

**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 least 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.

- (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  $\beta_1$  and  $\beta_2$  with OLS method and explain the meaning of the model.
- (2 points) Find  $R^2$  and explain its meaning.
- (1 points) If  $X_i = 60$ , estimate the value of  $\hat{Y}_i$  and explain its meaning.
- (3 points) Find the estimators of  $\text{var}(u_i)$ ,  $\text{var}(\hat{\beta}_1)$  and  $\text{var}(\hat{\beta}_2)$
- (2.5 points) What are the 95-percent confident intervals for  $\beta_2$ ? Interpret the meaning.
- (2.5 points) Test the hypothesis whether coefficients (both  $\beta_1$  and  $\beta_2$ ) are different from zero at 0.05 level of significance.

①

$$\begin{aligned} \textcircled{a} \quad \hat{\beta}_2 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ &= \frac{46131.6183}{23153.3861} = 1.9924 \# \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x} \\ &= 69.1478 - (1.9924)(86.0926) \\ &= -102.3632 \# \end{aligned}$$

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

$$\hat{y}_i = -102.3632 + 1.9924 x_i \#$$

explain: The intercept of this model is at -102.3632 and the slope is 1.9924. When  $x_i$  increases by 1 unit,  $\hat{y}_i$  increases by 1.9924. #

②

$$\begin{aligned} r^2 &= \frac{ESS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum (y_i - \bar{y})^2} \\ &= 1 - \frac{2610.9211}{94525.1748} = 0.9724 \# \end{aligned}$$

explain:  $r^2$  values of 0.9724 suggests that  $x_i$  explain about 97.24% of the variation in  $y_i$ .  
>> near 1 = fit well! #

③

$$x_i = 60$$

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

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

$$= 17.1808 \#$$

④

$$\text{Var}(u_i)$$

$$\rightarrow \text{var}(u_i | x_i) = \sigma^2$$

$$= \frac{\sum \hat{u}_i^2}{n-k} = \frac{2610.9211}{46-2}$$

$$= 59.3391 \#$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \sigma^2$$

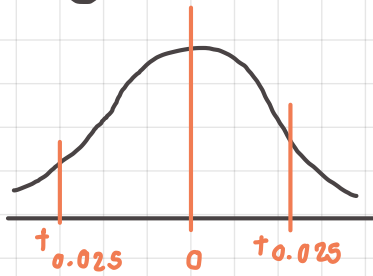
$$\uparrow \sum (x_i - \bar{x})^2$$

$$= \frac{364023.3}{(46)23153.3861} (59.3391) = 20.2814 \#$$

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

$$= \frac{59.3391}{23153.3861} = 0.002562 \#$$

e)



Degree of freedom

$$n - k = 46 - 2 = 44$$

$$P[\hat{\beta}_2 - \sigma \leq \beta_2 \leq \hat{\beta}_2 + \sigma] = 1 - \alpha$$

$$P[-t_{\frac{\alpha}{2}} \leq \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \leq t_{\frac{\alpha}{2}}] = 1 - \alpha$$

$$P[\hat{\beta}_2 - (t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}) \leq \beta_2 \leq \hat{\beta}_2 + \sigma_{\hat{\beta}_2}] = 1 - \alpha$$

$$P[1.9924 - (2.021)(0.0506) \leq \beta_2 \leq 1.9924 + (2.021)(0.0506)] = 95\%$$

$$P[1.8901 \leq \beta_2 \leq 2.0947] = 95\%$$

To interpret:  $\beta_2$  is between 1.8901 and 2.0947 which 95%.

f)

$$H_0: \beta_1 = 0 \rightarrow \text{Null } \checkmark$$

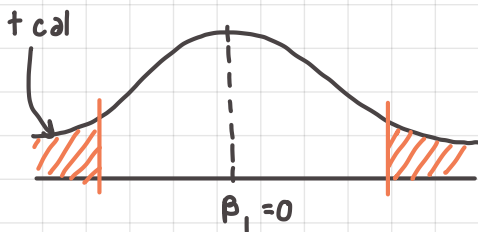
$$H_a: \beta_1 \neq 0 \rightarrow \text{Alternative}$$

$$S_{\hat{\beta}_1}^2 = 20.2814$$

$$S_{\hat{\beta}_1} = 4.5035$$

$$t_{\text{cal}} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} = \frac{-102.3632 - 0}{4.5035}$$

