Carson Konzak

ECN/PUB 480/580
Assignment #4
Due: Thursday March 17, 2022 by end of the day

Directions: Answer each question electronically in a MS Word or .pdf file. Compile your answers into a single computer file and then upload it to Canvas under "Assignment #4." Contact me if you have any questions.

Go to Canvas to download the data set entitled wage.dta from where this assignment is posted. This is a data set containing observations on wages and some other variables for 1,000 workers. We are going to look at what explains the wage a worker earns, whether females earn a lower wage than males, and if they do, then potentially why.

The data set includes the following variables:

wage: average hourly earnings, in dollars, for each individual
educ: number of years of education for each individual
age: age of each individual, in years
exper: number of years of work experience for each individual
marr: dummy variable taking the value of 1 if the individual is married, 0 otherwise
female: dummy variable taking the value of 1 if the individual is female, 0 otherwise

---

Start by opening the data set in Stata and generate 5 new variables using the `gen` command. To help you out, I give the command you need to use in parentheses.

a) the natural logarithm of wages (`gen lwage = log(wage)`)
b) experience squared (`gen expersq = exper^2`)
c) a new dummy variable for `male`, which takes on the value of 1 if the individual is a male and 0 otherwise (`gen male = 1-female`)
d) a new interaction dummy variable `marrfe` taking on the value of 1 if the individual is a married female, and 0 otherwise (`gen marrfe = marr*female`)
e) a new interaction dummy variable that will be equal to 1 if the individual is a married male, and 0 otherwise (`gen marrma = marr*male`)

The point of this assignment is to conduct a regression analysis to see if the "wage gap" between male and female workers persists after we start controlling for other variables, namely education and experience.

```
. gen lwage = log(wage)

. gen expersq = exper^2

. gen male = 1-female

. gen marrfe = marr*female

. gen marrma = marr*male
```

---

1. Create a table in MS Word (with a title and description) that summarizes the data by gender. That is, give the average wage, average education, average experience, and average age separately for men and women. You can use the `sum` command along with an `if` statement to do this (no commas). For example, `if female == 1` means only females. Use proper units for the variables in your summary statistics and round to two places past the decimal. Do NOT use the Stata output as your table! **(3 points)**

Summary Statistics by Gender

Gender	Average Wage ($/hr)	Average Education (yr)	Average Experience (yr)	Average Age (yr)
Male	11.95	12.34	8.24	34.78
Female	10.26	12.79	7.73	36.21

---

2. Consider the following regression:

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 expersq + \beta_4 female + u_i$$

This regression is a log-linear regression. This was "case 2" for using natural logarithms in regression in Lecture #12. Estimate this regression using Stata's `reg` command. Include your output along with the command used to generate it.

```
. reg lwage educ exper  expersq  female
```

```
      Source |       SS           df       MS      Number of obs   =     1,000
-------------+----------------------------------   F(4, 995)       =     82.87
       Model |  57.9794329         4  14.4948582   Prob > F        =    0.0000
    Residual |  174.036045       995  .174910598   R-squared       =    0.2499
-------------+----------------------------------   Adj R-squared   =    0.2469
       Total |  232.015477       999  .232247725   Root MSE        =    .41822

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0748992   .0052042    14.39   0.000     .0646868    .0851117
       exper |   .0317794   .0033939     9.36   0.000     .0251193    .0384394
     expersq |  -.0004707   .0000762    -6.18   0.000    -.0006203   -.0003211
      female |  -.1365068    .026626    -5.13   0.000    -.1887564   -.0842572
       _cons |   1.215849    .069698    17.44   0.000     1.079077    1.352621
------------------------------------------------------------------------------
```

---

a. Interpret the $\hat{\beta}_1$ and $\hat{\beta}_2$ from each equation. Based on the coefficients, how is the independent variable related to the dependent variable? In particular, how will the dependent variable change if the independent variable increases by 1 unit? Be specific and use specific units! Refer to Case 2 for the interpretation of a $\hat{\beta}$ for a log-linear regression. **(3 points)**

$\hat{\beta}_1 = 0.0749 \Rightarrow \Delta \log(wage) = 0.0749 \cdot \Delta \text{ education} \Rightarrow \Delta wage = e^{0.0749 \cdot \Delta \text{Education}}$

$\hat{\beta}_2 = 0.0318 \Rightarrow \Delta \log(wage) = 0.0318 \cdot \Delta \text{ experience} \Rightarrow \Delta wage = e^{0.0318 \cdot \Delta \text{Experience}}$

From Lecture

$100 \cdot \hat{\beta}_1 = 100 \cdot \frac{\Delta \log(wage)}{\Delta \text{Education}} = 7.49 \Rightarrow$ gaining 1 year of education is expected to increase wage by 7.49%

$100 \cdot \hat{\beta}_2 = 100 \cdot \frac{\Delta \log(wage)}{\Delta \text{Experience}} = 3.18 \Rightarrow$ Experience 3.18%

