PUBPOL 639: ASSIGNMENT 5 - Solutions
Winter 2011

Due: Monday, April 11th at the start of class

NOTE: This assignment explores various transformations that let you estimate non-linear relationships using OLS. Be sure to copy your do and log files into the back of your solutions. Courier 8 or 9 pt font works well. Your log file may be long, so don't worry about cleaning/formatting it.

In 1995, the newly elected government (headed by Nelson Mandela) has identified racial differences in earnings as a major concern. The administration believes that education is the key to reducing observed differences in pay. Prior to implementing new policy, the South African Labor and Development Unit (SALDRU) is hired to asses the current value of education in the labor market. Your task as one of the primary researchers in the unit is to complete the preliminary analysis below using the 1994 October Household Survey Data (ohs94.dta).

Final Notes:

- After completing the analysis, be sure to create two tables to present your regression output (one for each part)
- In running the models below, use the following omitted categories throughout:
    - Race => blacks — Union => non-union member
    - Gender => females — Location => rural

PART I

1. Estimate a model relating monthly income to education (continuous), age, race, and gender.

    a. Interpret the coefficient on education

A one year increase in school completed is associated with a 182.08 rand increase in monthly income, holding age, race, and gender constant.

    b. Interpret the coefficients on the race indicator variables

Coloured respondents earn 166.11 rand more than black respondents on average controlling for age, education, and gender.

Indian respondents earn 316.37 rand more than black respondents on average controlling for age, education, and gender.

White respondents ear 1041.62 rand more than black respondents on average controlling for age, education, and gender.

2. Now run the same regression, but including a quadratic age variable.

    a. Holding race, gender, and education constant, what is the predicted change in income associated with a one-year increase in age for a 34 year old respondent?

A one-year increase in age for a 34 year old respondent is associated with a 43.77 rand increase in average monthly income. (See do-file for calculation)

b. Holding race, gender, and education constant, what is the predicted change in income associated with a one-year increase in age for a 54 year old respondent?

A one-year increase in age for a 54 year old respondent is associated with a 1.95 rand decrease in average monthly income. (See do-file for calculation)

c. Can we reject the linear model from question 1 in favor of this quadratic model? Explain.

To test whether the quadratic or linear specification fits better we need to test the null hypothesis that the coefficient on the age squared term is equal to zero. Looking at the p-value associated with the age squared term, it is well below an alpha level of .01, which allows for the rejection of the null hypothesis. This suggests that we are confident that a non-linear relationship does exist.

3. To simplify the task of interpreting quadratic terms, write a program (called "coef2") that will provide you with the information found in question 2, parts (a) and (b). At the very least, the program should work for the model estimated in question 2, but if possible, it should work with any model estimated with any number of independent variables. Once the program is completed, confirm your answers to question 2 parts (a) and (b) using the program.

(See do-file & log file for details)

4. Now estimate a model relating income to education (continuous), race, and interactions between education and race.

   a. Which racial group receives the largest return for their educational attainment?

*Looking at the interaction terms, white respondents receive the largest return from an additional year of education as the white*education coefficient is the largest positive value observed relative to the other interactions.*

   b. Calculate the change in predicted income associated with a one year increase in education for each racial group, holding race constant.

For black respondents, an additional year of education is associated with a 127.41 rand increase in monthly income.

*For coloured respondents, an additional year of education is associated with a 150.51 rand increase in monthly income. This value is found summing the coefficient of education and the coloured*education interaction coefficient.*

*For indian respondents, an additional year of education is associated with a 186.03 rand increase in monthly income. This value is found summing the coefficient of education and the indian*education interaction coefficient.*

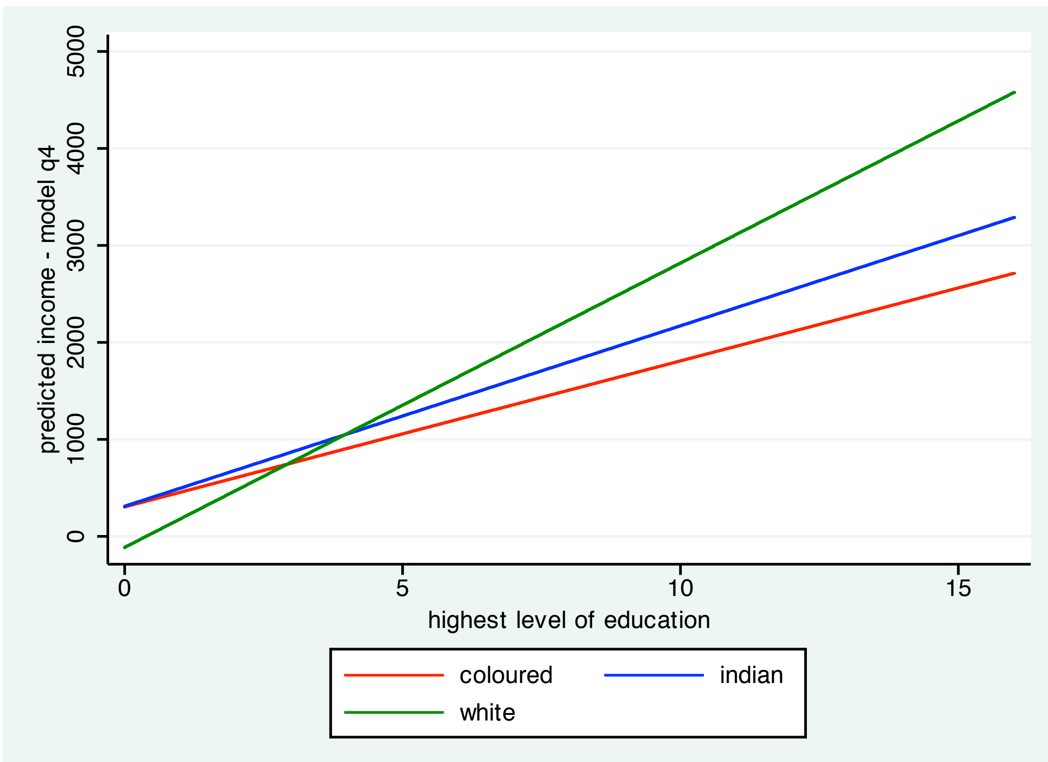
*For white respondents, an additional year of education is associated with a 293.15 rand increase in monthly income. This value is found summing the coefficient of education and the white*education interaction coefficient.*

(See do-file for calculation)

c. Check your calculations by plotting the predicted (fitted) relationship between income and education for each racial group. To do this, you should first generate a variable called pinc (predicted income) by typing "predict pinc" after your regression. You can then plot this predicted value by education separately by racial group by typing:

```
twoway line pinc educ if race == 2, lc(red) ||
       line pinc educ if race == 3, lc(blue) ||
       line pinc educ if race == 4, lc(green) ||,
       legend(lab(1 "coloured") lab(2 "indian") lab(3 "white")) ;
```

Note: You may need to sort the data by education first depending on version of Stata, otherwise can get crazy graph.



PART II

Prior to extending your analysis in Part I, the head researcher at SALDRU suggests you use a log transform on income given the distribution is positively skewed. Using this suggestion, complete the items below.

5. Now regress log(income) on education (continuous), race, and gender.

   a. Interpret the coefficient on education

A one year increase in school completed is associated with a 13.85% increase in monthly income, holding race and gender constant. [Important to note here that percentage change is calculated using exponentiation given level is great than 10% - (See do-file for calculations)]

b. Interpret the coefficients on the race indicator variables

Coloured respondents earn 17.25% percent more than black respondents on average controlling for education and gender.

Indian respondents earn 41.36% percent more than black respondents on average controlling for education and gender.

