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.

****

6304640474

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

1 a) $\beta_1 = \bar{Y} - \hat{\beta}_2 \bar{x}$, $\hat{\beta}_2 = \dfrac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} = \dfrac{\sum x_i (Y_i - \bar{Y})}{\sum x_i (x_i - \bar{x})}$

find $\hat{\beta}_2 = \dfrac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$

$= \dfrac{46.131.6183}{23,153.3861}$

$= 1.9924$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x}$

$= 69.1478 - (1.9924)(86.0826)$

$= 69.1478 - 171.5110$

$= -102.3632$

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

$Y_i = -102.3632 + 1.9924 x_i$

Therefore: -102.3632 which is $\hat{\beta}_1$ is autonomous or intercept 1.9924 or $\hat{\beta}_2$ is slope of SRF ; So when X increase by 1 , Y increase by 1.9924

1 b) $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS} = 1 - \dfrac{\sum \hat{U}_i^2}{\sum (Y - \bar{Y})^2} = 1 - \dfrac{\sum (Y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2}$

$R^2 = 1 - \dfrac{\sum \hat{u}^2}{\sum (Y - \bar{Y})^2}$

$= 1 - \dfrac{(2,610.9211)}{94,525,1748}$

$= 1 - (0.0276)$

$= 0.9724$

$\therefore$ explanation : $R^2 = 0.9724$

The SRF fits well with the data because R^2 or coefficient determination is near 1.

1 c.) $\hat{Y}_i = -102.3632 + 1.9929(x_i)$

$\hat{Y}_i = -102.3632 + (1.9929)(60)$

$\hat{Y}_i = 17.18$

$\therefore$ when x_i equals to 60, $\hat{Y}_1$ will be 17.18

1.D.) $Var(u_i) = 59.3391$

$$Var(\hat{\beta}_1) = \frac{\sum x^2}{n\sum(x_i x^2)^2}(\sigma)^2 = \frac{(369{,}023.30)}{46(23153.3861)} \cdot (59.3391)$$

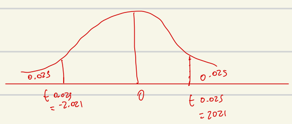
$$= 20.2614$$

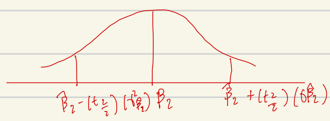
$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} = \frac{59.3391}{23{,}153.3861} = 0.0026$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-u} = \frac{2610.9211}{46-2} = \frac{2610.9211}{44} = 59.3391$$

1 e.) $\beta_2 \mid \alpha = 0.05$, only estimator is σ will use t-stat instead of Z

$$t = \frac{\hat{\beta}_2 - \beta_2}{\sigma\hat{\beta}_2}$$





Degree of freedom
$= 46-2$
$= 44$

Sd

$$P\left[-t\frac{2}{2} < \frac{\hat{\beta}_2 - \beta_2}{\sigma\hat{\beta}_2} < t\frac{2}{2}\right] = 1-2$$

$$P\left[\hat{\beta}_2 - (t\tfrac{2}{2})\underbrace{(\sigma\hat{\beta}_2)}_{9.1016} < \beta_2 < \hat{\beta}_2 + (t\tfrac{2}{2})(\sigma\hat{\beta}_2)\right] = 95\%$$

$$P[(1.9924) - (2.021)(4.5035) < \beta_2 < (1.9928) + (2.021)(4.5035)] = 95\%$$

$$P[(1.9924) - (9.1016) < \beta_2 < (1.9939) + (9.1016)] = 95\%$$

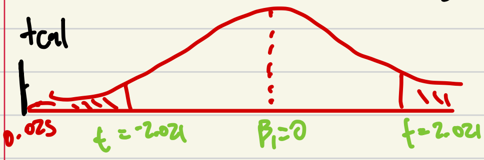
$$P[-7.1092 < \beta_2 < 11.094] = 95\%$$

$\therefore$ β_2 can be between -7.1092 and 11.094, which 95% of the observation of β_2 falls in this interval

1f.) $H_0: \beta_1 = 0$ - null hypothesis
$H_a: \beta_1 \neq 0$

$$\delta\beta_1 = \sqrt{20.2814}$$

- $t_{cal} = \dfrac{\hat{\beta}_1 - \beta_1}{\delta\hat{\beta}_1} = \dfrac{-102.3632 - 0}{4.5035} = -22.73$
- lower bound $= t_{\frac{2}{2}} = -2021$
- upper bound $= t_{\frac{2}{2}} = 2021$



n-k=44

∴ when tcal falls to rejection area, we can reject null hypothesis of $\beta_1 = 0$.

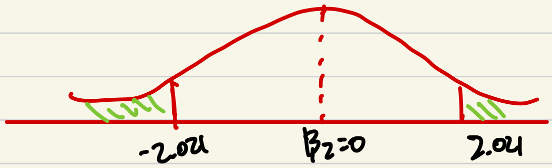
$H_0: \beta_2 = 0$ null hypothesis
$H_a: \beta_2 \neq 0$

$$\delta\hat{\beta}_2^2 = 0.0026 \rightarrow \delta\hat{\beta}_2 = \sqrt{0.0026} = 0.05099$$

$$t_{cal} = \frac{\hat{\beta}_2 - 0}{\delta\hat{\beta}_2} = \frac{1.9924}{0.05099} = 39.0743$$

lower bound $= t_{\frac{2}{2}} = -2.021$

upper bound $= t_{\frac{2}{2}} = 2.021$



when tcall falls to rejection area, we can reject null hypothesis of $\beta_2 = 0$

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?

a.) No, because if there is only of point of data, the SRF can have many slopes.

B.) Yes, β_2 can tell us how X and Y are related because β_2 is the slope of PRF. So, if there are no both β_2 or $\beta_2 = 0$, the PRF will be a horizontal line

c.) when we found out that β_2 is not zero, we can pretty much clarify that x and Y are related

d.) The advantage of using interval estimation over point estimation is that it covers bigger range of values. This make us not be over confident that the population value is exactly equal to the point estimate. For example the exact point mean would be at 32, and our 95% confidence interval is between 30 and 34.

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
ESS	Model	50.060869	1	50.060869	Prob > F	=	0.0000
RSS	Residual	166.152715	306	.542982728	R-squared	=	0.2315 $\to R^2$
					Adj R-squared	=	0.2290
TSS	Total	216.213584	307	.704278775	Root MSE	=	.73687

	lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
depend / indepen	main_hr	.0318017	.003312	9.60	0.000	.0252844	.0383189
$\hat{\beta}_1$	_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.) $\hat{\beta}_1 = 7.658082$

B.) $\hat{\beta}_2 = 0.0318017$ by 1 unit of hour

x_i = unit of hrs.

c.) 24 hours: $\widehat{lwage} = 7.658082 + 0.0318017x_i$

se: $(0.1256392)\ (0.003312)$

1 day: $\widehat{lwage} = 7.658082 + \overbrace{(24)(0.0318017)}^{0.7632} x_i$ x_i = unit day of hrs.

$\delta\hat{\beta}_2 = (0.003312)(24)$
$= 0.07949$

∴ when we are working for 1 day, lwage will increase by 0.7632

when our x changes from hours to days, the confidence interval will change because our $\delta\hat{\beta}_2$ changed to 0.07949