'jessica.vanparys@hunter

```
*==============================================================================
*Title: wage2_monica.do
*Author: Monica Elgawly
*Date Modified: 4/15/2018
*Assignment 4 Part 1
*==============================================================================

cap log close
clear
set more off

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

*==============================================================================
*1a) Use the wage2.dta to estimate the following model where the return to education
*depends upon the amount of work experience (and vice versa):
* log(wagei) = B0 + B1edu1 + B2experi + B3(educi)(experi) + ui
*Show that the marginal effect of another year of education equals B1 + B3exper.
*Plug in the values of B1 and B3 to get the general form for the marginal effect
*of education.

use "/Users/monicaelgawly/Downloads/wage2.dta"
gen educexper = educ*exper
reg lwage educ exper educexper, r

Take the derivative of the equation with regard to edu we get B1*B3*exper
From the regression estimate: 0.44 + .003*exper
*==============================================================================


*==============================================================================
*b) Test whether the return to education depends on the level of experience.

H0 = B3 = 0, Ha = B3 != 0
p-value of B3 hat indicates that B3 is statistically significant at 5% level.
*==============================================================================

*==============================================================================
*c) Now allow education, experience, job tenure, marriage status, race and geographic
*location to determine wages by estimating the following model:
*log(wagei) = B0 + B1edui + B2experi + B3tenurei + B4marriedi + B5blacki + B6southi
*+ B7urbani + ui. Report the results in standard form. Holding other factors fixed,
*what is the approximate difference in monthly salary between married people and
*nonmarried people? Is this difference statistically significant at the 5% level?

reg lwage educ exper tenure married black south urban, r

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

*==============================================================================
*d) Modify the model in part(c) by allowing log(wage) to differ across 4 groups of
* people: married and black, married and nonblack, single and black, and single and
*nonblack. What is the estimated wage differential between married nonblack people
*and nonmarried nonblack people?

gen mb = 0
replace mb = 1 if married == 1 & black == 1
gen nmb = 0
replace nmb = 1 if married == 0 & black == 1
gen mnb = 0
replace mnb = 1 if married == 1 & black == 0
```

```
gen nmnb = 0
replace nmnb = 1 if married == 0 & black == 0

if Benchmark category is not married, not black
reg lwage educ exper tenure south urban mb nmb mnb, r
married, nonblack people earn on average 20% more than nonmarried nonblack

if Benchmark category is married, not black
reg lwage educ exper tenure south urban mb nmb nmnb, r
nonmarried, nonblack people earn on average 19% less than married nonblack
*===========================================================================

*===========================================================================
*e) On a graph with education on the x-axis and log(wage) on the y-axis, draw the
*sample regression functions for the four groups of people (1) married and black,
*(2) married and nonblack, (3) single and black, and (4) single and nonblack, holding
*constant experience, tenure, and geographic location (i.e., south and urban). You do
*not need to create this graph in Stata, but you will use the results from part (d)
*to draw the graph on paper.
*===========================================================================

clear
```