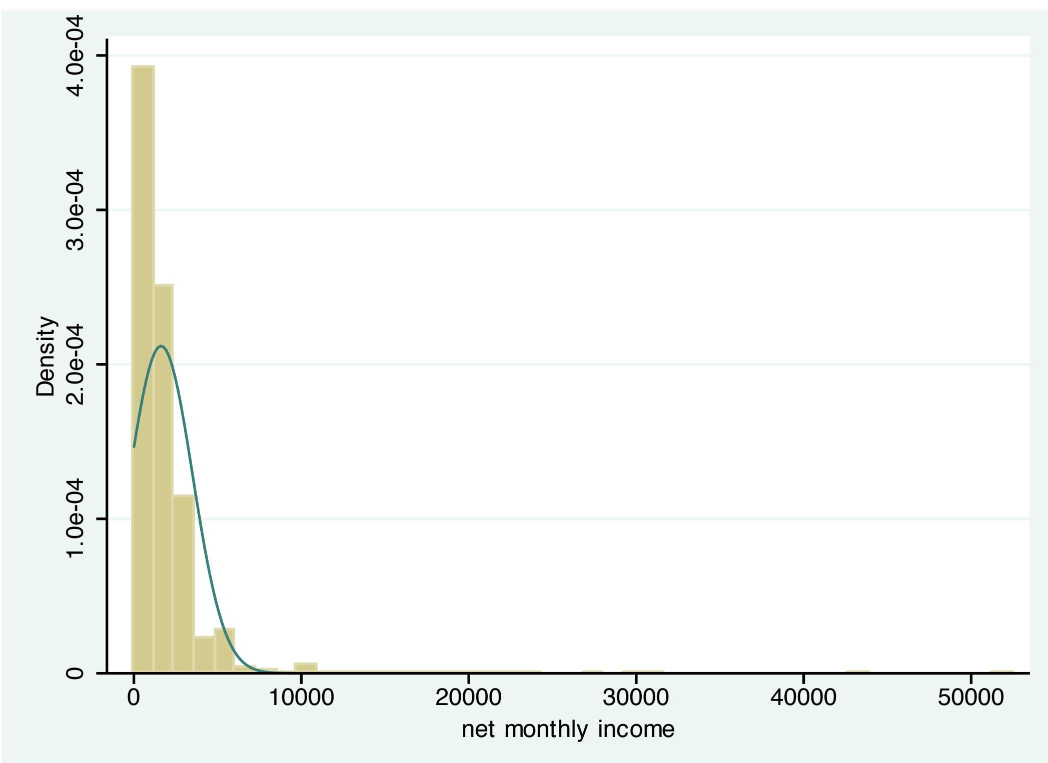
White respondents earn 73.27% percent more than black respondents on average controlling for education and gender.

[Important to note here that percentage difference is calculated using exponentiation given level is great than 10% - (See do-file for calculations)]

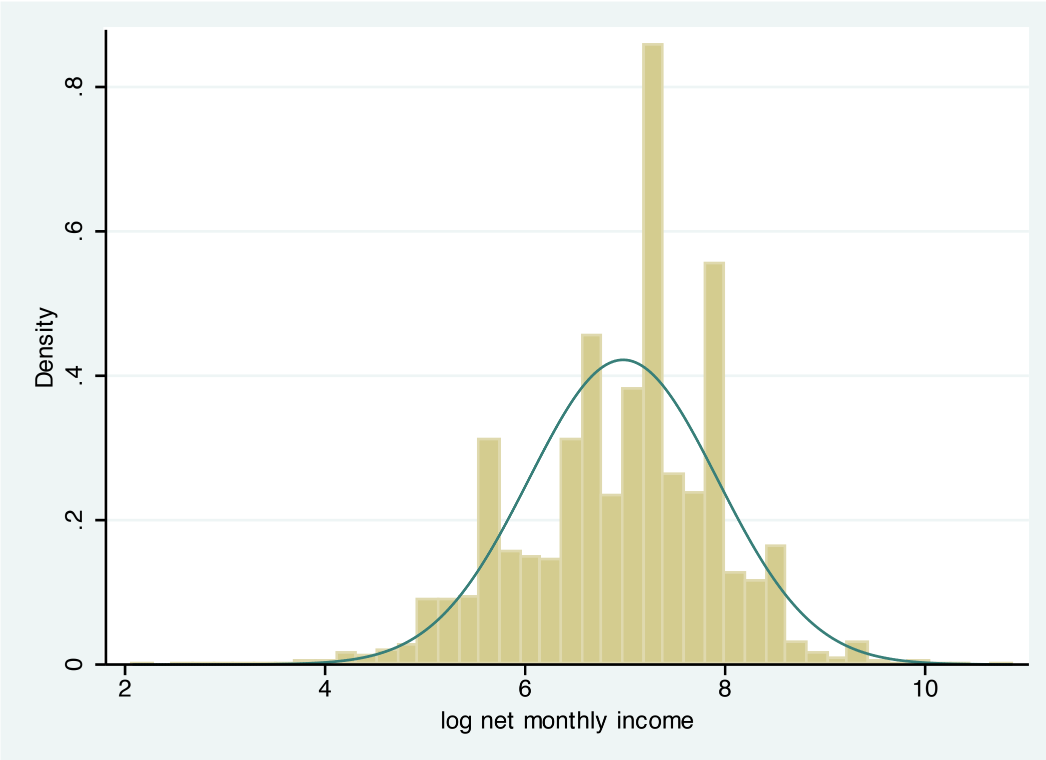
c. Does the transformation on income seem appropriate given the data and your findings? Explain.

Yes, the transformation is appropriate given the highly positively skewed distribution of monthly income. The transformation will normalize the distribution.

Original Distribution



Transformed Distribution



6. Now regress log(income) on education (continuous), age, race, and gender.

   a. Interpret the coefficient on education

A one year increase in school completed is associated with a 14.48% increase in monthly income, holding race, age, and gender constant. [Important to note here that percentage change is calculated using exponentiation given level is great than 10% - (See do-file for calculations)]

   b. Why has the coefficient on education changed from the regression in question 5? Explain in terms of the framework used in class.

The coefficient on education increases from .129 in model Q5 to .135 in model Q6. Omitting age results in a negative omitted variable bias on the education coefficient in model Q5.

Indeed, given the sign of the coefficient in model Q6, age is positively correlated with log income holding race, gender, and education constant. Additionally, given the history of the country, young and middle age adults tend to have higher levels of education than older adults producing a negative relationship between age and education. This combination of the positive and negative associations leads to our negative bias.

7. Now regress log(income) on education (continuous), age, age-squared, race, gender, union membership, location (urban/rural), and interactions between education and race.

   a. Using the model specified, in addition to previous results, summarize the general overall findings.

In general it is quite clear that racial differences in average income are substantial even after controlling for numerous independent factors. Relative to blacks, all other racial groups are earning at least 15% more on average, with whites earning 134% more. Thus, even when education levels are controlled (along with numerous additional factors) large differences in income are still observed by race.

While racial differences dominate the substantive findings, other distinctions in workers matter as well. As expected, education has a strong positive relationship with earnings. Additionally, it appears that the returns to education are fairly similar for the different racial groups (with the findings actually suggesting that blacks receive a slightly larger return than any other group). Gender differences are also quite large, with male workers earning 32% more female workers controlling for all other factors. Similar differences are observed for unionization and location as well.

b. Discuss how the findings inform the administration's assumption (listed below). Additionally, provide the administration with your recommendation on how to best reduce racial differences in earnings.

<u>Assumption</u>: The administration assumes that increasing educational attainment will reduce racial differences in earnings.

It is clear why the administration may have made this assumption given that Indian and white respondents have almost twice the years of school completed as black and coloured workers on average. Although looking at the regression analysis, even when education is controlled for large racial differences are still observed. This suggests that even if education levels are similar, differences in earnings would still persist. If we are confident that model Q7 specification is adequate, then the administration may need to consider the behavior of employers as a possible cause for the observed differences in earnings between the racial groups.