---

b. Explain the signs on the coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$. What do the signs indicate about the relationship between *exper* and *lwage*? **Hint:** refer Lecture #13, which covers using quadratic terms in regression. Is the relationship between *lwage* and *exper* a straight linear relationship or does the relationship taper over time? Why do you think this is the case? **(3 points)**

$\Delta \log(\text{wage}) = \Delta \text{experience} \cdot (0.0317 - \Delta \text{EXPERIENCE} \cdot 0.0005)$

$\Rightarrow \quad \Delta \text{wage} = e^{\Delta \text{EXPERIENCE}(.0317 - \Delta \text{EXPERIENCE} \cdot .0005)}$

From Lecture

$100 \cdot \dfrac{\Delta \log(\text{wage})}{\Delta \text{Experience}} = (0.0317 - \Delta \text{Experience} \cdot 0.0005) \cdot 100$

The signs indicate that experience initally increases wage & Then reaches A turning point such that it then decreases wage.

↳ The relationship in non-linear, it is quadratic.

↳ Eventually Experience hurts an employee; the company can find better talent for the same price; or equal talent for a lower price. Thus as capatilist do, they cut the employees wage to maximize profits.

---

c. Interpret $\hat{\beta}_4$, which is the estimated coefficient on the *female* dummy variable. Recall that the estimated coefficients on dummy variables have an "if-then" interpretation. In other words, *if* the dummy variable equals 1, *then* the y-variable changes by the associated $\hat{\beta}$. Recall that this is a log-linear regression, so the interpretation is that if the dummy variable equals 1, then the y-variable changes by $\hat{\beta}_4$x100%. $\hat{\beta}_4$x100% would thus be the "wage gap," since it says how much more or less someone earns, in percentage terms, if that person is female. **(3 points)**

$\hat{\beta}_4 = -0.1365 \Rightarrow \Delta \log(\text{wage}) = -0.1365 \cdot \delta \text{ female}$

$\Rightarrow \Delta \text{ wage} = \dfrac{1}{e^{0.1365 \cdot \delta \text{ female}}}$

From Lecture

$\dfrac{\Delta \log(\text{wage})}{\delta \text{ female}} \cdot 100 = -0.1365 \cdot 100 = -13.65$

$\Rightarrow$ Being A Female is predicted to decrease your wage by 13.65 %

Eg: If Female; then wage decreases 13.65%

---

3. a. Which of the estimated coefficients (that is, the $\hat{\beta}_s$) are statistically significant and why? Recall that statistically significant means you can reject the null hypothesis that a particular $\hat{\beta}$ is equal to zero. Explain your answer referring to the t-statistics, critical values, and use a 5% level of significance. You can either use the t-statistics in the Stata output or calculate them yourself. You can ignore the constant ($\hat{\beta}_0$). Use a two-tailed test. **(3 points)**

$\hat{\beta}_i$	|t|
Education	14.39
Experience	9.36
Experience Squared	6.18
Female	5.13

Conducting a 2-tailed hypothesis test with a 95% confidance interval & assuming gaussian distribution we have a CRITICAL VALUE of 1.96, & all $\hat{\beta}_i$ are statistically significant.

Eg; ALL ASSOCIATED |t-values| are ≥ the critical value.

---

b. Explain what the R^2 and F-statistics mean for this regression. That is, for the F-statistic, do you reject the H_0 for the F-test? **(2 points)**

$R^2 = 0.2499$ → 25% of the variation in wage is explained by education, experience, experience^2, & gender.

$F = 82.87$; $F_{4,1000}$ @ 90% Confidence = 3.76

↳ reject H_0 for the F-test as $|82.87| \geq 3.76$

F-STATISTIC is testing the possibility all x's are irrelevant, eg $\hat{\beta}_i = 0$ I

---

4. Estimate the regression for only single people. Do this by including the statement `if marr == 0` at the end of the `reg` command (no commas). Is the wage gap (as given by the $\hat{\beta}$ on the variable *female*) statistically significant? Compare the $\hat{\beta}$ *female* in this case to that in question #2. Is the "wage gap" for single women more or less than for women in general? How much more or less? **(3 points)**

```
. reg lwage educ exper  expersq  female if marr == 0

      Source |       SS           df       MS      Number of obs   =       469
-------------+----------------------------------   F(4, 464)       =     30.53
       Model |  19.3050208         4   4.8262552   Prob > F        =    0.0000
    Residual |  73.3502891       464   .15808252   R-squared       =    0.2084
-------------+----------------------------------   Adj R-squared   =    0.2015
       Total |  92.6553099       468  .197981431   Root MSE        =     .3976

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0704699   .0079729     8.84   0.000     .0548025    .0861373
       exper |   .0316466   .0047075     6.72   0.000     .0223961    .0408972
     expersq |  -.0004615   .0000974    -4.74   0.000    -.0006529   -.0002702
      female |  -.0191568   .0369228    -0.52   0.604    -.0917134    .0533999
       _cons |   1.122722   .1055533    10.64   0.000        .9153    1.330143
------------------------------------------------------------------------------
```

The wage gap is not statistically significant with a p-value of 0.6

$\hat{\beta}_{\#2} = -0.1365$

$\hat{\beta}_{\#4} = -0.0192$

If considering only single individuals, the wage gap between men & women is (-0.0192+0.1365) 11.73 percentage points smaller than that of for men & women regardless of their relationship status.

