IAC801. Homework, Due: April 17, 7pm

April 8, 2020

This is the first problem set. Also please submit weekly Stata assignment along with answers for this problems if you haven't submitted them yet.

Qestion 1. Simple regression model (15 points, each 1 point unless indicated otherwise)

This question is designed to use a simple regression model. Consider the following population model:

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

where, in a most typical example of economic applications, y is outcome of interest, x is the main factor that affected by a policy variable, and u includes all other relevant factors in the determination of y. Thus, mathematical relationship between outcome and policy is approximated by the model of equation (1).

(1) Propose your own research question that could be potentially studied by using equation (1). Also write your research question as a hypothesis form. [Note: If you have your own dataset, please estimate your model using a simple regression model.]

(2) List potential values of y and x could take in practice. Imagine you collect data on y and x. Identify your target population and propose a realistic sample you can have an access.

(3) Suppose that you estimate the population model of (1) using the sample at hand and OLS estimation method.

(4) Interpret the results. Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$ in your model.

(5) List some potential factors that could be in u.

(6) Can you trust $\hat{\beta}_1$ estimate you obtained from OLS estimation (e.g., reg y x in Stata)? Justify your answer? What are four assumptions to justify your result is valid.

```
. reg y x

      Source |       SS       df       MS              Number of obs =     250
-------------+------------------------------           F(  1,   248) =  195.30
       Model |  6628.96683     1  6628.96683           Prob > F      =  0.0000
    Residual |  8417.89056   248  33.9431071           R-squared     =  0.4406
-------------+------------------------------           Adj R-squared =  0.4383
       Total |  15046.8574   249  60.4291461           Root MSE      =  5.8261

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    1.78483   .1277173    13.97   0.000     1.533281    2.036379
       _cons |   3.046159   .3701413     8.23   0.000     2.317137     3.77518
------------------------------------------------------------------------------
```

(7) The above table shows an example of estimation results from simple regression model. Graphically describe the regression result.

(8) Describe about std. Err. of $\hat{\beta}_1$ and sampling distribution of $\hat{\beta}_1$ in your context in question (1) and (2). (2 points)

(9) Also provide a theorem of Central Limit Theorem (CLT) on a random variable x. And extend this CLT result to $\hat{\beta}_1$. (2 points)

(10) Interpret the R-squared value in the above table.

(11) Interpret the p-value for coefficient on _cons.

(12) Explain 95% conf. interval of $\hat{\beta}_1$.

(13) Identify the elements used in the construction of t-statistic testing the null hypothesis.

Qestion 2. OLS estimator properties (15 points, each 1 point unless indicated otherwise)

(1) What is an estimator? What is the difference between parameters and estimates?

(2) What are the two criteria we use to **evaluate** the properties of a certain estimator?

(3) Provide conditions for the OLS estimator to be unbiased.

(4) Which assumption is the most critical one in practice? Also provide definition for correlation and causal relation. Why distinction between these two are important?

(5) Identify the potential **sources** of violation of assumption that discussed in (4).

```
. reg lwage educ

      Source |       SS       df       MS              Number of obs =     759
-------------+------------------------------           F(  1,   757) =  150.00
       Model |  43.4598948     1  43.4598948           Prob > F      =  0.0000
    Residual |     219.334   757  .289741084           R-squared     =  0.1654
-------------+------------------------------           Adj R-squared =  0.1643
       Total |  262.793895   758  .346693793           Root MSE      =  .53828

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0992929   .0081073    12.25   0.000     .0833773    .1152084
       _cons |   1.142179   .1090192    10.48   0.000     .9281635    1.356195
------------------------------------------------------------------------------
```

(6) What is the return of education estimated in the Table?

(7) Find the predicted lwage for high school graduate. Also find the predicted lwage for collage graduate. Find the average return of having college education.

(8) How reliable the estimate you calculated in (7)? Why? There could be multiple answers here. Justify your answer as much as you can. (4 points)

(9) Interpret the R-squared value.

(10) What is the condition, besides those conditions in (3), for the precision of $\hat{\beta}_1$ estimator? Is this condition reasonable? Graphically describe this condition with a figure. (2 points)

(11) Why do you take log for dependent variable.