8. Using the existing data, are there improvements that could be made to the "final" model specified in question 7? If so, describe any changes you believe would improve the model.

There are several possible changes to model Q7 that could potentially improve the specification. See list below.

- *use categorical measurement of education -> it seems the response to education in the labor market would be better capture looking at degree thresholds vs. a constant return to an additional year of schooling*
- *add province indicators -> any regional difference in labor markets is lost in model Q7, including provincial indicator variables could possibly capture these differences*
- *interactions with gender -> there are large gender differences in many of the distributions we are working with, would be useful to look at interactions with several of the other regressors including: race, union membership, location, education, and age*
- *interact everything with race -> given the role of race in this country it may be the case that each of our regressors varies by race, instead of constructing interactions with every other independent variable, could run models separately for each racial group (may make sense to do the same for gender as well)*

9. Describe how any data limitations impact the analysis in questions 1-7. Be as specific as possible. [Note: I would like you to briefly discuss possible variables omitted from the data file, but a majority of the discussion should focus on measurement and operationalization of variables in the data file]

The OHS data is a bit limited in terms of the information collected. Ideally, would have been beneficial to include additional variables that most likely explain a share of the differences in monthly income. Those variables would include: labor force experience, sector of employment, employment industry, and quality of schooling.

The existing data also has limitations that could possibly be impacting our analysis. The most problematic limitation is the measurement of education. There was no continuous form of the variable in the data and so we had to construct one based on the existing categorical variable. The new proxy for continuous education is not very precise and could create misleading results if the imprecision is large enough. Ideally, we would have wanted a data set with continuous education. Additionally, the treatment of those respondents with missing income, zero income, unemployed, and those not looking for work was highly problematic. Based on the coding of the data it was extremely difficult to identify each of the previous groups. To make the data usable for the analysis, users are forced to drop all 9,999,999 and 0 income values. It is somewhat unclear who remains in the sample after this procedure and could lead to misleading inferences.

10. You are reminded that the OHS data is not a simple random sample and that weights should be utilized. Use the weights provided to re-estimate the "final model" requested in question 7. How, if at all, does this change your findings? Note: To use the weights insert the following code at the end of your regress command -> [aw=weight]

Using the sample weights reduces several of the large group differences that were observed in model Q7. Racial, gender, and location differences are all reduced once the data is properly weighted. All these reductions are quite subtle though, still leading to the same general set of conclusions. Given the proportional shares of these groups in the data it seems this is a proportional sample, if alternatively we would have seen a disproportional sample, where groups are quite different in their relative size in the sample vs. the population, then we could have seen larger differences with and without using sample weights.

<u>OPTIONAL CHALLENGE QUESTIONS</u>

Note: The questions below are completely optional and are not required.

I. Write a program that finds the difference in estimated standard errors for a given regression model when the robust option is and is not utilized.

(See do-file & log file for details)

II. Provide evidence for the improvements suggested in question 8. More specifically, actually run the analysis suggested to provide evidence of your claim.

(See do-file & log file for details)

Do file

```
* ==================================================================
* Public Policy 639 Assignment 5 - Solutions
* ==================================================================

clear all
set mem 300m
capture log close
log using assignment5solns.log, text replace


** Open Data File
use http://www-personal.umich.edu/~thomasjl/pp639/ohs94.dta, clear


** Data Management

* INCOME
recode income (0 = .) (9999999 = .), gen(inc)
lab var inc "net monthly income"

gen lninc = ln(inc)
lab var lninc "log net monthly income"

* EDUCATION
recode educ (14/15 = .) (13 = 16) (11 = 9) (12 = 10), gen(ed)
lab var ed "highest level of education"

* RACE
gen black = (race==1) if !missing(race)
lab var black "black indicator variable"

gen coloured = (race==2) if !missing(race)
lab var coloured "coloured indicator variable"

gen indian = (race==3) if !missing(race)
lab var indian "indian indicator variable"

gen white = (race==4) if !missing(race)
lab var white "white indicator variable"

* RACE-EDUCATION Interactions
gen bed = black*ed
lab var bed "black * education"

gen ced = coloured*ed
lab var bed "coloured * education"

gen ied = indian*ed
lab var ied "indian * education"

gen wed = white*ed
lab var wed "white * education"

* GENDER
recode gender (2=0 "female") (1=1 "male"), gen(male)
lab var male "male indicator variable"

* UNION
recode union (1=1 "yes") (2=0 "no"), gen(unionm)
lab var unionm "union member indicator"

* LOCATION
recode urban (1=1 "urban") (2=0 "rural"), gen(urb)
lab var urb "urban indicator"

* AGE
gen age2 = age*age
lab var age2 "age-squared"
```

```
* SAMPLE OF INTEREST
gen sample = (age>=25 & age<=65) if !missing(age)
lab var sample "sample of interest indicator"



** ANALYSIS

* Note: For all regression models, I first use Stata's Factor Variable syntax to run model
*       then I use the standard syntax form to run the same model, output is identical


* PART I

* Question 1
reg inc ed age i.race male if sample==1, r
reg inc ed age coloured indian white male if sample==1, r
est store q1

* Question 2
reg inc ed age age2 i.race male if sample==1, r
reg inc ed age age2 coloured indian white male if sample==1, r
est store q2

di "Impact of 1 year change for 25 = " ((_b[age]*26) + (_b[age2]*26*26)) - ((_b[age]*25) +
