
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 β_1 and β_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 β_2 ? Interpret the meaning.
- (2.5 points) Test the hypothesis whether coefficients (both β_1 and β_2) are different from zero at 0.05 level of significance.

$$1.a.) \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}, \hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i (Y_i - \bar{Y})}{\sum X_i (X_i - \bar{X})}$$

$$\begin{aligned} \text{Find } \hat{\beta}_2 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ &= \frac{46131.6183}{23153.9861} = 1.9924 \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X} \\ &= 69.1478 - (1.9924)(86.0826) \\ &= 69.1478 - 171.5110 \\ &= -102.3632 \end{aligned}$$

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i & \therefore \hat{\beta}_1 &= \text{autonomous or intercept} \\ &= -102.3632 + 1.9924 X_i & \hat{\beta}_2 &= \text{slope of SRF} \end{aligned}$$

So, when X_i increases by 1, Y_i increases by 1.9924

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

$$\begin{aligned} R^2 &= 1 - \frac{\sum u^2}{\sum (Y_i - \bar{Y})^2} \\ &= 1 - \frac{(2610.9211)}{94525.1748} \\ &= 1 - (0.0276) \\ &= 0.9724 \end{aligned}$$

$$\therefore R^2 = 0.9724$$

⚡ The SRF nicely fit to the data because the coefficient determination is close to 1.

$$\begin{aligned} 1.c.) \hat{Y}_i &= -102.3632 + 1.9924(X_i) \\ &= -102.3632 + 1.9924(60) \\ &= 17.18 \end{aligned}$$

$$\therefore \text{when } X_i = 60 \\ \hat{Y}_i = 17.18$$

$$1.d.) \text{var}(u_j) = \hat{\sigma}^2 = \frac{\sum u^2}{n-u}$$

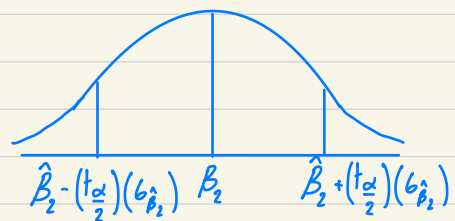
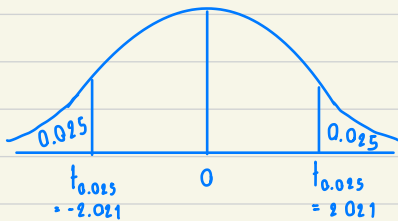
$$= \frac{2610.9211}{46-2} = \frac{2610.9211}{44} = 59.3391$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X^2}{n \sum (X_i - \bar{X})^2} (\hat{\sigma})^2 = \frac{364023.3000}{46(23153.3861)} \times 59.3391$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2} = \frac{59.3391}{23153.3861} = 0.0026$$

1.e.) $\alpha = 0.05$ because it just estimate, we can t-stat instead of z, $t = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}}$

Degree of freedom = $46 - 2 = 44$



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

$$P\left[\hat{\beta}_2 - \left(t_{\frac{\alpha}{2}}\right)(\hat{\sigma}_{\hat{\beta}_2}) < \beta_2 < \hat{\beta}_2 + \left(t_{\frac{\alpha}{2}}\right)(\hat{\sigma}_{\hat{\beta}_2})\right] = 95\%$$

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

$$P\left[1.9924 - 9.1016 < \beta_2 < 1.9924 + 9.1016\right] = 95\%$$

$$P\left[-7.1092 < \beta_2 < 11.0940\right] = 95\%$$

$\therefore \beta_2$ is between $(-7.1092, 11.0940)$ which is 95% of observation, falling in this interval.

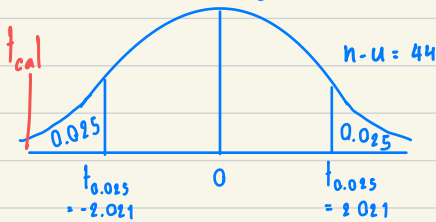
1. f.) $H_0: \beta_1 = 0 \rightarrow$ null hypothesis

$H_a: \beta_1 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{-102.3632 - 0}{4.5035} = -22.7397$$

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

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



\therefore when t_{cal} fall into rejection area null hypothesis can be rejected which is $\beta_1 = 0$

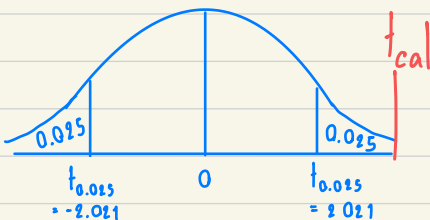
$H_0: \beta_1 = 0 \rightarrow$ null hypothesis

$H_a: \beta_1 \neq 0$

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

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

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



\therefore when t_{cal} fall into rejection area null hypothesis can be rejected which is $\beta_2 = 0$

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?

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

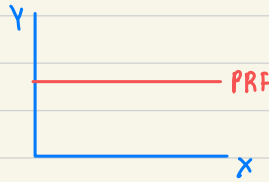
D.V	lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
I.V	main_hr	β_2 .0318017	.003312	9.60	0.000	.0252844 .0383189
	_cons	β_1 7.658082	.1256392	60.95	0.000	7.410856 7.905308

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.a.) Cannot, the SRF will have many slopes if the data has only 1 point.

2.b.) Because β_2 is a slope of PRF, it can inform about the relationship of X and Y.
The PRF will be a horizontal line when $\beta_2 = 0$ (no slope)



2.c.) When β_2 is not zero in most of times, it can tell that X relates to Y.

2.d.) When finding the estimator, if the result is a single number, it will have an error because the group of data can be variable when drawing the sample. Hence, the advantage of an interval estimation is it can cover the range of existed data when estimate them in interval.

3.a.) $\hat{\beta}_1 = 7.658082$

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

3.c.) 24 hours $\rightarrow \widehat{I}_{\text{wage}} = 7.658082 + 0.0318017 (X_i)$ (hour)

1 day $\rightarrow \widehat{I}_{\text{wage}} = 7.658082 + 24(0.0318017)(X_i)$ (day)
0.7632

$$6\hat{\beta}_2 = (0.003312)(24) = 0.07949$$

\therefore I_{wage} will increase by 0.7632 when working 1 day.
The size of confidence interval will change due to $6\hat{\beta}_2$ change to 0.07949, changing x from hour to day.