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

6204640780
Pinmanee Wangcharoen

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.

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.

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 $\operatorname{var}(u_i)$, $\operatorname{var}(\hat{\beta}_1)$ and $\operatorname{var}(\hat{\beta}_2)$

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.

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.

$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i$, $\hat{\beta}_2 = \bar{Y} - \hat{\beta}_2 \bar{X}$

Estimator of β_2 : $\hat{\beta}_2 = \dfrac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = \dfrac{46,131.6183}{23,153.3861} = 1.9924$ Ans

$\hat{\beta}_1 = \bar{Y} - \beta_1 \bar{X}$

$= 69.1428 - (1.9924)(86.0826)$

$= -102.3632$ Ans

$\bar{Y} = 69.1478$, $\bar{X} = 86.0826$, $\hat{Y}_i = -102.3632 + 1.9924 X_i$ Ans

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

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

$r^2 = \dfrac{(\Sigma x_i y_i)^2}{\Sigma x_i^2 \Sigma y_i^2} = \dfrac{(46,131.6183)^2}{23,153.3861 \times 94,525.1748} = 0.9723 = 97.23\%$ Ans

$\Sigma(\overset{\text{DATA}}{Y_1} - \overset{\text{แทนสม}}{\hat{Y}})^2$

$\therefore$ r^2 is the coefficience of determination. The implication is to measure the fitness of the proportion of the total variation in y explanied by the regression model. In this case, 97.23% of y is explained by the regression model Ans

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

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

$= 17.1808$ Ans

$\therefore$ When X_i with 60, the average is $\hat{Y}_i = 17.1808$.

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

$\hat{\sigma}^2 = \dfrac{\Sigma \hat{u}^2}{n-k} = \dfrac{2610.9211}{46-2} = 59.3391$ Ans

$\text{Var}(\hat{\beta}_1) = \hat{\sigma}^2_{\beta_1} = \dfrac{\Sigma X^2}{n\Sigma(x-\bar{x})^2} \hat{\sigma}^2 = \dfrac{364023.30\,(59.3291)}{46\,(23,153.2001)} = 20.2780$ Ans

$\text{Var}(\hat{\beta}_2) = \hat{\sigma}^2_{\beta_2} = \dfrac{\hat{\sigma}^2}{\Sigma(x-\bar{x})^2} \hat{\sigma}^2 = \dfrac{59.3391}{23.1533861} = 0.00256$ Ans

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

$\alpha = 0.05$

$t_{\frac{\alpha}{2}, n-k}$, $Df = n - k = 46 - 2 = 44$

$t_{\frac{\alpha}{2}, 44} = 2.0154$

$\hat{\beta}_2 - \left(t_{\frac{\alpha}{2}, n-k} \cdot \hat{\sigma}_{\hat{\beta}_2}\right) \leq \beta_2 \leq \hat{\beta}_2 + \left(t_{\frac{\alpha}{2}, n-k} \cdot \hat{\sigma}_{\hat{\beta}_2}\right)$

$1.9924 \pm (2.0154)(\sqrt{0.00256})$

$1.9924 \pm (2.0154)(0.0506)$

1.9924 ± 0.10198

$1.89042 \leq \beta_2 \leq 2.09438$ Ans

$\therefore$ We can estimate that 95 out of 100 times, β_2 value is between 1.89042 and 2.09438 Ans

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.

$H_0 : \beta_1 = 0$ $\quad$ $H_0 : \beta_2 = 0$ $\quad$ $t_{\frac{\alpha}{2}, kk} = \pm 2.154$