(_b[age2]*25*25))
di "Impact of 1 year change for 34 = " ((_b[age]*35) + (_b[age2]*35*35)) - ((_b[age]*34) +
(_b[age2]*34*34))
di "Impact of 1 year change for 54 = " ((_b[age]*55) + (_b[age2]*55*55)) - ((_b[age]*54) +
(_b[age2]*54*54))



* Question 3


/*
Note: There are numerous ways to write this program.  I have included a couple different
      versions to show you what is possible.
*/


* Version #1

/*
This version of the program is similar to a calculator. The user simply provides all
the information and Stata does the calculation.  When running the program users need
to provide

1st = coefficient of non-squared variable
2nd = coefficient of squared variable
3rd = larger value of non-squared variable
4th = smaller value of non-squared variable

This program can be run at any time, do not need data set open, do not need to have
run regression model prior to using command.
*/

capture program drop coef2
program define coef2
di ((`1'*`3') + (`2'*`3'*`3')) - ((`1'*`4') + (`2'*`4'*`4'))
end


* Confirm Results in Question 2, parts (a) and (b)
coef2 83.97987 -.7883657 35 34
coef2 83.97987 -.7883657 55 54




* Version #2

/*
```

```
This version will work with any model specification. When running the program user need
to provide four pieces of information:

1st = variable name of non-squared distribution
2nd = variable name of squared distribution
3rd = value of non-squared variable
4th = amount of positive change observed

This is a post-estimation command meaning it must be run after running regression. Additionally,
note that if you run command "return list" after program "coef2" the predicted change has been
stored in a scalar value.
*/

capture program drop coef2
program define coef2, rclass
args b1 b2 x dx
local y = `x' + `dx'
di "Impact of `dx' unit change in `b1' for `b1'=`x' is: " ((_b[`b1']*`y') + (_b[`b2']*`y'*`y')) -
((_b[`b1']*`x') + (_b[`b2']*`x'*`x'))
return scalar dy_`x'to`y' = ((_b[`b1']*`y') + (_b[`b2']*`y'*`y')) - ((_b[`b1']*`x') +
(_b[`b2']*`x'*`x'))
end


* Confirm Results in Question 2, parts (a) and (b)
coef2 age age2 34 1
coef2 age age2 54 1



* Question 4
reg inc c.ed##i.race if sample==1, r
reg inc ed coloured indian white ced ied wed if sample==1, r
est store q4

di "Change in Predicted Income for Blacks = " _b[ed]
di "Change in Predicted Income for Coloureds = " _b[ed] + _b[ced]
di "Change in Predicted Income for Indians = " _b[ed] + _b[ied]
di "Change in Predicted Income for Whites = " _b[ed] + _b[wed]

predict pinc
lab var pinc "predicted income - model q4"

twoway line pinc ed if race == 2, lc(red) || line pinc ed if race == 3, lc(blue) || line pinc ed if
race == 4, lc(green) ||, legend(lab(1 "coloured") lab(2 "indian") lab(3 "white"))

* Note: It is possible to also include the reference category in the graph as well (in this case
blacks)
twoway line pinc ed if race == 1, lc(black) || line pinc ed if race == 2, lc(red) || line pinc ed if
race == 3, lc(blue) || line pinc ed if race == 4, lc(green) ||, legend(lab(1 "black") lab(2 "coloured")
lab(3 "indian") lab(4 "white"))

est table q1 q2 q4, b(%9.3f) stats(r2_a N) star


* PART II

* Question 5
reg lninc ed i.race male if sample==1, r
reg lninc ed coloured indian white male if sample==1, r
est store q5

* part a - interpretation of education coefficient
di "Percent Change in Y = " 100 * ( exp(_b[ed]) - 1)

* part b - interpretation of race indicator coefficients
di "Percentage of earnings difference for coloureds = " 100 * ( exp(_b[coloured]) - 1)
di "Percentage of earnings difference for indians = " 100 * ( exp(_b[indian]) - 1)
di "Percentage of earnings difference for whites = " 100 * ( exp(_b[white]) - 1)

* part c
hist inc if sample==1, norm
hist lninc if sample==1, norm
```

```
* Question 6
reg lninc ed age i.race male if sample==1, r
reg lninc ed age coloured indian white male if sample==1, r
est store q6

* part a - interpretation of education coefficient
di "Percent Change in Y = " 100 * ( exp(_b[ed]) - 1)


* Question 7
reg lninc c.ed##i.race age age2 male unionm urb if sample==1, r
reg lninc ed age age2 coloured indian white male unionm urb ced ied wed if sample==1, r
est store q7

* Question 10
reg lninc c.ed##i.race age age2 male unionm urb if sample==1 [aw=weight], r
reg lninc ed age age2 coloured indian white male unionm urb ced ied wed if sample==1 [aw=weight], r
est store q10

est table q7 q10, b(%9.3f) stats(r2_a N) star



* Optional Challenge Question I


/*
This program produces differences for all se's in a given model. When running the
program user needs to provide variable specification as if running model in Stata
(dependent variable followed by independent variables).
*/

* Note: Program requires matrix program - matewmf to execute
*       can download ado file off CTools site

capture program drop sediff
program define sediff, rclass

quietly reg `*', robust
quietly matrix vr = e(V)
quietly matrix xr = vecdiag(vr)
quietly matrix ver = xr'
quietly matewmf ver ser, function(sqrt)

quietly reg `*'
quietly matrix vnr = e(V)
quietly matrix xnr = vecdiag(vnr)
quietly matrix venr = xnr'
quietly matewmf venr senr, function(sqrt)

quietly matrix diff = senr - ser

quietly matrix all = senr,ser,diff
quietly matrix colnames all = "SE, ~rob" "SE, rob" Diff

matrix list all

end


* Test SEDIFF command
sediff lninc ed age age2 male unionm urb if sample==1




* Optional Challenge Question II

* Race Specific Models

/* These Models provide the race specific effects of each of the independent variables.
   This is like specifying interactions between race and other iv's
*/
```

```
reg lninc ed age age2 male unionm urb if sample==1 & race==1, r
est store black

reg lninc ed age age2 male unionm urb if sample==1 & race==2, r
est store coloured

reg lninc ed age age2 male unionm urb if sample==1 & race==3, r
est store indian

reg lninc ed age age2 male unionm urb if sample==1 & race==4, r
est store white

est table black coloured indian white, b(%9.3f) stats(r2_a N) star


* Gender Specific Models

reg lninc ed age age2 coloured indian white unionm urb ced ied wed if sample==1 & male==0, r
est store female

reg lninc ed age age2 coloured indian white unionm urb ced ied wed if sample==1 & male==1, r
est store male

est table male female, b(%9.3f) stats(r2_a N) star


* Categorical Education

* Create new education variable
gen ed0 = (ed<10) if !missing(ed)
gen ed1 = (ed==10) if !missing(ed)
gen ed2 = (ed==16) if !missing(ed)

reg lninc ed1 ed2 age age2 coloured indian white male unionm urb ced ied wed if sample==1, r



log close
```

