B.E. International Program

Faculty of Economics, Thammasat University

Midterm Examination: 2/2020
Subject: EE325 Introductory Econometrics (section 046403)
Instructor: Ajarn Weerawat Phattarasukkumjorn
Date: Tuesday 16 March, 2020 **Time: 12.00 – 14.00 hrs.**

Seat No……………………….
ID.No…………………………

Instructions

1. The exam has 3 pages including this page.
2. There are 3 main questions. **ANSWER ALL QUESTIONS with 4 decimal places** in answer sheet. Every question should be answer with sufficient details.
3. Total score is 30 points (30% of your letter grade)
4. A simple, scientific or financial calculator is allowed.
5. Students must answer in blue or black ink.
6. One A-4 paper that students can write anything on both sides (No photocopy, no post it or staple on top of the paper)
7. Statistics tables are provided.

For all questions, answer up to 4 decimal places.

1. (15 points) Given this information,

$n = 46$	$\sum_{i=1}^{n} X_i = 3{,}959.80$	$\sum_{i=1}^{n} Y_i = 3{,}180.80$
$\bar{X} = 86.0826$	$\bar{Y} = 69.1478$	
$\sum_{i=1}^{n} (X_i)^2 = 364{,}023.30$		$\sum_{i=1}^{n} X_i Y_i = 319{,}943.18$
$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 23{,}153.3861$		$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 94{,}525.1748$
$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 46{,}131.6183$		$\sum_{i=1}^{n} \hat{u}_i^{\,2} = 2{,}610.9211$

answer the following questions. Show your work.

a) (4 points) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID(0, \sigma^2)$, find the estimators of β_1 and β_2 with OLS method and explain the meaning of the model.

> $\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{46{,}131.6183}{23{,}153.3861} = \mathbf{1.9924}$

> $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 69.1478 - (1.9924)86.0826 = \mathbf{-102.3632}$

Then, our SRF is $\hat{Y}_i = -102.3632 + 1.9924 X_i$ meaning that the intercept is -102.3632 and the slope of this function is 1.9924.

b) (2 points) Find R^2 and explain its meaning.

> $r^2 = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^{\,2}}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = 1 - \frac{2{,}610.9211}{94{,}525.1748} = 1 - 0.0276 = \mathbf{0.9724}$

The r^2 is a measure of 'goodness of fit' or the measure of the proportion of the total variation in Y explained by the regression model. In this case, 97.24% of Y is explained by the regression model.

c) (1 points) If $X_i = 60$, estimate the value of $\hat{Y}_i$ and explain its meaning.

From the SRF, $\hat{Y}_i = -102.3632 + 1.9924 X_i$, replacing X_i with 60, we get,

> $\hat{Y}_i = -102.3632 + 1.9924(60) = \mathbf{17.1808}$

which means that when $X_i = 60$, the average of $\hat{Y}_i$ will be 17.1808.

d) (3 points) Find the estimators of $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$

> $\text{var}(u_i) = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^{\,2}}{n-k} = \frac{2{,}610.9211}{46-2} = \mathbf{59.3391}$

> $\text{var}(\hat{\beta}_1) = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (X_i)^2}{n\sum_{i=1}^{n} (X_i-\bar{X})^2}\hat{\sigma}^2 = \frac{364{,}023.30}{(46)23{,}153.3861} 59.3391 = \mathbf{20.2814}$

> $\text{var}(\hat{\beta}_2) = \hat{\sigma}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (X_i-\bar{X})^2} = \frac{59.3391}{23{,}153.3861} = \mathbf{0.0026}$

e) (2.5 points) What are the 95-percent confident intervals for β_2? Interpret the meaning.

> $\text{P}\left[\left(\hat{\beta}_2 - t_{\frac{\alpha}{2}} \cdot \text{se}(\hat{\beta}_2)\right) \le \beta_2 \le \left(\hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot \text{se}(\hat{\beta}_2)\right)\right] = 1 - \alpha$

The values that we need to figure out are:

> $\alpha = 0.05$ as determined by the question. Then, $1 - \alpha = 0.95$ or 95%.

> $\text{se}(\hat{\beta}_2) = \sqrt{\text{var}(\hat{\beta}_2)} = \sqrt{0.0026} = 0.0510$

> $t_{\frac{\alpha}{2}} = 2.021$ taken from the t-table. (Note: the degrees of freedom is n-k or 46 – 2 = 44 or roughly 40 as 44 is not available in the stat table)

Plug all the values into the first equation, we get

> $\text{P}[(1.9924 - (2.021 \text{ x } 0.051)) \le \beta_2 \le (1.9924 + (2.021 \text{ x } 0.051))] = 0.95$

> $\mathbf{P[1.8893 \le \beta_2 \le 2.0955] = 0.95}$

which means that 95 out of 100 times, this confidence interval will contain the true β_2.

f) (2.5 points) Test the hypothesis whether coefficients are different from zero at 0.05 level of significance.

Testing $\boldsymbol{\beta_1}$: set up the null and alternative hypothesis as

> $\text{H}_0\text{: } \beta_1 = 0$ and $\text{H}_\text{a}\text{: } \beta_1 \ne 0$

> Compute $se(\hat{\beta}_1) = \sqrt{\text{var}(\hat{\beta}_1)} = \sqrt{20.2814} = 4.5035$

> Compute $t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{-102.3632-0}{4.5035} = -22.7297$

> Find the critical values when $\alpha = 0.05$, then $t_{\frac{\alpha}{2}} = \pm 2.021$.

