842   .0749096    10.40   0.000     .6323254    .9262429
------------------------------------------------------------------------------

. estat hettest;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =     8.56
         Prob > chi2  =   0.0034

. regress lnwage female nonwhite union education exper exp_squared,r;

Linear regression                                      Number of obs =    1289
                                                       F(  6,  1282) =  145.17
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3705
                                                       Root MSE      =   .4663

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.242462    .026376    -9.19   0.000    -.2942069   -.1907171
    nonwhite |  -.1305487   .0343429    -3.80   0.000    -.1979231   -.0631744
       union |   .1731559   .0296575     5.84   0.000     .1149732    .2313385
   education |   .0951149   .0051261    18.56   0.000     .0850585    .1051713
       exper |   .0390986   .0038906    10.05   0.000      .031466    .0467313
 exp_squared |  -.0006316    .000097    -6.51   0.000    -.0008219   -.0004412
       _cons |   .7792842   .0749449    10.40   0.000     .6322561    .9263123
------------------------------------------------------------------------------

. regress lnwage female nonwhite union college some_college exper;

      Source |       SS       df       MS              Number of obs =    1289
-------------+------------------------------           F(  6,  1282) =   93.97
       Model |  135.261411     6  22.5435684           Prob > F      =  0.0000
    Residual |  307.569666  1282  .239913936           R-squared     =  0.3054
-------------+------------------------------           Adj R-squared =  0.3022
       Total |  442.831077  1288  .343812948           Root MSE      =  .48981

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2451276   .0274773    -8.92   0.000     -.299033   -.1912221
    nonwhite |  -.1462069    .038297    -3.82   0.000    -.2213387   -.0710752
       union |   .1939992   .0380702     5.10   0.000     .1193126    .2686858
     college |   .5682876   .0316352    17.96   0.000     .5062251    .6303501
some_college |   .3140907    .050996     6.16   0.000     .2140459    .4141355
       exper |   .0110464   .0011965     9.23   0.000      .008699    .0133938
       _cons |    2.06904   .0332687    62.19   0.000     2.003773    2.134307
------------------------------------------------------------------------------

. estat hettest;
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =    20.68
         Prob > chi2  =   0.0000

. regress lnwage female nonwhite union college some_college exper,r;

Linear regression                                      Number of obs =    1289
                                                       F(  6,  1282) =  109.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3054
                                                       Root MSE      =  .48981

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2451276   .0274473    -8.93   0.000    -.2989741    -.191281
    nonwhite |  -.1462069   .0372502    -3.92   0.000     -.219285   -.0731289
       union |   .1939992   .0319926     6.06   0.000     .1312356    .2567628
     college |   .5682876   .0340563    16.69   0.000     .5014755    .6350997
some_college |   .3140907   .0475559     6.60   0.000     .2207947    .4073867
       exper |   .0110464   .0012677     8.71   0.000     .0085594    .0135333
       _cons |    2.06904    .029912    69.17   0.000     2.010358    2.127722
------------------------------------------------------------------------------

. /*
> reject none
> c. Run a regression omitting union and nonwhite from model (vi) from part b.
> This is the restricted model that can be used to test if union and nonwhite are
> jointly significant. Save the RSS as rss_r using ereturn list and the scalar
> command.*/
>
> regress lnwage female college some_college exper,r;

Linear regression                                      Number of obs =    1289
                                                       F(  4,  1284) =  129.50
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2853
                                                       Root MSE      =  .49648

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2611761   .0276074    -9.46   0.000    -.3153368   -.2070155
```

```
     college |   .5812343    .033983    17.10   0.000      .514566    .6479026
some_college |   .3205216   .0486417     6.59   0.000     .2250958    .4159475
       exper |   .0122078   .0012532     9.74   0.000     .0097492    .0146664
       _cons |   2.059682   .0296149    69.55   0.000     2.001583     2.11778
------------------------------------------------------------------------------

. scal rss_r = e(rss);

. /*d. Now, run model (vi) from part b again. Save the RSS for this model as rss_u
>  using ereturn list and the scalar command.*/
>
> regress lnwage female nonwhite union college some_college exper,r;

Linear regression                                      Number of obs =    1289
                                                       F(  6,  1282) =  109.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3054
                                                       Root MSE      =  .48981

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2451276   .0274473    -8.93   0.000    -.2989741    -.191281
    nonwhite |  -.1462069   .0372502    -3.92   0.000     -.219285   -.0731289
       union |   .1939992   .0319926     6.06   0.000     .1312356    .2567628
     college |   .5682876   .0340563    16.69   0.000     .5014755    .6350997
some_college |   .3140907   .0475559     6.60   0.000     .2207947    .4073867
       exper |   .0110464   .0012677     8.71   0.000     .0085594    .0135333
       _cons |    2.06904    .029912    69.17   0.000     2.010358    2.127722
------------------------------------------------------------------------------

