Assignment 1

Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

6304640235

Instructions

(1) Please read the instruction carefully. Also take this habit with you into the exam room.
(2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at lease a paragraph for your analysis, or a graph when necessary, even when the question does not indicate so.
(3) Handing and submitting assignments are only available via BE Moodle.

Answering the questions and preparing answer sheets

(1) Answers are to be handwritten, in either digital or analog form, in a blank canvas or any clean paper. Make sure that your handwriting is clearly visible and readable.
(2) There is no need to rewrite the question. Just indicate the question number clearly for each of the answer, such as 1.a).
(3) Default decimal point is 4.
(4) Choose precise wordings, especially when you want to interpret the meaning of a test, confidence interval, or coefficients.
(5) When done, for the digital case, collage all the pages into a single PDF file. For those who write on sheets of paper, take photo of all pages then convert all of them into a single PDF file as well.
(6) Name your PDF file as StudentID_YourNickname, such as 640123456_Bo.

Submitting your answers

(1) Make sure your file does not exceed 10MB. This is the maximum file size for BE Moodle upload.
(2) Login to BE Moodle, head into the course, then the assignment topic.
(3) Choose your file to submit. Done. There will be timestamp for your upload date and time, so please make sure to not submit later than that.

EE325 Introductory Econometrics **Section 2** Weerawat Phattarasukkumjorn
Assignment 1
Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

---

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.

$\hat{\beta}_1$

a) (4 points) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID\left(0, \sigma^2\right)$, find the estimators of β_1
$\hat{\beta}_2$
and β_2 with OLS method and explain the meaning of the model.

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

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

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

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

f) (2.5 points) Test the hypothesis whether coefficients (both β_1 and β_2) are different from zero at 0.05 level of significance.

answer the following questions. Show your work.

6304640235

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 ($\hat{\beta}_1$) and β_2 ($\hat{\beta}_2$) with OLS method and explain the meaning of the model.

$$\beta_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{46,191.6183}{23,153.3861}$$

$$= 1.9924 \ *$$

$$\beta_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$= 69.1478 - (1.9924)(86.0826)$$

$$= 69.1478 - 171.5110$$

$$= -102.3632 \ *$$

$\therefore$ $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

$Y_i = -102.3632 + 1.9924 x_i$

$\hat{\beta}_1$ or -102.3632 is intercept

$\hat{\beta}_2$ or 1.9924 is slope of SRF, when x_i increase by 1 y increase by 1.9924

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

$$r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y - \bar{Y})^2} = \frac{1 - \sum (Y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$r^2 = \frac{1 - \sum u^2}{\sum (Y - \bar{Y})^2}$$

$$= 1 - \frac{(2,610.9811)}{94,525.1798}$$

$$= 1 - (0.0276)$$

$$= 0.9724$$

$\therefore r^2$ tell how much SRF fitted the data so

X_i explain about 0.9724% of the variation in Y_i

if it gets near to 1 means that the line fitted the data well

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.

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$= -102.3632 + 1.9924(60)$$

$$= 17.1808$$

$\therefore$ $\hat{Y}_i$ estimat of $E(Y|X_i)$ so in this case

give $x_i = 60$, $\hat{y}_i$ will $= 17.1808$

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

6304640235

$Var(u_i) = \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{2,610.9211}{46-2} = 59.3391$

$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n \sum x_i^2} \hat{\sigma}^2 = \frac{364,023.30}{46(23,153.3861)} (59.3391) = 20.2814$

$\hat{\sigma}^2_{\hat{\beta}_2} = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{59.3391}{364,023.30} = 0.000163$

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

$CI = 95\% = 1-\alpha$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $d.f. = 46-2 = 44$

$= \hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_2}$

$= 1.9924 \pm (2.021)(0.0128)$

lower = 1.9665

upper = 2.0183

$\therefore P[1.9665 \leq \beta_2 \leq 2.0183] = 0.95$

when the CI is 95%, the value of β_2 will be between 1.9665 (lower bound) and 2.0183 (upper bond)

f) (2.5 points) Test the hypothesis whether coefficients (both β_1 and β_2) are different from zero at 0.05 level of significance.

$\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$

β_1

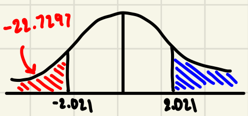
① $H_0: \beta_1 = 0 \rightarrow$ null hypothesis

$H_a: \beta_1 \neq 0$

② $t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{-102.3632}{4.5035} = -22.7297$

③ $d.f. = 46-2 = 44$

$t_{\frac{\alpha}{2}} = \pm 2.021$



④ $t_{cal} = -22.7297$ which falls into the rejection area so we reject null hypothesis.

We are 95% sure that β_1 is not 0

β_2

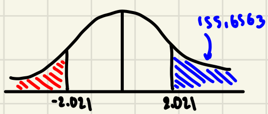
① $H_0: \beta_2 = 0 \rightarrow$ null hypothesis

$H_a: \beta_2 \neq 0$

② $t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{1.9924 - 0}{0.0128} = 155.6563$

③ $d.f. = 46-2 = 44$

$t_{\frac{\alpha}{2}} = \pm 2.021$



④ $t_{cal} = 155.6563$ which means that it falls into the rejection area so we will reject the null hypothesis.

We are 95% sure that β_2 is not 0

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?

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.

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.

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

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)

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

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.

****

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

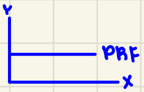
6304640235

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

No , if there is only one point , SRF is linear so its can be many slope. Moreover linear need at least a point.

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.

β_2 can tell us How X and Y relate because β_2 is slope of PRF if there are no β_2 or $\beta_2 = 0$, the PRF will be horizontal line if the β_2 is a, when x increase by 1, dependent variable or Y increase by a



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.

When β_2 is not zero in most of times, this can be clarified that x is relate to Y

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

When we try to find the estimator. If it came up in point which is single number there will be error because when we drawn the sample, the group of data can be vary. Therefore, if we estimate the min interval, the data interval can cover the rage of data.

6304640235

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)

$lwage = 7.6581 + 0.0318\ (\cancel{main}\_hr)$

$lwage = 7.6581$

$wage = 2117.7299$

$\therefore$ the nominal wage for a person who works 0 hr. a week is 2,117.7299

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

$lwage = 7.6581 + 0.0318\ (main\_hr)$

$\frac{d\, lwage}{d\, main\_hr} = 0.0318$

$\frac{d\, \widehat{wage}}{wage} = 0.0318\ d main\_hr$

> multiply by 100 on both side to get percentage

$= \%\Delta\, \widehat{wage} = 0.0318\ d main\_hr \times 100$

$= 3.18\%$

$\therefore$ if the person work an hour more (main_hr), we expect wages to increase by 3.18%

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.

hr $\ln \widehat{wage} = 7.658082 + 0.0318017\ (main\_hour)$; if work 0 hour, worker get 7.658082 baht

se = (0.1256392) (0.003312) if work 1 more hr, worker get 3.18017 %

day $\ln \widehat{wage} = 7.658082 + (0.0318017)(24)(main\_day)$; if work 0 day, worker get 7.658082 baht

se = (0.1256392) $\underbrace{(0.003312)(24)}_{\hat{\sigma}_{\hat{\beta}_2}}$ if work 1 more day, worker get 76.32401%