Log file

```
-----------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  assignment5solns.log
  log type:  text

.
.
. ** Open Data File
. use http://www-personal.umich.edu/~thomasjl/pp639/ohs94.dta, clear
(October Household Survey - 1994)

.
.
. ** Data Management
.
. * INCOME
. recode income (0 = .) (9999999 = .), gen(inc)
(75110 differences between income and inc)

. lab var inc "net monthly income"

.
. gen lninc = ln(inc)
(103657 missing values generated)

. lab var lninc "log net monthly income"

.
```

```
. * EDUCATION
. recode educ (14/15 = .) (13 = 16) (11 = 9) (12 = 10), gen(ed)
(6833 differences between educ and ed)

. lab var ed "highest level of education"

.
. * RACE
. gen black = (race==1) if !missing(race)

. lab var black "black indicator variable"

.
. gen coloured = (race==2) if !missing(race)

. lab var coloured "coloured indicator variable"

.
. gen indian = (race==3) if !missing(race)

. lab var indian "indian indicator variable"

.
. gen white = (race==4) if !missing(race)

. lab var white "white indicator variable"

.
. * RACE-EDUCATION Interactions
. gen bed = black*ed
(353 missing values generated)

. lab var bed "black * education"

.
. gen ced = coloured*ed
(353 missing values generated)

. lab var bed "coloured * education"

.
. gen ied = indian*ed
(353 missing values generated)

. lab var ied "indian * education"

.
. gen wed = white*ed
(353 missing values generated)

. lab var wed "white * education"

.
. * GENDER
. recode gender (2=0 "female") (1=1 "male"), gen(male)
(69872 differences between gender and male)

. lab var male "male indicator variable"

.
. * UNION
. recode union (1=1 "yes") (2=0 "no"), gen(unionm)
(24077 differences between union and unionm)

. lab var unionm "union member indicator"

.
. * LOCATION
. recode urban (1=1 "urban") (2=0 "rural"), gen(urb)
(56262 differences between urban and urb)

. lab var urb "urban indicator"

.
```

```
. * AGE
. gen age2 = age*age

. lab var age2 "age-squared"

.
. * SAMPLE OF INTEREST
. gen sample = (age>=25 & age<=65) if !missing(age)

. lab var sample "sample of interest indicator"

.
.
.
. ** ANALYSIS
.
. * Note: For all regression models, I first use Stata's Factor Variable syntax to run model
. *       then I use the standard syntax form to run the same model, output is identical
.
.
. * PART I
.
. * Question 1
. reg inc ed age i.race male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  6, 23982) =  598.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2663
                                                       Root MSE      =  1612.5

------------------------------------------------------------------------------
             |               Robust
         inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   182.0849   4.576782    39.78   0.000     173.1141    191.0557
         age |   18.32571    1.29814    14.12   0.000     15.78128    20.87015
             |
        race |
          2  |   166.1174   17.79169     9.34   0.000     131.2445    200.9902
          3  |    316.371    36.3839     8.70   0.000     245.0562    387.6857
          4  |   1041.626   31.10463    33.49   0.000      980.659    1102.593
             |
        male |   569.3923   21.57462    26.39   0.000     527.1046    611.6799
       _cons |  -1081.013   77.23482   -14.00   0.000    -1232.398   -929.6275
------------------------------------------------------------------------------

. reg inc ed age coloured indian white male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  6, 23982) =  598.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2663
                                                       Root MSE      =  1612.5

------------------------------------------------------------------------------
             |               Robust
         inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   182.0849   4.576782    39.78   0.000     173.1141    191.0557
         age |   18.32571    1.29814    14.12   0.000     15.78128    20.87015
    coloured |   166.1174   17.79169     9.34   0.000     131.2445    200.9902
      indian |    316.371    36.3839     8.70   0.000     245.0562    387.6857
       white |   1041.626   31.10463    33.49   0.000      980.659    1102.593
        male |   569.3923   21.57462    26.39   0.000     527.1046    611.6799
       _cons |  -1081.013   77.23482   -14.00   0.000    -1232.398   -929.6275
------------------------------------------------------------------------------

. est store q1

.
. * Question 2
. reg inc ed age age2 i.race male if sample==1, r
```

```
Linear regression                                      Number of obs =   23989
                                                       F(  7, 23981) =  514.82
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2681
                                                       Root MSE      =  1610.6

------------------------------------------------------------------------------
             |               Robust
         inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   181.6978   4.579382    39.68   0.000     172.7219    190.6737
         age |   83.97987   9.041891     9.29   0.000      66.2572    101.7025
        age2 |  -.7883657   .1116576    -7.06   0.000    -1.007221   -.5695098
             |
        race |
          2  |   168.7718   17.79439     9.48   0.000     133.8937    203.6499
          3  |   311.8314   36.29858     8.59   0.000     240.6839    382.9789
          4  |   1042.646    31.0672    33.56   0.000     981.7527     1103.54
             |
        male |   572.3591   21.55121    26.56   0.000     530.1173    614.6008
       _cons |  -2367.886   184.2838   -12.85   0.000    -2729.094   -2006.679
------------------------------------------------------------------------------

. reg inc ed age age2 coloured indian white male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  7, 23981) =  514.82
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2681
                                                       Root MSE      =  1610.6

------------------------------------------------------------------------------
             |               Robust
         inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   181.6978   4.579382    39.68   0.000     172.7219    190.6737
         age |   83.97987   9.041891     9.29   0.000      66.2572    101.7025
        age2 |  -.7883657   .1116576    -7.06   0.000    -1.007221   -.5695098
    coloured |   168.7718   17.79439     9.48   0.000     133.8937    203.6499
      indian |   311.8314   36.29858     8.59   0.000     240.6839    382.9789
       white |   1042.646    31.0672    33.56   0.000     981.7527     1103.54
        male |   572.3591   21.55121    26.56   0.000     530.1173    614.6008
       _cons |  -2367.886   184.2838   -12.85   0.000    -2729.094   -2006.679
------------------------------------------------------------------------------

. est store q2

.
. di "Impact of 1 year change for 25 = " ((_b[age]*26) + (_b[age2]*26*26)) - ((_b[age]*25) +
(_b[age2]*25*25))
Impact of 1 year change for 25 = 43.773222

. di "Impact of 1 year change for 34 = " ((_b[age]*35) + (_b[age2]*35*35)) - ((_b[age]*34) +
(_b[age2]*34*34))
Impact of 1 year change for 34 = 29.58264

. di "Impact of 1 year change for 54 = " ((_b[age]*55) + (_b[age2]*55*55)) - ((_b[age]*54) +
(_b[age2]*54*54))
Impact of 1 year change for 54 = -1.9519868

.
.
.
. * Question 3
.
.
. /*
> Note: There are numerous ways to write this program.  I have included a couple different
>       versions to show you what is possible.
> */
.
.
. * Version #1
.
```

```
. /*
> This version of the program is similar to a calculator. The user simply provides all
> the information and Stata does the calculation.  When running the program users need
> to provide
>
> 1st = coefficient of non-squared variable
> 2nd = coefficient of squared variable
> 3rd = larger value of non-squared variable
> 4th = smaller value of non-squared variable
>
> This program can be run at any time, do not need data set open, do not need to have
> run regression model prior to using command.
> */
.
. capture program drop coef2

. program define coef2
  1. di ((`1'*`3') + (`2'*`3'*`3')) - ((`1'*`4') + (`2'*`4'*`4'))
  2. end

.
.
. * Confirm Results in Question 2, parts (a) and (b)
. coef2 83.97987 -.7883657 35 34
29.582637

. coef2 83.97987 -.7883657 55 54
-1.9519913

.
.
.
.
. * Version #2
.
. /*
> This version will work with any model specification. When running the program user need
> to provide four pieces of information:
>
> 1st = variable name of non-squared distribution
> 2nd = variable name of squared distribution
> 3rd = value of non-squared variable
> 4th = amount of positive change observed
>
> This is a post-estimation command meaning it must be run after running regression. Additionally,
> note that if you run command "return list" after program "coef2" the predicted change has been
> stored in a scalar value.
> */
.
. capture program drop coef2

. program define coef2, rclass
  1. args b1 b2 x dx
  2. local y = `x' + `dx'
  3. di "Impact of `dx' unit change in `b1' for `b1'=`x' is: " ((_b[`b1']*`y') + (_b[`b2']*`y'*`y')) -
((_b[`b1']*`x') + (_b[`b2']*`x'*`x'))
  4. return scalar dy_`x'to`y' = ((_b[`b1']*`y') + (_b[`b2']*`y'*`y')) - ((_b[`b1']*`x') +
(_b[`b2']*`x'*`x'))
  5. end

.
.
. * Confirm Results in Question 2, parts (a) and (b)
. coef2 age age2 34 1
Impact of 1 unit change in age for age=34 is: 29.58264

. coef2 age age2 54 1
Impact of 1 unit change in age for age=54 is: -1.9519868

.
.
.
. * Question 4
. reg inc c.ed##i.race if sample==1, r
```

```
Linear regression                                      Number of obs =   23989
                                                       F(  7, 23981) = 1065.10