. scal rss_u = e(rss);

. /*e. Now, calculate the F-statistic manually using the saved scalars from
>  parts c and d the STATA scalar command: sca fstat= Display the f-statistic
>  by typing sca list fstat*/
>
> scal fstat = ((rss_r - rss_u)/2)/(rss_u/1289);

. scal list;
     fstat =  18.717925
     rss_u =  307.56967
     rss_r =  316.50227

.  /*f. Run model (vi) from part b again. Use the STATA command test to test for
> joint significance of union and nonwhite in this model.*/
>
> regress lnwage female nonwhite union college some_college exper,r;
```

```
Linear regression                                      Number of obs =    1289
                                                       F(  6,  1282) =  109.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3054
                                                       Root MSE      =  .48981

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2451276   .0274473    -8.93   0.000    -.2989741    -.191281
    nonwhite |  -.1462069   .0372502    -3.92   0.000     -.219285   -.0731289
       union |   .1939992   .0319926     6.06   0.000     .1312356    .2567628
     college |   .5682876   .0340563    16.69   0.000     .5014755    .6350997
some_college |   .3140907   .0475559     6.60   0.000     .2207947    .4073867
       exper |   .0110464   .0012677     8.71   0.000     .0085594    .0135333
       _cons |    2.06904    .029912    69.17   0.000     2.010358    2.127722
------------------------------------------------------------------------------

. test union=nonwhite;

 ( 1)  - nonwhite + union = 0

       F(  1,  1282) =   45.90
            Prob > F =    0.0000

. /*g. Now, create a special variable for use in testing if the coefficients on
> college and some_college are the same. This variables is all_college =
> college+some_college*/
> gen all_college = 0;

. replace all_college = college+some_college;
(453 real changes made)

. /*h. Run a regression such that: dependent variable: lnwage; explanatory
>  variables are female nonwhite union college all_college exper*/
>
> regress lnwage female nonwhite union college all_college exper;

      Source |       SS       df       MS              Number of obs =    1289
-------------+------------------------------           F(  6,  1282) =   93.97
       Model |  135.261411     6  22.5435684           Prob > F      =  0.0000
    Residual |  307.569666  1282  .239913936           R-squared     =  0.3054
-------------+------------------------------           Adj R-squared =  0.3022
       Total |  442.831077  1288  .343812948           Root MSE      =  .48981

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
      female |  -.2451276   .0274773    -8.92   0.000     -.299033   -.1912221
    nonwhite |  -.1462069    .038297    -3.82   0.000    -.2213387   -.0710752
       union |   .1939992   .0380702     5.10   0.000     .1193126    .2686858
     college |   .2541969   .0548669     4.63   0.000     .1465581    .3618357
 all_college |   .3140907    .050996     6.16   0.000     .2140459    .4141355
       exper |   .0110464   .0011965     9.23   0.000      .008699    .0133938
       _cons |    2.06904   .0332687    62.19   0.000     2.003773    2.134307
------------------------------------------------------------------------------

. /*i. Now, run the model (vi) from part b again. Use the STATA command test to
> test whether the coefficients on college and some_college are the same in this
> model.*/
>
> regress lnwage female nonwhite union college some_college exper,r;

Linear regression                                      Number of obs =    1289
                                                       F(  6,  1282) =  109.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3054
                                                       Root MSE      =  .48981

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2451276   .0274473    -8.93   0.000    -.2989741    -.191281
    nonwhite |  -.1462069   .0372502    -3.92   0.000     -.219285   -.0731289
       union |   .1939992   .0319926     6.06   0.000     .1312356    .2567628
     college |   .5682876   .0340563    16.69   0.000     .5014755    .6350997
some_college |   .3140907   .0475559     6.60   0.000     .2207947    .4073867
       exper |   .0110464   .0012677     8.71   0.000     .0085594    .0135333
       _cons |    2.06904    .029912    69.17   0.000     2.010358    2.127722
------------------------------------------------------------------------------

. test college = some_college;

 ( 1)  college - some_college = 0

       F(  1,  1282) =   21.97
            Prob > F =    0.0000

. /*j. Test for multicollinearity of the model (vi) from part b by using the vif
> command.*/
>
> vif;

    Variable |       VIF       1/VIF
-------------+----------------------
```

```
     college |      1.06    0.941918
       exper |      1.05    0.956468
       union |      1.04    0.960187
some_college |      1.04    0.964904
    nonwhite |      1.02    0.980143
      female |      1.01    0.986111
-------------+----------------------
    Mean VIF |      1.04