$$\rightarrow -22.73$$



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

• upper bound:  $t_{\frac{\alpha}{2}} = 2.021$

$$H_0: \beta_2 = 0 \rightarrow \text{Null } \checkmark$$

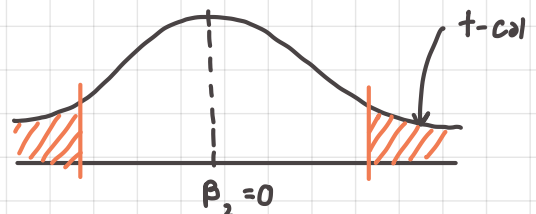
$$H_a: \beta_2 \neq 0 \rightarrow \text{Alternative}$$

$$S_{\hat{\beta}_2}^2 = 0.002562$$

$$S_{\hat{\beta}_2} = 0.0506$$

$$t_{\text{cal}} = \frac{\hat{\beta}_2 - 0}{S_{\hat{\beta}_2}} = \frac{1.9924}{0.05099}$$

$$\rightarrow 39.0743$$



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

• upper bound:  $t_{\frac{\alpha}{2}} = 2.021$

Ans: when the t-cal fall into rejection area we can reject null hypothesis that  $\beta_1 = 0 / \beta_2 = 0$ .

2. (8 points) Answer the following question without any mathematical proof. A 3-6-line paragraph for a question is sufficient.
- (2 points) If we have only one data point, can we create a sample regression function? Why?
  - (2 points) Does a significant  $\beta_2$  sufficient for us to believe that  $X$  and  $Y$  are causally related? Provide an example to support your answer.
  - (2 points) When we test a hypothesis and find that  $\beta_2$  is significantly different from zero, what does the result actually suggest? Answer with specific wording from statistical perspective.
  - (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
Model	50.060869	1	50.060869	F(1, 306)	=	92.20
Residual	166.152715	306	.542982728	Prob > F	=	0.0000
				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.

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

(a) No because a only one point of data makes SRF can be many slope plus SRF is linear so it needs at least 2 points.

---

(b)  $\beta_2$  can tell how  $x$  and  $y$  relate because  $\beta_2$  is slope of PRF

If there are no  $\beta_2$  or  $\beta_2 = 0$ , the PRF is horizontal line

If the  $\beta_2$  is 2, when  $x$  increase by 1

, dependent variable or  $y$  increase by 2



(c) when  $\beta_2 \neq 0$  in most of times, this can be imply that  $x$  is relate to  $y$ .

---

(d) When we draw the sample with one single number there will be error  $\rightarrow$  data can vary!

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
ESS Model	50.060869	1	50.060869	F(1, 306)	=	92.20
RSS Residual	166.152715	306	.542982728	Prob > F	=	0.0000
TSS Total	216.213584	307	.704278775	R-squared	=	0.2315
				Adj R-squared	=	0.2290
				Root MSE	=	.73687

dep.(y)	lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
indep.(x)	main_hr	.0318017	.003312	9.60	0.000	.0252844 .0383189
	_cons	7.658082	.1256392	60.95	0.000	7.410856 7.905308

$$\ln \widehat{wage} = \beta_1 + \beta_2 (\text{main\_hr})$$

$$\textcircled{a} \quad \ln \widehat{wage} = \beta_1 + \beta_2 (0)$$

$$\ln \widehat{wage} = \beta_1$$

$$\beta_1 = 7.658082$$

$$\textcircled{b} \quad \text{wage will increase by } 100 (0.0318017) \text{ or } 3.18017\%$$

$$\textcircled{c} \quad \text{hr } \ln \widehat{wage} = 7.658082 + 0.0318017 (\text{main\_hour})$$

$$\text{se} = (0.1256392) (0.003312)$$

if work 0 hour, worker get 7.658082 baht  
if work 1 more hr, worker get 3.18017%.

$$\text{day } \ln \widehat{wage} = 7.658082 + (0.0318017) (24) (\text{main\_day})$$

$$\text{se} = (0.1256392) \underbrace{(0.003312) (24)}_{\sigma_{\hat{\beta}_2}}$$

if work 0 day, worker get 7.658082 baht  
if work 1 more day, worker get 76.3240%