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2456
                                                       Root MSE      =  1635.1

------------------------------------------------------------------------------
             |               Robust
         inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   127.4118    3.08015    41.37   0.000     121.3745    133.4491
             |
        race |
          2  |  -31.01306   27.36814    -1.13   0.257    -84.65633    22.63022
          3  |  -24.37371   111.6997    -0.22   0.827    -243.3122    194.5648
          4  |  -448.5553   219.4672    -2.04   0.041    -878.7249   -18.38576
             |
   race#c.ed |
          2  |   23.10366   5.198957     4.44   0.000     12.91338    33.29395
          3  |   58.61983   14.39903     4.07   0.000     30.39683    86.84283
          4  |   165.7409   23.57246     7.03   0.000     119.5374    211.9444
             |
       _cons |   335.1206    16.1871    20.70   0.000     303.3929    366.8483
------------------------------------------------------------------------------

. reg inc ed coloured indian white ced ied wed if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  7, 23981) = 1065.10
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2456
                                                       Root MSE      =  1635.1

------------------------------------------------------------------------------
             |               Robust
         inc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   127.4118    3.08015    41.37   0.000     121.3745    133.4491
    coloured |  -31.01306   27.36814    -1.13   0.257    -84.65633    22.63022
      indian |  -24.37371   111.6997    -0.22   0.827    -243.3122    194.5648
       white |  -448.5553   219.4672    -2.04   0.041    -878.7249   -18.38576
         ced |   23.10366   5.198957     4.44   0.000     12.91338    33.29395
         ied |   58.61983   14.39903     4.07   0.000     30.39683    86.84283
         wed |   165.7409   23.57246     7.03   0.000     119.5374    211.9444
       _cons |   335.1206    16.1871    20.70   0.000     303.3929    366.8483
------------------------------------------------------------------------------

. est store q4

.
. di "Change in Predicted Income for Blacks = " _b[ed]
Change in Predicted Income for Blacks = 127.41183

. di "Change in Predicted Income for Coloureds = " _b[ed] + _b[ced]
Change in Predicted Income for Coloureds = 150.51549

. di "Change in Predicted Income for Indians = " _b[ed] + _b[ied]
Change in Predicted Income for Indians = 186.03166

. di "Change in Predicted Income for Whites = " _b[ed] + _b[wed]
Change in Predicted Income for Whites = 293.15271

.
. predict pinc
(option xb assumed; fitted values)
(353 missing values generated)

. lab var pinc "predicted income - model q4"

.
. twoway line pinc ed if race == 2, lc(red) || line pinc ed if race == 3, lc(blue) || line pinc ed if
race == 4, lc(green) ||, legend(lab(1 "coloured") lab(2 "indian") lab(3 "white"))
```

```
.
. * Note: It is possible to also include the reference category in the graph as well (in this case
blacks)
. twoway line pinc ed if race == 1, lc(black) || line pinc ed if race == 2, lc(red) || line pinc ed if
race == 3, lc(blue) || line pinc ed if race == 4, lc(green) ||, legend(lab(1 "black") lab
> (2 "coloured") lab(3 "indian") lab(4 "white"))

.
. est table q1 q2 q4, b(%9.3f) stats(r2_a N) star

-----------------------------------------------------------
    Variable |      q1            q2            q4         
-------------+---------------------------------------------
          ed |  182.085***     181.698***     127.412***   
         age |   18.326***      83.980***                  
    coloured |  166.117***     168.772***     -31.013      
      indian |  316.371***     311.831***     -24.374      
       white | 1041.626***    1042.646***    -448.555*     
        male |  569.392***     572.359***                  
        age2 |                  -0.788***                  
         ced |                                 23.104***   
         ied |                                 58.620***   
         wed |                                165.741***   
       _cons | -1081.013***   -2367.886***     335.121***   
-------------+---------------------------------------------
        r2_a |     0.266          0.268          0.245     
           N |     23989          23989          23989     
-----------------------------------------------------------
                   legend: * p<0.05; ** p<0.01; *** p<0.001

.
.
. * PART II
.
. * Question 5
. reg lninc ed i.race male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  5, 23983) = 3341.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4396
                                                       Root MSE      =  .70745

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |    .129736   .0014448    89.80   0.000     .1269042    .1325679
             |
        race |
          2  |   .1591345   .0117975    13.49   0.000     .1360108    .1822583
          3  |   .3461424   .0156513    22.12   0.000     .3154648    .3768199
          4  |   .5496654   .0125073    43.95   0.000     .5251503    .5741805
             |
        male |   .2732818   .0095804    28.53   0.000     .2545036      .29206
       _cons |   5.673454   .0140189   404.70   0.000     5.645976    5.700932
------------------------------------------------------------------------------

. reg lninc ed coloured indian white male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  5, 23983) = 3341.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4396
                                                       Root MSE      =  .70745

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |    .129736   .0014448    89.80   0.000     .1269042    .1325679
    coloured |   .1591345   .0117975    13.49   0.000     .1360108    .1822583
      indian |   .3461424   .0156513    22.12   0.000     .3154648    .3768199
       white |   .5496654   .0125073    43.95   0.000     .5251503    .5741805
```

```
        male |   .2732818   .0095804    28.53   0.000     .2545036      .29206
       _cons |   5.673454   .0140189   404.70   0.000     5.645976    5.700932
------------------------------------------------------------------------------

. est store q5

. 
. * part a - interpretation of education coefficient
. di "Percent Change in Y = " 100 * ( exp(_b[ed]) - 1)
Percent Change in Y = 13.852779

. 
. * part b - interpretation of race indicator coefficients
. di "Percentage of earnings difference for coloureds = " 100 * ( exp(_b[coloured]) - 1)
Percentage of earnings difference for coloureds = 17.249569

. di "Percentage of earnings difference for indians = " 100 * ( exp(_b[indian]) - 1)
Percentage of earnings difference for indians = 41.360385

. di "Percentage of earnings difference for whites = " 100 * ( exp(_b[white]) - 1)
Percentage of earnings difference for whites = 73.267316

. 
. * part c
. hist inc if sample==1, norm
(bin=43, start=8, width=1220.7442)

. hist lninc if sample==1, norm
(bin=43, start=2.0794415, width=.2043983)

. 
. 
. * Question 6
. reg lninc ed age i.race male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  6, 23982) = 2863.24
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4472
                                                       Root MSE      =  .70264

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |    .135212   .0014695    92.01   0.000     .1323317    .1380922
         age |   .0087547   .0004904    17.85   0.000     .0077934     .009716
             |
        race |
          2  |   .1705928   .0117013    14.58   0.000     .1476576    .1935281
          3  |   .3339924   .0155913    21.42   0.000     .3034324    .3645524
          4  |    .518292   .0125666    41.24   0.000     .4936607    .5429233
             |
        male |   .2659966   .0095272    27.92   0.000     .2473226    .2846705
       _cons |   5.303452   .0247531   214.25   0.000     5.254934    5.351969
------------------------------------------------------------------------------

. reg lninc ed age coloured indian white male if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F(  6, 23982) = 2863.24
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4472
                                                       Root MSE      =  .70264

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |    .135212   .0014695    92.01   0.000     .1323317    .1380922
         age |   .0087547   .0004904    17.85   0.000     .0077934     .009716
    coloured |   .1705928   .0117013    14.58   0.000     .1476576    .1935281
      indian |   .3339924   .0155913    21.42   0.000     .3034324    .3645524
       white |    .518292   .0125666    41.24   0.000     .4936607    .5429233
```

```
       male |   .2659966   .0095272    27.92   0.000     .2473226    .2846705
      _cons |   5.303452   .0247531   214.25   0.000     5.254934    5.351969
------------------------------------------------------------------------------

. est store q6

.
. * part a - interpretation of education coefficient
. di "Percent Change in Y = " 100 * ( exp(_b[ed]) - 1)
Percent Change in Y = 14.477942

.
.
. * Question 7
. reg lninc c.ed##i.race age age2 male unionm urb if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F( 12, 23976) = 1769.07
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4828
                                                       Root MSE      =  .67968

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1249175   .0019636    63.62   0.000     .1210688    .1287662
             |
        race |
          2  |    .145534   .0226989     6.41   0.000     .1010428    .1900253
          3  |   .5108085   .0444032    11.50   0.000     .4237755    .5978415