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1
  Statistics/Data Analysis
```

Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

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

Notes:
1. Unicode is supported; see help unicode advice.

1 . log using "/Users/monicaelgawly/Downloads/wage2_monica.log"

name: <unnamed>
log: /Users/monicaelgawly/Downloads/wage2_monica.log
log type: text
opened on: 15 Apr 2018, 00:01:35

2 . use "/Users/monicaelgawly/Downloads/wage2.dta"

3 . gen educexper = educ*exper

4 . reg lwage educ exper educexper, r

Linear regression

Number of obs	=	935
F(3, 931)	=	50.45
Prob > F	=	0.0000
R-squared	=	0.1349
Root MSE	=	.39233

lwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0440498	.0178328	2.47	0.014	.0090527	.0790469
exper	-.0214959	.0193845	-1.11	0.268	-.0595383	.0165465
educexper	.003203	.0014829	2.16	0.031	.0002928	.0061132
_cons	5.949455	.2468364	24.10	0.000	5.465035	6.433875

5 . reg lwage educ exper tenure married black south urban, r

Linear regression

Number of obs	=	935
F(7, 927)	=	50.83
Prob > F	=	0.0000
R-squared	=	0.2526
Root MSE	=	.36547

lwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0654307	.0064093	10.21	0.000	.0528524	.0780091
exper	.014043	.0032386	4.34	0.000	.0076872	.0203988
tenure	.0117473	.0025387	4.63	0.000	.006765	.0167295
married	.1994171	.0396937	5.02	0.000	.121517	.2773171
black	-.1883499	.0367035	-5.13	0.000	-.2603816	-.1163182
south	-.0909036	.027363	-3.32	0.001	-.1446043	-.037203
urban	.1839121	.0271125	6.78	0.000	.1307031	.237121
_cons	5.395497	.1131274	47.69	0.000	5.173482	5.617513

6 . tab educ

years of education	Freq.	Percent	Cum.
9	10	1.07	1.07
10	35	3.74	4.81
11	43	4.60	9.41
12	393	42.03	51.44
13	85	9.09	60.53
14	77	8.24	68.77
15	45	4.81	73.58
16	150	16.04	89.63
17	40	4.28	93.90
18	57	6.10	100.00
Total	935	100.00	

7 . sum educ

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	935	13.46845	2.196654	9	18

8 . sum exper

Variable	Obs	Mean	Std. Dev.	Min	Max

```
       exper |        935    11.563644   4.374586          1         23

  9 . sum tenure

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      tenure |        935    7.234225    5.075206          0         22

 10 . sum married

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     married |        935    .8930481    .3092174          0          1

 11 . sum black

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       black |        935    .1283422    .3346495          0          1

 12 . sum south

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       south |        935    .3411765    .4743582          0          1

 13 . sum urban

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       urban |        935    .7176471    .4503851          0          1

 14 . gen mb = 0

 15 . replace mb = 1 if married == 1 & black == 1
(102 real changes made)

 16 . gen nmb = 0

 17 . replace nmb = 1 if married == 0 & black == 1
(18 real changes made)

 18 . gen mnb = 0

 19 . replace mnb = 1 if married == 1 & black == 0
(733 real changes made)

 20 . gen nmnb = 0
```



```
 21 . replace nmnb = 1 if married == 0 & black == 0
(82 real changes made)

 22 . reg lwage educ exper tenure south urban mb nmb nmnb, r

Linear regression                               Number of obs     =        935
                                                F(8, 926)         =      44.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2528
                                                Root MSE          =      .3656

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0654751   .0064153    10.21   0.000     .0528848    .0780654
       exper |   .0141462   .0032479     4.36   0.000     .0077721    .0205203
      tenure |   .0116628   .0025493     4.57   0.000     .0066597     .016666
       south |  -.0919894   .0274988    -3.35   0.001    -.1459566   -.0380222
       urban |   .1843501   .0271884     6.78   0.000      .130992    .2377081
          mb |  -.1794663   .0403216    -4.45   0.000    -.2585986   -.1003339
         nmb |  -.4297347   .0731179    -5.88   0.000    -.5732308   -.2862386
        nmnb |  -.1889147   .0448302    -4.21   0.000    -.2768953   -.1009341
       _cons |   5.592708   .1093308    51.15   0.000     5.378143    5.807273
------------------------------------------------------------------------------