> Concluding the test: t_{cal} is beyond the critical value, therefore, we can **reject the null hypothesis**. In other words, we can make sure that β_1 is not zero 95 out of 100 times.

Testing $\boldsymbol{\beta_2}$: set up the null and alternative hypothesis as

> $\mathrm{H_0}$: $\beta_2 = 0$ and $\mathrm{H_a}$: $\beta_2 \neq 0$

> Compute $t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{1.9924 - 0}{0.051} = 39.0667$

> Find the critical values when $\alpha = 0.05$, then $t_{\frac{\alpha}{2}} = \pm 2.021$.

> Concluding the test: t_{cal} is beyond the critical value, therefore, we can **reject the null hypothesis**. In other words, we can make sure that β_2 is not zero 95 out of 100 times.

2. **(8 points)** Answer the following question without any mathematical proof. A 3-6-line paragraph for a question is sufficient.

   a) (2 points) If we have only one data point, can we create a sample regression function? Why?

We simply cannot create an SRF due to the fact that there is no other reference point(s) at all. Therefore, a single point of data cannot settle how much the slope would be since all the slopes are possible, leading to infinite answers of $\hat{\beta}_2$ and also $\hat{\beta}_1$.

   b) (2 points) Does a significant β_2 sufficient for us to believe that X and Y are causally related? Provide an example to support your answer.

No. Regression relation only reveals statistical relevance between two variables or more but does not provide any rationale underneath the significance. For instance, raining day might be accidentally the same day when SET index drops but there is no theory or any intuition suggest that these two variables are connected. To conclude that X and Y are causally related due to the significance might be too risky since we might have a spurious regression.

   c) (2 points) When we test a hypothesis and find that β_2 is significantly different from zero, what does the result actually suggest? Answer with specific wording from statistical perspective.

The only interpretation is that we can make sure that, according to our estimation $(1-\alpha)\%$ of the time, the true β_2 will not be zero. We might expect that on average when X increase for 1 unit, Y will change by $\hat{\beta}_2$ units. However, note that the wording here is "expect", not that we can be so sure of X causing Y.

   d) (2 points) What is (are) (an) advantage(s) of an interval estimation over point estimate?

Since our $\hat{\beta}_2$ is estimated from only a sample, this might not be the exact value of the true β_2. Nevertheless, multiple assumptions that we imposed allow us to conclude that the expected value of $\hat{\beta}_2$ is β_2 and β_2 is normally distributed. Interval estimation, based on t-distribution since we do not know the true variance, **allow us to predict with a margin of error** with an exact probability of each interval of values.

3. **(7 points)** Given that the dependent variable is natural log of wage (lwage) in Thai Baht and the independent variable is hours worked per week (main_hr), the result of estimation is shown in the table below here.

```
      Source |       SS           df       MS      Number of obs   =       308
-------------+----------------------------------   F(1, 306)       =     92.20
       Model |   50.060869         1   50.060869   Prob > F        =    0.0000
    Residual |  166.152715       306  .542982728   R-squared       =    0.2315
-------------+----------------------------------   Adj R-squared   =    0.2290
       Total |  216.213584       307  .704278775   Root MSE        =    .73687

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     main_hr |   .0318017    .003312     9.60   0.000     .0252844    .0383189
       _cons |   7.658082   .1256392    60.95   0.000     7.410856    7.905308
------------------------------------------------------------------------------
```

Answer the following questions. Show your work.

a) (2 points) On average, how much is the nominal wage for a person who works 0 hour a week? (Note that this is a point estimation, not a prediction)

Our dependent variable is in the log scale, therefore when $X = 0$

> $\ln(\widehat{wage_i}) = 7.6580$

To interpret the wage in nominal term, we need to find the anti-log which is to

> $e^{\ln(\widehat{wage_i})} = e^{7.6580} = \mathbf{2,117.5182}$

In conclusion, when hours worked is zero, a person gains on average 2,117.5182 Thai Baht.

b) (2 points) If a person works an hour more, how much, on average, wage change do we expect?

This is log-lin function. When a person works an hour more, we expect that wage will go up by $\hat{\beta}_2$ x 100 percent or 0.0318 x 100 = **3.18 percent**.

c) (3 points) If you want to change the reading of unit from hours worked to days worked, what values in the main_hr row will differ? Calculate the changes to all the values in that row, **disregarding the constant row**. You might want to impose an assumption here. State that clearly before calculation.

First of all, we may assume that people work 9 hours a day on average (your assumption on hours worked per day may vary which will affect the multiplier). Therefore, the interpretation changes from an increase in an hour of work to a 9-hour increase of work (a day of work).

Which means that the coefficient ($\hat{\beta}_2$), its standard error and the 95% confidence interval must be all scale up 9 times.

> $\hat{\beta}_2$ becomes 0.0318 x 9 = **0.2862**

> se($\hat{\beta}_2$) becomes 0.0033 x 9 = **0.0297**

> The 95% confidence interval will be 0.0253 x 9 = **0.2277** for the lower bound and 0.0383 x 9 = **0.3447** for the upper bound.

The rest of values in the table, disregarding the constant, remain the same. (NOTE: disregarding the constant term make sense and the calculations in this row are still correct in this context since $\hat{\beta}_1$ is given by $\hat{\beta}_2$)

****