Qestion 3. Motivation for multiple regression (15 points, each 1 point unless indicated otherwise)

Consider an extension of the log(wage) equation we used for simple regression:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + u$$

where IQ is IQ score (in the population, it has a mean of 100 and sd 15).

(1) Why do we want to include IQ?

(2) Interpret β_0, β_1, and β_2.

(3) Provide conditions for the OLS estimator for β_1 to be unbiased.

(4) What are/is the key assumption for the OLS estimator for β_1 in practice?

(5) Provide definition for ceteris paribus condition.

(6) Also interpret the following equation. $E(u|educ, IQ) = 0$. (2 points)

(7) In the estimation and interpretation of $\hat{\beta}_1$, using ceteris paribus condition interpret the slope coefficients. Justify you answer using the formula in the lecture. (3 points)

(8) True or False: The beauty of multiple regression is that it gives us the ceteris paribus interpretation without having to find two people with the same value of IQ who differ in education by one year. Justify your answer.

(9) The relationship between Simple and Multiple Regression Estimates:

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \tag{2}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{3}$$

where a tilda denote the simple regression and hat denotes multiple regression. Both models estimate parameters using the same n observations. Show that $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$ where $\tilde{\delta}_1$ be the slope from the regression x_2 on x_1.

(10) Use the regression result (reg lwage educ) table above and the following result, identify all elements of $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$ where $x_1 = educ$ and $x_2 = IQ$.

(11) Verbally explain what these terms in the equation$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$ imply.

(12) In your research question in 1-(1), discuss the same problem could arise in your context. (Note: For the rest of semester, one of our main tasks are to learn variety of methods to address this particular problem properly.)

```
. reg lwage educ IQ

      Source |       SS       df       MS              Number of obs =     759
-------------+------------------------------           F(  2,   756) =   88.30
       Model |  49.7567642     2  24.8783821           Prob > F      =  0.0000
    Residual |   212.99188   756  .281735291           R-squared     =  0.1894
-------------+------------------------------           Adj R-squared =  0.1872
       Total |  262.748644   758  .346634095           Root MSE      =  .53079

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0728327   .0097582     7.46   0.000     .0536763    .0919892
          IQ |   .0076329   .0016143     4.73   0.000     .0044639    .0108019
       _cons |   .7284623   .1386158     5.26   0.000     .4563446     1.00058
------------------------------------------------------------------------------

. corr educ IQ
(obs=759)

             |     educ       IQ
-------------+------------------
        educ |   1.0000
          IQ |   0.5734   1.0000
```

Qestion 4. Application in International Trade (15 points)

In this question, we study the effect of exporting (x) on firms' productivity(y) using the paper titled as "Exporting and Firm Performance: A Randomized Experiment", Quarterly Journal of Economics, Volume 132, Issue 2, May 2017, Pages 551–615.

(1) What is two central challenges to identify the causal effect of exporting on firm performance? Explain as much details as possible. (3 points)

$$y = \beta_0 + \beta_1 x + \beta_2 y_{initial} + u$$

where x is 1 is an Indian firm is **randomly** provided with an opportunity to export their rug to US/Europe and 0 if an Indian firm is not provided with an opportunity; y is firm's productivity measure before export opportunity becoming available, $y_{initial}$.

(2) What is the condition we have to impose to justify OLS estimation result? (2 points)

(2) What is implementation method? Describe implementaion process in Stata. (2 points)

(3) Describe ceteris paribus condition and interpret β_1 using this condition. (3 points)

(4) Interpret the following result table that estimating the above equation as much as you can. (5 points)

a. Interpret coefficient on indicator for ever exported.

b. Interpret the R-squared value.

c. Perform a hypohesis test.

d. What is the t-value?

e. Provide 95% confidence interval for the coefficient on "indicator for ever exported"

IMPACT OF INTERVENTION ON FIRMS KNOWINGLY EXPORTING

	(1) ITT
Indicator for ever exported	0.55***
	(0.06)
R-squared	0.33
Control mean	0.20
Observations	191