 23 .
```

wage2_monica.do
Monica Elgawly
4/18/2018
Assignment 4 Part I

1a) to show that the marginal effect of another year of education $= \beta_1 + \beta_3 exper_i$

our model:

$$\log(wage_i) = \beta_0 + \beta_1 \overset{X_{i1}}{educ_1} + \beta_2 \overset{X_{i2}}{exper_i} + \beta_3 (\overset{X_{i1}}{educ_i})(\overset{X_{i2}}{exper_i}) + u_i$$

- unlike OLS, we can't directly use the coefficients from probit model to measure marginal impact of the regressors on the dependent variable. however, the intuitions of inc/dec (more/less likely) for both models is the same.
- under OLS, the coefficient of regressor tells us the marginal effect of that variable on the dependent variable bc it is the first derivative of the dependent variable with respect to that regressor

ex/ in a linear model $y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + u_i$ the coefficient $\hat{\beta}_1$ tells the effect of one unit change in X_1 on Y because $dy/dx = \tilde{\beta}_1$.

* The marginal effect modelled by the probit model depends on where the observation is in the distribution.

Go to LC#16 interactions btwn 2 binary variables: slides $\frac{5}{16}$ through $\frac{10}{16}$

now educ and exper are not binary/dummy variables

they are continuous variables: LC#17 slide 13/18

* if $\beta_3 > 0$ then the return to education increases with addition years of experience, compared with people with less education
* we say the marginal effect of education (on wages) depends on experience & vice versa.

$$\frac{d\log(wage_i)}{deduc_i} = \beta_1 + \beta_3 exper_i$$

Key Concept 8.5 P. 326/827 The coefficient on $X_1 \times X_2$ is the effect of a one-unit increase in X_1 and X_2 above & beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone. This is true whether X_1 and/or X_2 are continuous or binary.

323 KeyC 8.1 Eqn 8.4 Eqn 8.35

- The expected change in Y, ΔY, associated with the change in X_1, ΔX_1, holding $X_2 \cdots X_K$ constant is the difference between the value of the population regression function before and after changing X_1, holding $X_2 \cdots X_K$ constant. That is, the expected change in Y is the difference $= \Delta Y = f(X_1 + \Delta X_1, X_2, \cdots X_k) - f(X_1, X_2, \cdots X_K)$. (8.4)
  * The same applies to estimator/estimation.

The ΔY applied to $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} * X_{2i}) + u_i$ (8.35) is $\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1$

This then is the mathematical proof of the marginal effect as follows:

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$$

The general form of the marginal effect: $\frac{\Delta Y}{\Delta X_1} = .044 + .003(exper_i)$

$$\frac{\Delta \log(wage_i)}{\Delta edu_{i1}} = 0.44 + .003(exper_i)$$