$r_{\#4} = -0.0192$ points smaller than that of for men & women regardless of their relationship status.

---

5. Estimate the regression for only people who are single and less than 30 years old. Do this by including the statement `if marr == 0 & age < 30` at the end of the `reg` command (no commas). Given how young this subsample is, **do not** include the variable *expersq*. Include the variable *female* in both equations and interpret the results as you did in question #4. Is the wage gap statistically significant in this case? Is the "wage gap" for single women more or less than for single women in general? For women in general? How much more or less? Why do you think this is? **(3 points)**

```
. reg lwage educ exper female if marr == 0 & age < 30
```

Source	SS	df	MS	Number of obs	=	260
Model	2.2684138	3	.756137934	F(3, 256)	=	7.60
Residual	25.4646975	256	.099471475	Prob > F	=	0.0001
				R-squared	=	0.0818
Total	27.7331113	259	.10707765	Adj R-squared	=	0.0710
				Root MSE	=	.31539

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0427246	.0093576	4.57	0.000	.0242969	.0611524
exper	-.001074	.0026475	-0.41	0.685	-.0062876	.0041396
female	.0066582	.0399923	0.17	0.868	-.0720976	.0854141
_cons	1.462023	.1170881	12.49	0.000	1.231444	1.692601

Without Experience Squared, The Wage Gap is not statistically significant for single individuals & for single individuals under 30.

The Wage Gap for single women under 30 is 0.66%, meaning single men under 30 are expected to mak 0.66 percentage points less.

The Wage Gap for single women in general is -1.92%, & for women in general is -13.65%

The Absolute Gap is decreased, it is (1.92-.66) 1.26 percentage points closer to Ø than general single women & (13.65-.66) 12.99 percentage points closer to Ø than general women alone.

---

1 point extra credit: Refer to the statement "Given how young this subsample is, **do not** include the variable *expersq*" from question 11. What do I mean by this statement? That is, why would including the square of experience be inappropriate here? You can answer this in just one or two sentences.

The turning point phenomena does not appear until ages later than 30. Thus we are better of using a linear model for young adults & experience.

---

6. Estimate the regression, including *expersq*, for everyone (e.g. don't include the "if" statement), but include both the *female* dummy variable and the *marrfe*.

```
. reg lwage educ exper  expersq  female marrfe
```

Source	SS	df	MS		
Model	58.4509962	5	11.6901992	Number of obs	= 1,000
Residual	173.564481	994	.174612154	F(5, 994)	= 66.95
Total	232.015477	999	.232247725	Prob > F	= 0.0000
				R-squared	= 0.2519
				Adj R-squared	= 0.2482
				Root MSE	= .41787

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0747913	.0052002	14.38	0.000	.0645867	.0849958
exper	.0312317	.0034074	9.17	0.000	.0245453	.0379182
expersq	-.000461	.0000764	-6.03	0.000	-.0006109	-.0003111
female	-.1689819	.0331398	-5.10	0.000	-.234014	-.1039499
marrfe	.0599392	.0364736	1.64	0.101	-.0116348	.1315133
_cons	1.220307	.0696913	17.51	0.000	1.083548	1.357066

---

a. Is the $\hat{\beta}$ on the interaction term statistically significant? How do you know **(2 points)**

Under A 2-tailed test, the $\hat{\beta}_{marrfe}$ is not statistically significant at a 90% Confidance level.

We know this as $1-p = 1-0.101 = 0.899 < 0.90$

Similarly $|1.64| < 1.646$, eg. t-statistic is not statistically significant.

---

b. Interpret the $\hat{\beta}$ on the interaction term. Does being married increase or decrease a female worker's wage? How much? **(3 points)**

$\hat{\beta}_{marrfe} = 0.0599 \Rightarrow \Delta \log(wage) = 0.0599 \cdot \delta\, marrfe$

$\Rightarrow \Delta wage = e^{.0599 \cdot \delta marrfe}$

From Lecture

$100 \cdot \hat{\beta}_{marrfe} = \frac{\Delta \log(wage)}{\delta\, marrfe} \cdot 100 = 5.99 \Rightarrow$ Being both married & female is expected to increase your wage by 5.99%

---

c. Compare the $\bar{R}^2$ if you have the interaction term in the regression versus if you don't. What does this mean about the interaction term? **(2 points)**

$\bar{R}^2_{w/\, marrfe} = 0.2482$

$\bar{R}^2_{w/o\ marrfe} = 0.2469$

As our $\bar{R}^2$ increases we say that the interaction term explains some of the variance is % change of wage & thus keep it in the regression :)