Assignment 1
Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

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.

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: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID\left(0, \sigma^2\right)$, find the estimators of β_1 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 $\operatorname{var}(u_i)$, $\operatorname{var}\left(\hat{\beta}_1\right)$ and $\operatorname{var}\left(\hat{\beta}_2\right)$

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.

---

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.

****

Question I

a) $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

$= 69.1478 - (0.8789)(86.0826)$

$= -6.5102$

$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$

$= \frac{319,943.18}{364,023.30}$

$= 0.8789$

An estimated Y-intercept is -6.5012 and estimated slope is 0.8789.

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

$= 1 - \frac{2,610.9211}{94,525.1748}$

$= 0.9724$

r^2 value suggest that X_i explains about 97.24 percent of variation in Y_i.

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

$= -6.5012 + (0.8789)(60)$

$= 46.2328$

When X_i is equal to 60, the $\hat{Y}_i$ is 46.6828.

d) $var(u) = \frac{\sum \hat{u}_i^2}{n-k}$

$= \frac{2,610.9211}{46-2}$

$= 59.3391$

$var(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$

$= \frac{\sigma^2}{n}$

$= \frac{59.3391}{46}$

$= 1.2900$

$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$

$= \frac{59.3391}{364,023.30}$

$= 0.0002$

e) 95-percent confident interval for $\hat{\beta}_2$ is

$\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \sigma_{\hat{\beta}_2}$

$= 0.8789 \pm (2.021)(\sqrt{0.0002})$

$= (0.8503, 0.9075)$

There is 95% certain that β_2 lies between 0.8503 and 0.9075.

Question I

f) (i) $H_0: \beta_1 = 0 \ , \ \beta_2 = 0$

$H_1: \beta_1 \neq 0 \ , \ \beta_2 \neq 0$

(ii) $t_{cal\beta_1} = \dfrac{-6.5102 - 0}{\sqrt{1.2900}}$

$= -5.7319$

$t_{cal\beta_2} = \dfrac{0.8789 - 0}{\sqrt{0.0002}}$

$= 62.1476$

(iii) Critical Value at $\alpha = 0.05$

for β_1;

$\mu = 0 \rightarrow \text{boundary} = \pm t_{\frac{\alpha}{2}}$

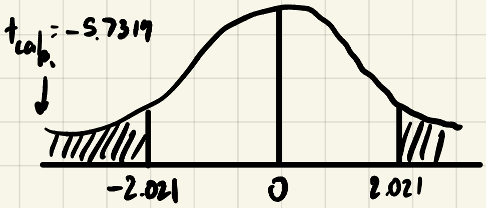
$t_{\frac{\alpha}{2}} = \pm 2.021$

for β_2;

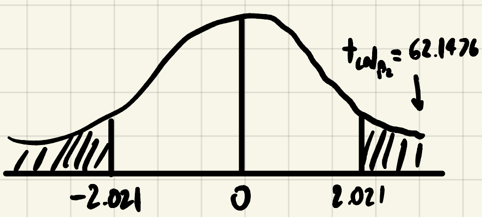
$\mu = 0 \rightarrow \text{boundary} = \pm t_{\frac{\alpha}{2}}$

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

β_1 Two-tail Test



β_2 Two-tail Test



(iv) As $t_{cal\beta_1}$ and $t_{cal\beta_2}$ both lied beyond the bondaries, we can ensure that, at 0.05 level of significant, both β_1 and β_2 are not zero.

Question II

a) No. If there is only one data point, we cannot create a sample regression function because the SRF itself is a representative of the relationship between multiple sets of dependent variable and multiple sets of independent varible. A single point of data cannot illustrate how dependent and indipendent varibles relate to each other.

b) No. Even though β_2 shows how change in independend variable affets change in dependend variable, it does not suggest that one variable causes another variable. Given an education-level-to-GDP relationship example, it provides information that these two variables are related but not cause each other without further information.

c) The result of β_2 significantly different from zero points that, under certain level of significant, we can ensure that β_2 is not zero i.e. null hypothesis that $\beta_2 = 0$ is rejected, at certain level of significant. Hence, it shows that independent and dependend variables are statistically related.

d) Interval estimation has benefits on providing further information on where the parameter are prone to locate. Thus, it provides more accuracy on the apporimation.

Question III

a) Let $\hat{X}_i$ be hours work per week in hour and $\hat{Y}_i$ be wage in Thai Baht.

SRF: $\hat{Y}_i = \beta_1 + \beta_2 X_i$

$$\ln \hat{Y}_i = 7.6581 + 0.3180 X_i$$

work 0 hr/week, the wage is

$$\begin{aligned} \ln \hat{Y}_i &= 7.6581 + 0.3180(0) \\ \ln \hat{Y}_i &= 7.6581 \\ \hat{Y}_i &= e^{7.6581} \\ &= 2{,}117.7299 \end{aligned}$$

Person who works 0 hours a week receives 2,117.7299 THB of nominal wage.

b) Let $\hat{X}_i$ be hours work per week in hour and $\hat{Y}_i$ be wage in Thai Baht.

SRF: $\ln \hat{Y}_i = 7.6561 + 0.3180 X_i$

When X_i changes by 1 unit, Y will change by:

$$\begin{aligned} \frac{d \ln \hat{Y}_i}{dx_i} &= \frac{d}{dx} 7.6581 + \frac{d}{dx_i} 0.3180 x_i \\ \frac{dY}{dx} \times \frac{1}{Y} &= 0.3180 \\ \frac{dY}{Y} &= 0.3180\, dX \\ \frac{dY}{Y} \times \frac{100}{100} &= 0.3180\, dX \\ \Delta \%Y &= 0.3180\, dX \end{aligned}$$

∴ When people work 1 hour more, their wage will increase by 0.3180 percent on average.

(c) Let $\hat{X}_i$ be hours work per week in hour and $\hat{Y}_i$ be wage in Thai Baht.

SRF: $\ln \hat{Y}_i = 7.6561 + 0.3180 X_i$

Suppose there is 24 hours a day.

SRF when the unit is days work:

$$\begin{aligned} \ln \hat{Y}_i &= 7.6581 + 0.3180 X_i \times 24 \\ \ln \hat{Y}_i &= 7.6581 + 0.7632 X_i \end{aligned}$$

se when the unit is days work:

$$0.003312 \times 24 = 0.0795$$

95% confident interval when the unit is days work:

$$\begin{aligned} (\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\beta}_2}) &= 0.7632 \pm 1.984 \times 0.0795 \\ &= (0.6055, 0.9209) \end{aligned}$$

When the unit is days work, following values will change: $\hat{\beta}_2$, se, and confidence interval. $\hat{\beta}_2$ is 0.7632. se is 0.0795. Confidence interval is between 0.6055 and 0.9209.