$H_a : \beta_1 \neq 0$ $\quad$ $H_a : \beta_2 \neq 0$

$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}}$

$= \frac{-102.3632 - 0}{\sqrt{20.2980}}$

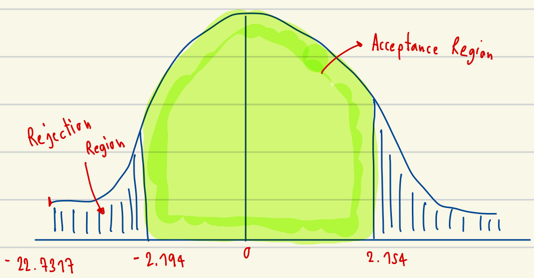
$= \frac{-102.3632}{4.5031}$

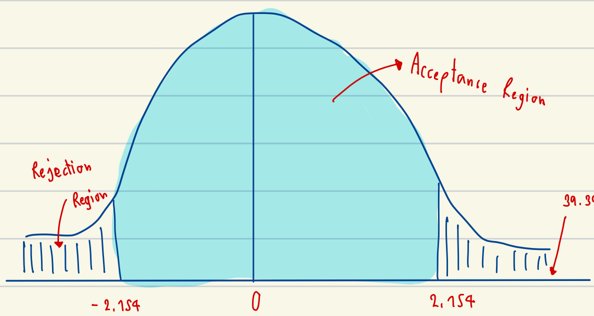
$= -22.7317$ Ans

$t(a) = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$

$t_{(a)} = \frac{1.9924 - 0}{0.0506}$

$t_{(a)} = 39.375$ Ans





$\therefore$ Both t_{cal} is the critical value so we can reject the null hypothesis. In otherwords, we can say for sure that both β_1 and β_2 is not equal to 95 out of 100. Ans

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
Model	50.060869	1	50.060869
Residual	166.152715	306	.542982728
Total	216.213584	307	.704278775

Number of obs	=	308
F(1, 306)	=	92.20
Prob > F	=	0.0000
R-squared	=	0.2315
Adj R-squared	=	0.2290
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.

$$Y = 7.658082 + 0.0318 (x_i) + (0.0318 \times 24)\, x_i$$

Day

****

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?

a) If there is only one data point, we cannot make a sample regression function. Hence, 2 data point is needed for full of the sample function.

b) It is sufficient because β_2 is the slope of population regression function.
If $\beta_2 = 0$ we can claim the population of the function is a holizontal line and $\beta_2 = 2$, it implies that when x rise by 1, y increase by 2.

c) The reason why people test against zero is that want to make sure that β_2 is not zero.
In other words, β_2 is not zero, X and Y are said to be related.

d) An estimate of a population permeter given by a single number is called a point estimate of the parameter.
Other estimate given by 2 numbers between the parameter may be considered to lie is called internal estimate of the parameter.
Moreover, interval estimate gives VS a range of value that is likely to contain the population parameter.
It is more preferable because it gives the range of value of population parameter that is more accurate at the point estimated (provided).

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.

a) $\ln \text{wage} = 7.658082 + 0.0318017 X_i$

$X = 0$; $\ln \text{wage} = 7.658082 + 0.0318077 (0)$

$\ln \text{wage} = 7.658082$

$\text{wage} = e^{7.658082} = 2,117.6917$

$\therefore$ workers work 0 hour / week, they will get the wage 7.658082 bath

2

b) $\ln \text{wage} = 7.658082 + 0.0318017 (1)$

$\text{wage} = e^{7.689883} = 2,186.1187$

$\text{wage}_{x=1} - \text{wage}_{x=0} = 68.4270$

$\therefore$ Workers work 1 hour / week, they will get 68.427 bath.

c) $\ln \text{wage} = 7.658082 + \frac{0.0318019}{24} (24 X_i)$

$= (0.1256392) \left(\frac{0.003312}{24}\right)$

$X_i = \text{day / weeked (24 hours)}$

$= 7.658082 + 0.001325 (X_i)$

$(0.1256392) \quad (0.000138)$

$\therefore$ If it scale X up from 1 hour to a day. It means if a person works for a day then the scale of working time will be multiplied by 24.
The scale follows the unit of working time.
In other words, the depreciation is 24 times in comparing with the original data.
Hence, the cofficient $\hat{\beta}_2$ and its standard error must be divided by 24.

2