b) According to information gathered from the regression already utilized in (a) the p-value for educexper is 0.031. This value of .031 is less than .05 which is the critical point of testing for 95% confidence. When $p \leq .05$, we have a significance value that suggests we have the data necessary

$p \leq .05$ sig
$p \leq .01$ high sig
$p > .05$ not sig

to conclude that the null hypothesis we proposed is to be questioned the way we assume a quarter has both a heads and tails side until when there are so many times of flipping resulting in only heads, we must reject the null that it's a normal quarter, so too here we reject the null, $H_0: \beta_3 = 0$ to test for the statistical significance of the interaction term of $educ_i \times exper_i$.

edu	contin
exp	contin.
tenure	contin.
married	binary.
black	binary.
south	binary
urban.	binary

c) See log file attached for standa[...] regression.

The approximate difference in mor[...] married:

~~• The variable wage represents~~

~~• To hold all else constant is to [...] tinuous variables and the value of zero~~

~~$\log(wag_i) = 5.395497 + .0654307[\beta_1]$ ... $[\beta_3]$ 117473 $(7.234225) + .1994171[\beta_4](0)_{married}$ ... $(0)_{south}$ $+ (.1839121)(0)_{urban} = \log(wage_i)$~~

Note: from lecture #12 & reiiterated elsewhere,
log-linear:
$\log(Y_i) = \beta_0 + \beta_1 X_{i1} + \mu_i$
Slide #36
* a 1 unit increase in X_{i1} increases Y by $100 \times \beta_1$ percent.

∴ Therefore, the approximate difference in monthly salary is represented by the idea that a change from not married (where in binary = 0) to married (where in binary = 1) increases wages — monthly wages and not natural log of wages by $\beta_4 \times 100$ percent which is 19.9%.

The p-value of 0 is less than the critical point of .01 so we have enough evidence to reject the null where we assume insignificance ($H_0: \beta_4 = 0$). Therefore the difference in monthly salary is statistically significant.

d) So in the regression we utilize

reg lwage educ exper tenure south urban mb nmb nmnb, r

We purpose leave out one of the four generated variables (it can be any one of the four) as this will be the reference point for interpreting the remaining four variables. (See Slide 14/37 of LC#11)

Leaving out one variable protects our results from the dummy variable trap & ensures imperfect multicollinearity. *It's difficult to estimate the wage differential between married, nonblack people & nonmarried, non-black people when all 4 are included in the regression because their high correlation increases the liklihood of high standard error of our estimated coefficients. (Slide #15 of 37 of LC#11)

∴ Comparing mnb to nmnb, we exclude mnb to utilize our regression results to state that non-married, non black people earn 18.89% or approx. 19% less (due to ⊖ sign) than married, nonblack.

