Labor Economics, Spring 2021

Problem Set 2
Due Friday, April 30 in class

<u>NOTE</u>: For questions using Stata, please type your answers into a word processing document (using MS Word or something similar). Insert the relevant part of your Stata log file into your document, and clearly explain how your Stata output answers the question.

1. (10 points) (Labor Demand)

   a. Why is the short-run demand for labor downward sloping? Why is the long-run demand for labor downward sloping? Which is more elastic? Why? (5 points)

      The short-run labor demand curve is downward sloping because of the diminishing marginal product of labor. In the short run, capital is fixed, so there is no room for substitution between capital and labor. That is, the short-run labor demand elasticity includes only the scale effect. In contrast, in the long run firms can substitute capital for labor and further decreases employment. Therefore, the labor demand should be more elastic in the long run.

   b. We discussed four factors that determine the elasticity of labor demand (attributed to Marshall). Provide an intuitive explanation for two of the four factors. (5 points)

      See page 43 of Chapter 4's slides.

2. (6 points) Throughout this class, we're going to be using Stata to calculate regression lines, but it is probably a good idea for everyone to compute one regression line by hand. Here's the data:

OBS	*Y*	*X*
1	16	10
2	12	7
3	8	4
4	4	4
5	20	10

   In a regression of *Y* on *X*,

   a. What is your estimate of the slope parameter? (2 points)
   b. What is your estimate of the intercept? (2 points)
   c. Give an explicit interpretation of both the slope and the intercept. (2 points)

Our estimate of the slope parameter is $\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$**.**

Note that the sample mean of *X* is 35/5 = 7, and the sample mean of *Y* is 60/5 = 12, so I can construct a new table:

OBS				
1	4	3	9	12
2	0	0	0	0
3	-4	-3	9	12
4	-8	-3	9	24
5	8	3	9	24

So, our estimate of the slope parameter is 72/36 = 2. Our estimate of the intercept is $\bar{Y}-\hat{\beta}_1\bar{X}$ **= 12 – 7×2 = –2. What these mean is that the estimated average of *Y* is -2 when *X* equals zero, and the estimated effect of a one-unit increase in *X* on the average of *Y* is 2.**

3. (19 points) Using the NLSY data set, run a regression of the log wage on *male*, a “dummy variable” that’s equal to 1 for men and 0 for women (type “*reg lwage male*” into Stata…you will have to create *lwage* as in Problem Set 1). You are estimating this model:

$$lwage = \beta_0 + \beta_1 male + u$$

   a. Interpret the estimate of the slope and the intercept, being as specific as possible – make sure that your interpretation recognizes the fact that *male* is a dummy variable. Is the sign of the slope what you would expect? (3 points)

```
. reg lwage male

      Source |       SS           df       MS      Number of obs   =     5,898
-------------+----------------------------------   F(1, 5896)      =    146.87
       Model |  69.3584016         1  69.3584016   Prob > F        =    0.0000
    Residual |  2784.36121     5,896  .472245795   R-squared       =    0.0243
-------------+----------------------------------   Adj R-squared   =    0.0241
       Total |  2853.71961     5,897  .483927354   Root MSE        =     .6872

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .2171958    .017922    12.12   0.000     .1820621    .2523294
       _cons |   1.689189   .0130078   129.86   0.000     1.663689    1.714689
------------------------------------------------------------------------------
```

The slope here is 0.217, which means that each one-unit increase in the variable “male” increases the average log wage by 0.217. Put another way, on average, we estimate that males earn approximately 21.7% more than females.

The constant is 1.6891, which is the estimated average value of lwage when male=0, i.e., for females.

b. Suppose you want to use the regression in (a) to test whether gender is related to *lwage*. As specifically as possible in terms of notation, what are the null and alternative hypotheses you want to test? At a significance level of 0.05, do you reject the null hypothesis (why or why not?), and what do you conclude (**HINT**: you will perform this test using just the regression output)? (3 points)

The null hypothesis is that the slope in the regression above, which is β_1, is equal to zero. The alternative is that the slope is not equal to zero. Since our p-value of 0.000 is less than 0.05, we reject the null hypothesis and conclude that gender is related to *lwage*.

c. What is the R^2 of this regression, and how do you interpret it? (2 points)

The r^2 of this regression is 0.0243, meaning that 2.43 percent of the variation in wages is due to variation in gender.

d. Now you decide that the true model is

$$lwage = \beta_0 + \beta_1 male + \beta_2 school + u$$

If you run this regression in Stata ("*reg lwage male school*"), what is your estimate of β_1 now? What is the interpretation of this estimate? WHY is it bigger than the estimate you got from part a (be as specific as possible)? (4 points)

```
. reg lwage male school

      Source |       SS           df       MS      Number of obs   =     5,898
-------------+----------------------------------   F(2, 5895)      =    332.77
       Model |  289.496622         2  144.748311   Prob > F        =    0.0000
    Residual |  2564.22299     5,895  .434982695   R-squared       =    0.1014
-------------+----------------------------------   Adj R-squared   =    0.1011
       Total |  2853.71961     5,897  .483927354   Root MSE        =    .65953

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .2529257   .0172735    14.64   0.000     .2190632    .2867881
      school |   .0817576   .0036343    22.50   0.000     .0746331    .0888821
       _cons |   .6302659   .0486983    12.94   0.000     .5347995    .7257324
------------------------------------------------------------------------------
```

Now our estimate of the slope on male is 0.2529, which means that, on average, we estimate that males earn approximately 25.3% more than females, HOLDING EDUCATION CONSTANT. To get full credit for this, all you had to say was that it is answering a different question than the estimate from the simple regression (or something similar). In the future, though, we'll want to be a lot more precise about this. For example, this estimate is bigger than the estimate we got from part a because the estimate in part a had a negative bias – in particular, we left out the important variable *school,* which is negatively correlated with *male* in the sample and positively correlated with *lwage*. Since a negative number times a positive number is a negative number, our estimate from the simple regression had a negative bias.

Another way to think about this is that males would earn 25.3% more than females if

they had the same amount of education. In fact, they have LESS education, which is why they only earn 21.7% more.

e. What is your estimate of β_2, and what is your interpretation of that estimate? (2 points)

Our estimate of the slope on school is 0.082, which means that, on average, we estimate that each year of education increases earnings by 8.2%, HOLDING GENDER CONSTANT.

f. What is the R^2 of this regression, and how do you interpret it? Why is it bigger than the R^2 from the model in part c? (3 points)

The r-squared here is 0.1014, which means that 10.14% of the variation in log(wage) can be explained due to variation in the variables school and male. This is bigger than the r-squared from the simple regression because the r-squared NEVER goes down when you add a variable. In this case, it went up because school explains some variation in lwage that male fails to explain.

g. Based on the estimates from this regression, what is our estimate of the population mean of *lwage* among men with 12 years of education (**HINT**: use the Stata output from the regression)? How about for women with 12 years of education? (2 points)

These are just predicted values. A man with 12 years of education has *male*=1 and *school*=12, so the predicted value of log(wage) in that case is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(12)$$

= .630 + .253 + .0817 * 12
= 1.8634

Similarly, a woman with 12 years of education has *male*=0 and *school*=12, so the predicted value of log(wage) in that case is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(12)$$

= .630 + 0 + .0817 * 12
= 1.6104