          4  |   .8522006    .044144    19.31   0.000     .7656757    .9387255
             |
   race#c.ed |
          2  |   .0010384   .0032788     0.32   0.751    -.0053882     .007465
          3  |   -.027876    .005041    -5.53   0.000    -.0377567   -.0179952
          4  |  -.0372341   .0044686    -8.33   0.000    -.0459928   -.0284754
             |
         age |   .0461053   .0036622    12.59   0.000     .0389272    .0532835
        age2 |  -.0004574    .000044   -10.39   0.000    -.0005436   -.0003711
        male |   .2784063   .0093511    29.77   0.000     .2600775    .2967351
      unionm |   .1932443    .009676    19.97   0.000     .1742788    .2122098
         urb |   .2725235   .0114755    23.75   0.000     .2500308    .2950161
       _cons |   4.432359   .0746123    59.41   0.000     4.286114    4.578603
------------------------------------------------------------------------------

. reg lninc ed age age2 coloured indian white male unionm urb ced ied wed if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F( 12, 23976) = 1769.07
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4828
                                                       Root MSE      =  .67968

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1249175   .0019636    63.62   0.000     .1210688    .1287662
         age |   .0461053   .0036622    12.59   0.000     .0389272    .0532835
        age2 |  -.0004574    .000044   -10.39   0.000    -.0005436   -.0003711
    coloured |    .145534   .0226989     6.41   0.000     .1010428    .1900253
      indian |   .5108085   .0444032    11.50   0.000     .4237755    .5978415
       white |   .8522006    .044144    19.31   0.000     .7656757    .9387255
        male |   .2784063   .0093511    29.77   0.000     .2600775    .2967351
      unionm |   .1932443    .009676    19.97   0.000     .1742788    .2122098
         urb |   .2725235   .0114755    23.75   0.000     .2500308    .2950161
         ced |   .0010384   .0032788     0.32   0.751    -.0053882     .007465
         ied |   -.027876    .005041    -5.53   0.000    -.0377567   -.0179952
         wed |  -.0372341   .0044686    -8.33   0.000    -.0459928   -.0284754
       _cons |   4.432359   .0746123    59.41   0.000     4.286114    4.578603
------------------------------------------------------------------------------

. est store q7
```

```
.
. * Question 10
. reg lninc c.ed##i.race age age2 male unionm urb if sample==1 [aw=weight], r
(sum of wgt is   6.8345e+06)

Linear regression                                      Number of obs =   23989
                                                       F( 12, 23976) = 1181.64
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4663
                                                       Root MSE      =  .68823

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1210572   .0022262    54.38   0.000     .1166937    .1254207
             |
        race |
          2  |   .1794265   .0263461     6.81   0.000     .1277864    .2310665
          3  |   .4276853   .0552518     7.74   0.000     .3193883    .5359822
          4  |   .7143135   .0511103    13.98   0.000     .6141342    .8144928
             |
   race#c.ed |
          2  |  -.0041331   .0037291    -1.11   0.268    -.0114424    .0031762
          3  |  -.0231736   .0062978    -3.68   0.000    -.0355177   -.0108294
          4  |  -.0212476   .0050907    -4.17   0.000    -.0312257   -.0112695
             |
         age |   .0406899   .0046307     8.79   0.000     .0316135    .0497663
        age2 |  -.0003831    .000056    -6.84   0.000    -.0004929   -.0002734
        male |   .2544953   .0113932    22.34   0.000      .232164    .2768266
      unionm |    .226595   .0119657    18.94   0.000     .2031416    .2500484
         urb |     .25336    .014074    18.00   0.000     .2257742    .2809458
       _cons |   4.579702    .093779    48.84   0.000     4.395889    4.763515
------------------------------------------------------------------------------

. reg lninc ed age age2 coloured indian white male unionm urb ced ied wed if sample==1 [aw=weight], r
(sum of wgt is   6.8345e+06)

Linear regression                                      Number of obs =   23989
                                                       F( 12, 23976) = 1181.64
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4663
                                                       Root MSE      =  .68823

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1210572   .0022262    54.38   0.000     .1166937    .1254207
         age |   .0406899   .0046307     8.79   0.000     .0316135    .0497663
        age2 |  -.0003831    .000056    -6.84   0.000    -.0004929   -.0002734
    coloured |   .1794265   .0263461     6.81   0.000     .1277864    .2310665
      indian |   .4276853   .0552518     7.74   0.000     .3193883    .5359822
       white |   .7143135   .0511103    13.98   0.000     .6141342    .8144928
        male |   .2544953   .0113932    22.34   0.000      .232164    .2768266
      unionm |    .226595   .0119657    18.94   0.000     .2031416    .2500484
         urb |     .25336    .014074    18.00   0.000     .2257742    .2809458
         ced |  -.0041331   .0037291    -1.11   0.268    -.0114424    .0031762
         ied |  -.0231736   .0062978    -3.68   0.000    -.0355177   -.0108294
         wed |  -.0212476   .0050907    -4.17   0.000    -.0312257   -.0112695
       _cons |   4.579702    .093779    48.84   0.000     4.395889    4.763515
------------------------------------------------------------------------------

. est store q10

.
. est table q7 q10, b(%9.3f) stats(r2_a N) star

--------------------------------------------
    Variable |      q7            q10
-------------+------------------------------
          ed |     0.125***       0.121***
         age |     0.046***       0.041***
```

```
        age2 |    -0.000***      -0.000***
    coloured |     0.146***       0.179***
      indian |     0.511***       0.428***
       white |     0.852***       0.714***
        male |     0.278***       0.254***
      unionm |     0.193***       0.227***
         urb |     0.273***       0.253***
         ced |     0.001         -0.004
         ied |    -0.028***      -0.023***
         wed |    -0.037***      -0.021***
       _cons |     4.432***       4.580***
-------------+----------------------------
        r2_a |     0.483          0.466
           N |     23989          23989
------------------------------------------
    legend: * p<0.05; ** p<0.01; *** p<0.001

.
.
.
. * Optional Challenge Question I
.
.
. /*
> This program produces differences for all se's in a given model. When running the
> program user needs to provide variable specification as if running model in Stata
> (dependent variable followed by independent variables).
> */
.
. * Note: Program requires matrix program - matewmf to execute
. *       can download ado file off CTools site
.
. capture program drop sediff

. program define sediff, rclass
  1.
. quietly reg `*', robust
  2. quietly matrix vr = e(V)
  3. quietly matrix xr = vecdiag(vr)
  4. quietly matrix ver = xr'
  5. quietly matewmf ver ser, function(sqrt)
  6.
. quietly reg `*'
  7. quietly matrix vnr = e(V)
  8. quietly matrix xnr = vecdiag(vnr)
  9. quietly matrix venr = xnr'
 10. quietly matewmf venr senr, function(sqrt)
 11.
. quietly matrix diff = senr - ser
 12.
. quietly matrix all = senr,ser,diff
 13. quietly matrix colnames all = "SE, ~rob" "SE, rob" Diff
 14.
. matrix list all
 15.
. end

.
.
. * Test SEDIFF command
. sediff lninc ed age age2 male unionm urb if sample==1

all[7,3]
          SE, ~rob     SE, rob        Diff
    ed   .00135888   .00143487  -.00007599
   age   .00374974   .00377664   -.0000269
  age2   .00004465   .00004536  -7.077e-07
  male    .0095298   .00957582  -.00004603
unionm   .01000496   .00962498   .00037997
   urb   .01085156   .01130037  -.00044881
 _cons   .07630478   .07626488   .00003991

.
.
```

```
.
.
. * Optional Challenge Question II
.
. * Race Specific Models
.
. /* These Models provide the race specific effects of each of the independent variables.
>    This is like specifying interactions between race and other iv's
> */
.
. reg lninc ed age age2 male unionm urb if sample==1 & race==1, r

Linear regression                                      Number of obs =   10635
                                                       F(  6, 10628) = 1176.90
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3841
                                                       Root MSE      =  .70546

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1204165   .0021075    57.14   0.000     .1162854    .1245476
         age |   .0371361   .0057712     6.43   0.000     .0258235    .0484487
        age2 |  -.0003474    .000069    -5.03   0.000    -.0004826   -.0002121
        male |   .2255936   .0150619    14.98   0.000     .1960695    .2551178
      unionm |   .3442738   .0148383    23.20   0.000      .315188    .3733596
         urb |   .2460719   .0150833    16.31   0.000     .2165058    .2756379
       _cons |   4.626514   .1173128    39.44   0.000     4.396559    4.856469
------------------------------------------------------------------------------

. est store black

.
. reg lninc ed age age2 male unionm urb if sample==1 & race==2, r

Linear regression                                      Number of obs =    5297
                                                       F(  6,  5290) =  627.15
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4150
                                                       Root MSE      =  .63457

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1109064   .0035638    31.12   0.000     .1039198     .117893
         age |   .0566034   .0072702     7.79   0.000     .0423508     .070856
