Assignment 1

Assigned on Sep 14^{th}, 2021. Due date Sep 27^{th}, 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,

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

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 $\mathrm{var}\left(u_i\right)$, $\mathrm{var}\left(\hat{\beta}_1\right)$ and $\mathrm{var}\left(\hat{\beta}_2\right)$

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.

****

1

1.a) find $\hat{\beta}_2 = \dfrac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2}$

$= \dfrac{46,131.6183}{23,153.3861}$

$= 1.9924$

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

$= 69.1478 - (1.9924)(86.0826)$

$= -102.3632$

$\therefore$ $\hat{\beta}_1$ is the intercept of SRF and Y axis so the SRF line will intercept y at -102.3632

$\hat{\beta}_2$ is the slope of SRF line this mean that when X increase in 1 unit y will increase 1.9924

1.b) $r^2 = 1 - \dfrac{\Sigma\hat{u}_i^2}{\Sigma(Y_i-\bar{Y})^2}$

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

$= 0.9724$

$\therefore$ r^2 tell how much SRF fitted the data so X_i explain about 97.24% of the variation in Y_i

if it gets near to 1 means that the line fitted the data well

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

$= -102.3632 + 1.9924(60)$

$= 17.1808$

$\therefore$ $\hat{Y}_i$ estimator of $E(Y|X_i)$ so in this case given $X_i = 60$, $\hat{Y}_i$ will $=$ 17.1808

1.d) $Var(u_i) = \sigma^2 = \dfrac{\Sigma\hat{u}_i^2}{n-k}$

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

$= 59.3391$

$\sigma^2_{\hat{\beta}_1} = \dfrac{\Sigma X_i^2}{n\Sigma x_i^2}\sigma^2$

$= \dfrac{364,023.30}{46(23,153.3861)}(59.3391)$

$= 20.2814$

$\sigma^2_{\hat{\beta}_2} = \dfrac{\sigma^2}{\Sigma X_i^2}$

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

$= 0.000163$

1.e) $CI = 95\% = 1-\alpha$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $d.f. = 46-2 = 44$

find lower and upper limit

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

$= 1.9924 \pm (2.021)(0.0128)$

lower $= 1.9665$

upper $= 2.0183$

$\therefore P[1.9665 \le \beta_2 \le 2.0183] = 0.95$

When the CI is 95%, the value of β_2 will be between 1.9665 (lower bound) and 2.0183 (upper bound)

1.f) $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$

β_1

① $H_0 : \beta_1 = 0$ → null hypothesis

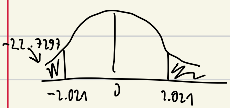
$H_a : \beta_1 \neq 0$

② $t_{cal} = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma \hat{\beta}_1} = \dfrac{-102.3632 - 0}{4.5035}$

$= -22.7297$

③ d.f. $= 46 - 2 = 44$

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



④ $t_{cal} = -22.7297$ which falls into the rejection area so we reject null hypothesis.

We are 95% sure that β_1 is not 0

β_2

① $H_0 : \beta_2 = 0$ → null hypothesis

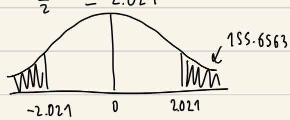
$H_a : \beta_2 \neq 0$

② $t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{\sigma \hat{\beta}_2} = \dfrac{1.9924 - 0}{0.0128}$

$= 155.6563$

③ d.f. $= 46 - 2 = 44$

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



④ $t_{cal} = 155.6563$ which means that it falls into the rejection area so we will reject the null hypothesis.

We are 95% sure that β_2 is not 0

2

2.a) Yes, the reason is that sample regression function will be plot to fitted with each data point the most. So if there is only 1 data point the SRF line will pass through the point and make no error term which is the distance between the line and the data point.

2.b) β_2 is the true value of the slope of the PRF line, which tell us how much y increase or decrease if x changes for 1 unit. So I think that β_2 is sufficient enough to tell us the relationship of X and Y. Example from 1.a) you can see the relationship of x and y that if x increase 1 unit, y will increase for 1.9924.

2.c) When β_2 is significant different from 0 mean that the result of t_{cal} fall in the rejection area. So we reject the null hypothesis (H_0) which state that $\beta_2 = 0$. The result will tell you how much percentage you can be sure that β_2 will not be 0. This is call the type I error.

2.d) Point estimation is when we select one point in the sample to represent a point in the population. This kind of estimation can tell the exact point estimate of the population value. However, it can not tell us the overall estimate of the population value like interval estimation can. It allows you to tell the confidence that an interval will include the true population value

3

3.a) $\widehat{lwage} = 7.6581 + 0.0318(\cancel{main\_hr})$ 0

$\widehat{lwage} = 7.6581$

$wage = 2117.7299$

$\therefore$ the nominal wage for a person who works 0 hr. a week is 2,117.7299

3.b) $\widehat{lwage} = \cancel{7.6581} + 0.0318(main\_hr)$ 0 $\beta_1 + \beta_2$

$\dfrac{d\ \widehat{lwage}}{d\ main\_hr} = 0.0318$

$\dfrac{d\ \widehat{wage}}{wage} = 0.0318\, d\, main\_hr$

> multiply by 100 on both side to get percentage

$= \%\Delta\ \widehat{wage}_i = 0.0318\ d\, main\_hr \times 100$

$= 3.18\%$

$\therefore$ if the person work an hour more (main_hr), we expect wages to increase by 3.18%

3.c) hour worked per week $\xrightarrow{\div 24}$ days worked per week

$lwage = Y$ ⇒ when $x \div 24$ $\hat{\beta}_2$ will be $\hat{\beta}_2 \times 24$

$main\_hr = X$ $se(\hat{\beta}_2)$ will be $se(\hat{\beta}_2) \times 24$

$\widehat{lwage} = 7.6581 + \overset{\hat{\beta}_2}{0.0318}(main\_hr)$ $\Rightarrow \hat{\beta}_2 \times 24 = 0.7632$

$se = (0.1256)\quad (0.003312)$ $\Rightarrow se(\hat{\beta}_2) \times 24 = 0.0795$

↓

$\widehat{lwage} = 7.6581 + 0.7632(main\_days)$

$se = (0.1256)\quad (0.0795)$

$\therefore$ the value that will change in the main_hr row is the Coef. and Std. Err (SE)

Coef will change from 0.0318 → 0.7632

SE will change from 0.003312 → 0.0795