. log close;
      name:  <unnamed>
       log:  C:\Users\cla-spa206.CAMPUS-DOMAIN\Downloads\lab6\lab6demo.log
  log type:  text
 closed on:  21 Mar 2013, 11:43:46
----------------------------------------------------------------------------------------------------
```

2 Questions

Q1 Explain (giving one or two reasons) why we might want to transform a variable using log (Note: This is natural log).

Although there are no well defined rules for using the log of a variable, we would consider using it when we suspect that the relationship between one or more dependent variables and the independent variables is not strictly linear. One case that begs for logarithmic linearization are economic growth models such as the Cobb-Douglas production function: $Q_i = AL^{\beta}K^{\alpha}$. Whatever the reason, it is important that any variable that is log transformed is only positive as the logarithm of a negative number is undefined.

Q2 Use your results from L1 for this problem.

a. Is there a problem with heteroskedasticity? Explain how you know.

Yes, heteroskedasticity was evident at the 95% confidence level for each model. When *estate hettest* was run for each, the p-value is less than 0.05 which means that we should reject the null hypothesis which is that the data is homoskedastic.

b. For each of the models below (which correspond to the models from part b of L1, interpret the coefficients identified here.

i. Dependent variable: wage; explanatory variables are female nonwhite union education exper. Interpret the coefficient on exper

For each unit increase of exper, there is a .1666065 increase in wage, ceteris paribus.

ii. Dependent variable: annual_wage; explanatory variables are female nonwhite union education exper. Interpret the coefficient on exper

There is an error in my regression I mistakenly included exp_squared in this model so the following value of the coefficient is incorrect. For each unit increase of exper, there is a 1018.71 increase in annual_wage, ceteris paribus.

iii. Dependent variable: lnwage; explanatory variables are female nonwhite union education exper Interpret the coefficient on exper

For each unit increase in exper, there is a 1.289% increase in wage, ceteris paribus.

iv. Dependent variable: lnwage; explanatory variables are female nonwhite union education lnexp. Interpret the coefficient on lnexp

For each percent increase in exper, there is a .2112% increase in wage, ceteris paribus. This is the wage elasticity with respect to exper.

v. Dependent variable: lnwage; explanatory variables are female nonwhite union education exper exp_squared - Interpret the effect of experience on wages, using exper and exp_squared

For each unit of exper, there is a 3.9% increase in wage. Also, for each additional unit of exper, there is 0.06% drop in wage - this signifies a decreased return to wages with respect to experience.

vi. Dependent variable: lnwage; explanatory variables are female nonwhite union college some_college exper - Interpret the effect of having at least 4 years of college on wages
For those who have some_college but have not graduated, we expect their wage to be 3.14% above those with only high school, ceteris paribus.

c. Why did we leave no_college out of the final model (estimated in L1, parts b, vi)?
If all three categorical variables are included then the model would have perfect collinearity. Stata will general omit one of the variables with warnings but it is generally better to choose one ourselves.

d. In L1, parts c, d, and e you used STATA to calculate an F-statistic. Write out the formula for that F-statistic here. What is the null hypothesis for this F-test? If the critical F-value with 2, 1282 degrees of freedom at the 5% level is 3.00, would you reject the null hypothesis? What is your conclusion here?
In this section of the lab we are testing for the joint significance of union and non-white. Our formula was $\frac{\frac{rss_r - rss_u}{2}}{\frac{rss_u}{1289}}$ where rss_r is the residuals sum-of-squares for the restricted model and rss_u is the residual sum-of-squares from the unrestricted model. The H_0 is that the variables are not jointly significant (taken together their coefficients are not different from zero). We would reject the null hypothesis at the 5% significance level and conclude that the variables are jointly significant.

e. Is the test in L1 part f the same as the test in part e? What is your conclusion based on the result in part f?

   We are testing for the same phenomenon as part f as in part e and we receive the same result in both as well: we can reject the null hypothesis and assume that there is joint significance between union and non-white.

f. In L1 part g and h, you created a new variable and ran a regression on that new variable. What is the null hypothesis you are testing? Based on that hypothesis, demonstrate (show) how we transform our original equation (L1 part b, vi) to the one in L1 part h.

   The H_0 is the coefficients for college and some_college are the same. We take the two variables in L1.b.vi, some_college and college and add the together to create a new variable called all_college. We use the new variable in a regression with either some_college or college. In this case our regression used college. The t-stat for the coefficient of the new variable can be used to test our null hypothesis.

g. Based on the results from L1 part h, what is your conclusion about the hypothesis you are testing (in other words, the hypothesis you outlined in Q2, part d)?

   Based on the results from L1.h, we can reject the null hypothesis and assume that the variables are different from each other at the 5% significance level.

h. What does the test in L1 part i tell us? Why couldnt we calculate this particular test using scalars?

   L1.i tells us the same thing with different numbers. We may reject the null hypothesis at the 5% significance level. I am not sure what the second question is asking.

i. Is there a problem with multicollinearity in the model in L1 part b iv? How do you know?

   The model in L1.b.iv is reasonably free of multicollinearity (all the values in the VIF output are less than 5).