        age2 |  -.0006169   .0000879    -7.02   0.000    -.0007893   -.0004446
        male |   .2188881    .018133    12.07   0.000       .18334    .2544361
      unionm |    .157337   .0202563     7.77   0.000     .1176262    .1970478
         urb |   .4315379   .0253078    17.05   0.000     .3819241    .4811516
       _cons |   4.452872    .144953    30.72   0.000     4.168705     4.73704
------------------------------------------------------------------------------

. est store coloured

.
. reg lninc ed age age2 male unionm urb if sample==1 & race==3, r

Linear regression                                      Number of obs =    2214
                                                       F(  6,  2207) =   80.88
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1876
                                                       Root MSE      =  .61521

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0909851   .0051466    17.68   0.000     .0808925    .1010777
         age |   .0433862   .0115289     3.76   0.000     .0207776    .0659949
        age2 |  -.0004636   .0001405    -3.30   0.001    -.0007392   -.0001881
        male |   .2322556   .0294124     7.90   0.000     .1745767    .2899344
```

```
      unionm |    .014504   .0269862     0.54   0.591    -.0384171    .0674251
         urb |    .315058   .0374204     8.42   0.000     .2416752    .3884408
       _cons |    5.16778   .2402847    21.51   0.000     4.696572    5.638988
------------------------------------------------------------------------------

. est store indian

.
. reg lninc ed age age2 male unionm urb if sample==1 & race==4, r

Linear regression                                      Number of obs =    5843
                                                       F(  6,  5836) =  206.36
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1905
                                                       Root MSE      =  .66557

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0814586   .0040335    20.20   0.000     .0735515    .0893657
         age |   .0516728   .0069174     7.47   0.000      .038112    .0652335
        age2 |  -.0005194   .0000825    -6.30   0.000    -.0006812   -.0003577
        male |   .4712262   .0179741    26.22   0.000     .4359904     .506462
      unionm |   -.050192   .0195029    -2.57   0.010     -.088425   -.0119591
         urb |   .0524171   .0309875     1.69   0.091    -.0083299     .113164
       _cons |   5.384006   .1455013    37.00   0.000     5.098769    5.669242
------------------------------------------------------------------------------

. est store white

.
. est table black coloured indian white, b(%9.3f) stats(r2_a N) star

--------------------------------------------------------------------------
    Variable |    black        coloured       indian         white
-------------+------------------------------------------------------------
          ed |     0.120***       0.111***       0.091***       0.081***
         age |     0.037***       0.057***       0.043***       0.052***
        age2 |    -0.000***      -0.001***      -0.000***      -0.001***
        male |     0.226***       0.219***       0.232***       0.471***
      unionm |     0.344***       0.157***       0.015         -0.050*
         urb |     0.246***       0.432***       0.315***       0.052
       _cons |     4.627***       4.453***       5.168***       5.384***
-------------+------------------------------------------------------------
        r2_a |     0.384          0.414          0.185          0.190
           N |     10635           5297           2214           5843
--------------------------------------------------------------------------
                                  legend: * p<0.05; ** p<0.01; *** p<0.001

.
.
. * Gender Specific Models
.
. reg lninc ed age age2 coloured indian white unionm urb ced ied wed if sample==1 & male==0, r

Linear regression                                      Number of obs =    8888
                                                       F( 11,  8876) =  517.29
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4240
                                                       Root MSE      =  .67393

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1457078   .0035833    40.66   0.000     .1386837     .152732
         age |   .0324039   .0062227     5.21   0.000     .0202059    .0446018
        age2 |  -.0003187    .000076    -4.19   0.000    -.0004678   -.0001696
    coloured |   .2397744    .044996     5.33   0.000     .1515719    .3279768
      indian |   .7853478    .083924     9.36   0.000     .6208373    .9498583
       white |   .9559935    .071689    13.34   0.000     .8154664    1.096521
      unionm |   .1789033   .0158157    11.31   0.000      .147901    .2099057
         urb |   .2002103   .0199326    10.04   0.000     .1611378    .2392828
```

```
         ced |  -.0093801   .0058798    -1.60   0.111     -.020906    .0021457
         ied |  -.0595112   .0091267    -6.52   0.000    -.0774016   -.0416208
         wed |  -.0631601   .0071904    -8.78   0.000    -.0772549   -.0490653
       _cons |   4.672836   .1262312    37.02   0.000     4.425393    4.920278
------------------------------------------------------------------------------

. est store female

.
. reg lninc ed age age2 coloured indian white unionm urb ced ied wed if sample==1 & male==1, r

Linear regression                                      Number of obs =   15101
                                                       F( 11, 15089) = 1447.87
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5162
                                                       Root MSE      =  .67841

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .1132717   .0023982    47.23   0.000      .108571    .1179725
         age |    .056154   .0045362    12.38   0.000     .0472625    .0650454
        age2 |  -.0005618    .000054   -10.40   0.000    -.0006677   -.0004558
    coloured |   .1115784   .0262733     4.25   0.000     .0600795    .1630772
      indian |   .3928093   .0518858     7.57   0.000     .2911067    .4945118
       white |   .8217936   .0560652    14.66   0.000     .7118991    .9316881
      unionm |   .1905523   .0122243    15.59   0.000     .1665912    .2145134
         urb |   .3150831   .0139297    22.62   0.000     .2877792     .342387
         ced |   .0046844   .0040786     1.15   0.251    -.0033101    .0126789
         ied |  -.0120209   .0060344    -1.99   0.046    -.0238492   -.0001927
         wed |  -.0227115   .0057146    -3.97   0.000    -.0339128   -.0115102
       _cons |   4.515269   .0922187    48.96   0.000     4.334509    4.696029
------------------------------------------------------------------------------

. est store male

.
. est table male female, b(%9.3f) stats(r2_a N) star

--------------------------------------------
    Variable |     male          female
-------------+------------------------------
          ed |     0.113***       0.146***
         age |     0.056***       0.032***
        age2 |    -0.001***      -0.000***
    coloured |     0.112***       0.240***
      indian |     0.393***       0.785***
       white |     0.822***       0.956***
      unionm |     0.191***       0.179***
         urb |     0.315***       0.200***
         ced |     0.005         -0.009
         ied |    -0.012*        -0.060***
         wed |    -0.023***      -0.063***
       _cons |     4.515***       4.673***
-------------+------------------------------
        r2_a |     0.516          0.423
           N |     15101           8888
--------------------------------------------
    legend: * p<0.05; ** p<0.01; *** p<0.001

.
.
. * Categorical Education
.
. * Create new education variable
. gen ed0 = (ed<10) if !missing(ed)
(353 missing values generated)

. gen ed1 = (ed==10) if !missing(ed)
(353 missing values generated)

. gen ed2 = (ed==16) if !missing(ed)
(353 missing values generated)
```

```
.
. reg lninc ed1 ed2 age age2 coloured indian white male unionm urb ced ied wed if sample==1, r

Linear regression                                      Number of obs =   23989
                                                       F( 13, 23975) = 1408.56
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4421
                                                       Root MSE      =  .70596

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ed1 |   .5790345   .0122306    47.34   0.000     .5550618    .6030073
         ed2 |   1.121462   .0361033    31.06   0.000     1.050697    1.192227
         age |   .0506855   .0038186    13.27   0.000     .0432007    .0581702
        age2 |  -.0005459   .0000459   -11.90   0.000    -.0006359    -.000456
    coloured |  -.2051571   .0218377    -9.39   0.000    -.2479604   -.1623538
      indian |   .5351784   .0498086    10.74   0.000     .4375505    .6328063
       white |   1.001512   .0597174    16.77   0.000     .8844617    1.118561
        male |   .2420462   .0096817    25.00   0.000     .2230694     .261023
      unionm |   .2589028   .0098452    26.30   0.000     .2396056       .2782
         urb |   .4202711   .0115939    36.25   0.000     .3975463     .442996
         ced |   .0641154    .003178    20.17   0.000     .0578864    .0703445
         ied |  -.0133156   .0056131    -2.37   0.018    -.0243175   -.0023136
         wed |  -.0379373   .0060644    -6.26   0.000    -.0498238   -.0260508
       _cons |    4.91553   .0768094    64.00   0.000     4.764978    5.066081
------------------------------------------------------------------------------

.
.
.
. log close
      name:  <unnamed>
       log:  assignment5solns.log
  log type:  text
----------------------------------------------------------------------------------------------------
```