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.

Wasawat Sonbutsaengurai
6304641829

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.

$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{319{,}943.18}{364{,}023.30} \approx 0.8789$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 69.1478 - 0.8789(86.0826) = -6.5101$

so $\hat{Y}_i = -6.5101 + 0.8789 X_i$

The intercept of this model is at -6.5101 and the slope is 0.8789.

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

$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{2{,}610.9211}{94{,}525.1748} = 0.9724$

r^2 values of 0.9724 suggests that X_i explain about 97.24 percent of the variation in Y_i.

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

$\hat{Y}_i = -6.5101 + 0.8789(60) \approx 46.2239$

when $X_i = 60$ $\hat{Y}_i$ is 46.2239

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

wasawat sonbutsaengurai
6304641829

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{\sum (Y_i - \hat{Y})^2}{n-k} = \frac{2610,9211}{46-2} = 59.3391$$

$$var(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 = \frac{364,023.30}{46(364,023.30)}(59.3391) = \frac{59.3391}{46} \approx 1.2900$$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{59.3391}{364,023.30} = 1.63 \times 10^{-4}$$

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

$$\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2} \rightarrow 0.8789 \pm 0.681 \cdot 1.63 \times 10^{-4}$$

$$0.8789 - 2.021 \cdot \sqrt{1.63 \times 10^{-4}} \leq \beta_2 \leq 0.8789 + 2.021 \cdot \sqrt{1.63 \times 10^{-4}}$$

$$0.8530 \leq \beta_2 \leq 0.9047$$

there are 95% that value β_2 would be between 0.8530 and 0.9047.

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.

Find $se(\hat{\beta}_1)$ and $se(\hat{\beta}_2)$

$$se(\hat{\beta}_1) = \sqrt{var(\hat{\beta}_1)} = \sqrt{1.29} = 1.1360$$

$$se(\hat{\beta}_2) = \sqrt{var(\hat{\beta}_2)} = \sqrt{1.63 \times 10^{-4}} = 0.0130$$

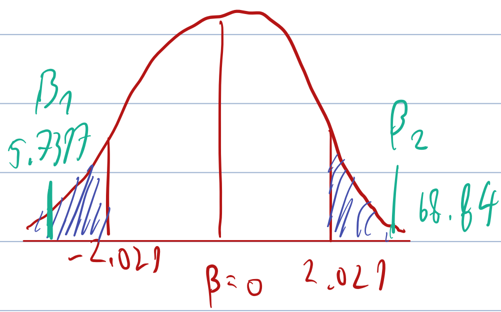
1. $H_0: \beta_1 = 0$ $H_1: \beta_1 \neq 0$ / $H_0: \beta_2 = 0$ $H_1 = \beta_2 \neq 0$

2. $$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = \frac{-6.5101 - 0}{1.1360} = -5.7310$$

wasawat sombutsaengurai
6304641829

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} = \frac{0.8789 - 0}{0.0130} = 68.84$$

③ when $\alpha = 0.05$ then $t_{\frac{\alpha}{2}} = \pm 2{,}021$



Ans. Since both β_1, β_2 are within rejection area so we reject H_0. we can say that β_1, β_2 is not 0 95 out of 100 percent.

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.

$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ ********************

Wasawat Sombutsaengurai
6304647829

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?

we can't create SRF, because SRF illustrates the relationship between multiple sets of multiple independent variable (Y) and multiple dependent variable (x). Therefore, only one data point can't show a function of relationship between x and y.

ด้วยพหูพจน์อะไรซักอย่าง

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.

No, because β_2 only shows the relationship of x and y but it can be coincidence relationship not a causal one. For example, Even though income is related to saving ; we can't conclude that 100 out of the time as income increase will affect people to sawe more.

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.

If level of significant is 95%, Although the result from β_2 is reject H_0, we can't say that β_2 is not 0 absolutely. In contrast, we also can't conclude that $\beta_2 = 0$ 100% as well.

d) (2 points) What is (are) (an) advantage(s) of an interval estimation over point estimate?

Interval estimation show us a range where the parameter is likely to be based on significant level. So it is more accurate than estimate only one point and still have some error gap for failure.

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.

wasawat sonbutsaengurai
6304647829

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)

$y = \beta_1 + \beta_2 X_i$

$\frac{d \ln y}{dx_i} = \beta_2$

$\frac{dy}{dx_i} \cdot \frac{1}{y} = \beta_2$

slope $= \frac{dy}{dx} \rightarrow \frac{dy}{dx_i} = \beta_2 Y$

elasticity $= \frac{dy}{dx} \cdot \frac{x}{y} \rightarrow \frac{dy}{dx} \cdot \frac{x}{y} = \beta_2 X$

$\ln \widehat{wage} = 7.658082 + \cancel{0.0318017}^{\,0} main\_hr \leftarrow (0)$

$\ln \widehat{wage} = 7.658082$

$\widehat{wage} = e^{7.658082}$

$\widehat{wage} = 2117.6918$

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

$\ln \widehat{wage} = 7.658082 + 0.0318017 main\_hr$

take diff.

$\frac{d \ln \widehat{wage}}{d main\_hr} = 0 + 0.0318017$

$\frac{d \widehat{wage}}{wage} = 0.0318017 \, d main\_hr$

$\times 100$ both side

$\% \Delta \widehat{wage} = 3.18017 \, d main\_hr$

Ans: If work hour increase by one hour, we expect wage to increase by 3.18017 percent.

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.

hour worked per week → days work $\hat{Y}_0$ lwage

1 hours → 24 hours $X_i \times 24 \to \frac{\hat{\beta}_2}{24}$

$lwage = 7.658082 + 0.0318017 \times 24\, main\_hr \to lwage = 7.658082 + 0.7632408\, main\_hr$

$(0.1256392)\ (0.003312) \times 24 \qquad (0.1256392)\ (0.079488)$

→ main_hr ↑ 1 hour → lwage ↑ by 0.0318017 Thai Baht

→ main_hr ↑ 24 hour → lwage ↑ by 0.7632408 Thai Baht

```
------------------------------------------------------------------------------
       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
```

d.f. = 308 − 1

new

Confidence interval : $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_2}$

upper = 0.7632408 + 1.984 (0.079488) = 0.920944992

lower = 0.7632408 − 1.984 (0.079488) = 0.605536608

1000,000,000 → 0.2535

1,000,000 → 2.535×10^{-4} 0.0002535

1 hour → 0.0318017

24 hour → 0.7632408

Wasawat sonbutsaengurai
6304647829