to conclude that the null hypothesis we proposed is to be questioned the way we assume a quarter has both a heads and tails side until when there are so many times of flipping resulting in only heads, we must reject the null that it is a normal quarter. So too here we reject the null, $H_0: \beta_3 = 0$ to test for the statistical significance of the interaction term of $educ_i \times exper_i$.

edu	contin
exp	contin.
tenure	contin.
married	binary.
black	binary
south	binary
urban.	binary.

c) See log file attached for standard form results of wage 2 data regression. The approximate difference in monthly salary between married & nonmarried:

~~• The variable wage represents monthy earnings.~~

~~• To hold all else constant is to utilize the average value for continuous variables and the value of zero for dummy variables.~~

~~$\log(wag_i) = 5.395497 + \overset{\beta_1}{.0654307}(13.46845) + \overset{\beta_2}{.014043}(11.56364) + \overset{\beta_3}{.0117473}(7.234225) + \overset{\beta_4}{.1994171}(\underset{married}{0}) + (-.1883499)(\underset{black}{0}) + (-.0909036)\underset{south}{(0)} + (-.1839121)\underset{urban.}{(0)} = \log(wage_i)$ is 6.52~~

incorrect!

∴ Therefore, the approximate difference in monthy salary is represented by the idea that a change from not married (where in binary = 0) to married (where in binary = 1) increases wages — monthly wages and not natural log of wages by $\beta_4 \times 100$ percent which is 19.9%.

The p-value of 0 is less than the critical point of .01 so we have enough evidence to reject the null where we assume insignificance ($H_0: \beta_4 = 0$). Therefore the difference in monthly salary is statistically significant.

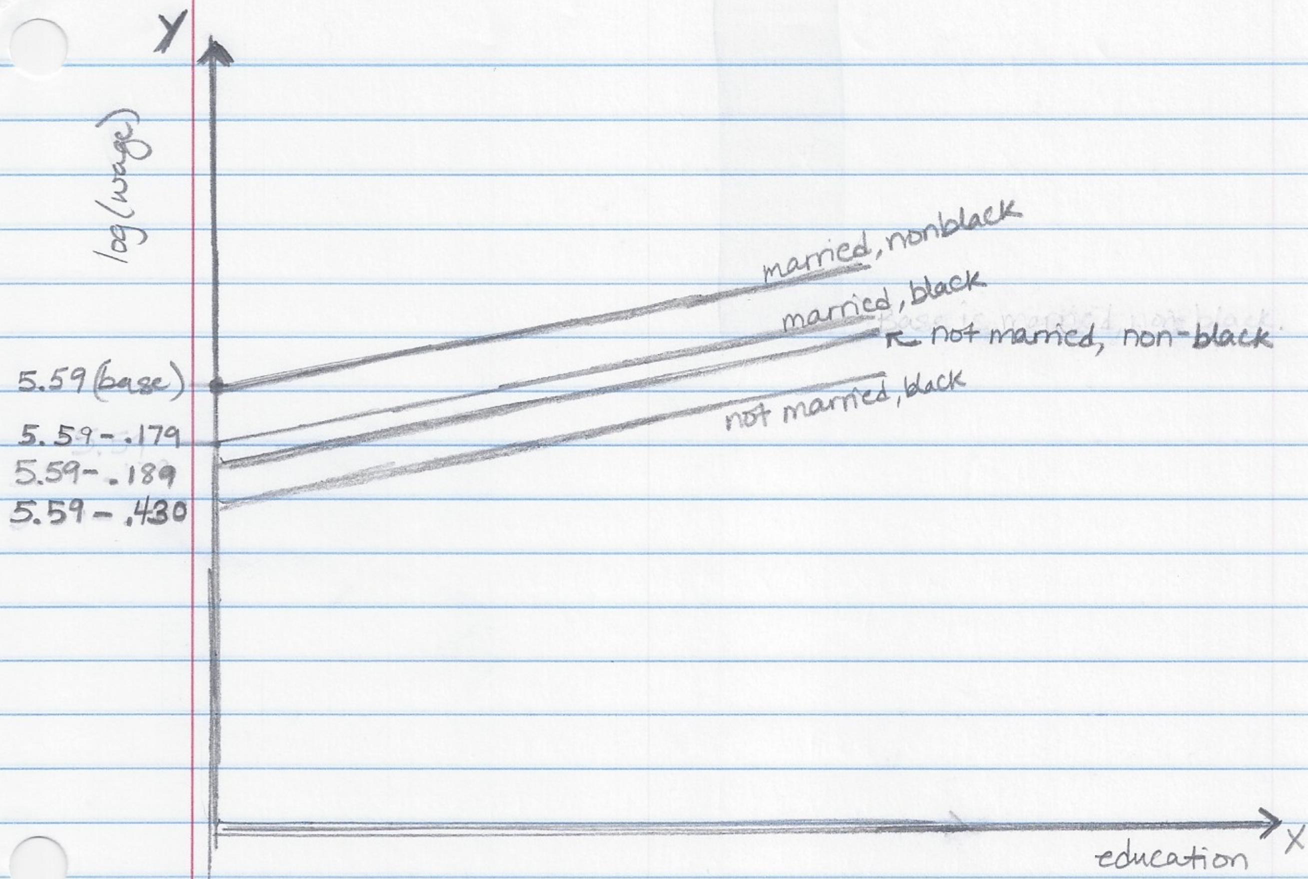
d) So in the regression we utilize

reg lwage educ exper tenure south urban mb nmb nmnb, r

We purpose leave out one of the four generated variables (it can be any one of the four) as this will be the reference point for interpreting the remaining four variables. (See Slide 14/37 of LC#11)

Leaving out one variable protects our results from the dummy variable trap & ensures imperfect multicollinearity. *It's difficult to estimate the wage differential between married, nonblack people & nonmarried, non-black people when all 4 are included in the regression because their high correlation increases the liklihood of high standard error of our estimated coefficients. (Slide #15 of 37 of LC#11)

∴ Comparing mnb to nmnb, we exclude mnb to utilize our regression results to state that non-married, non black people earn 18.89% or approx. 19% less (due to ⊖ sign) than married, nonblack.



See Lecture #17:

our slope doesn't change because the only change is the exchange of the values of our four dummy variables.

As discussed in (d) our married, nonblack component is the base value to how we structured our regression results. Since β_0 = the regression results of mnb, any deviation from this to any of the other 3 options utilizes the addition/subtraction of components.

Slide #3 of 18 on Lecture #17