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.

Assignment 1

Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

---

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.

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 $\operatorname{var}(u_i)$, $\operatorname{var}(\hat{\beta}_1)$ and $\operatorname{var}(\hat{\beta}_2)$

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.

a) estimator of β_1 and β_2 is $\hat{\beta}_1$ and $\hat{\beta}_2$

$\hat{\beta}_2 = \dfrac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \dfrac{46{,}131.6183}{23{,}153.3861} = 1.99243506$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 69.1478 - 1.9924(86.0826) = -102.3662$

$\hat{Y}_i = -102.3662 + 1.9924(X_i)$

it mean intercept of this model is -102.3662 and slope is 1.9924 when x increase 1 unit y decrease by 1.9924 unit

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

$= 1 - \dfrac{RSS}{TSS} = 1 - \dfrac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} = 1 - \dfrac{2610.9211}{94525.1748} = 0.9724$

R^2 value of 0.9724 suggests that X_i explain about 97.24% of variation in Y_i

c) when $X_i = 60$

$\hat{Y}_i = -102.3662 + 1.9924(60)$

$= 17.1778$

D) find Var (U) Var $(\hat{\beta}_1)$ Var $(\hat{\beta}_2)$

$$\sigma^2 = \frac{\sum \hat{u}^2}{n-U}$$

$$= \frac{2610.9211}{44} = 59.3391$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sum x^2}{n\sum(x_i-\bar{x})^2}\,\sigma^2 = \frac{364023.30}{46(23153.3861)} \cdot 59.3391$$

$$= 20.2814$$

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2} = \frac{59.3391}{23153.3861} = 0.0026$$

e β_2 | $\alpha = 0.05$ as we have only estimator using t-stat substitute z

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

β_2 can be between -7.1092 to 11.094 which 95% of observation β_2 fall in the interval

$$P\left(-t_{\frac{\alpha}{2}}(\sigma\hat{\beta}_2) < \hat{\beta}_2-\beta_2 < t_{\frac{\alpha}{2}}(\sigma\hat{\beta}_2)\right)$$

$$P\left(-\hat{\beta}_2-t_{\frac{\alpha}{2}}(\sigma\hat{\beta}_2) < -\beta_2 < -\hat{\beta}_2+t_{\frac{\alpha}{2}}(\sigma\hat{\beta}_2)\right)$$

$$P\left(\hat{\beta}_2-t_{\frac{\alpha}{2}}(\sigma\hat{\beta}_2) < \beta_2 < \hat{\beta}_2+t_{\frac{\alpha}{2}}(\sigma\hat{\beta}_2)\right)$$

$$= P\left(1.9924-(2.021)(4.5035) < \beta_2 < 1.9924+(2.021)(4.5035)\right)$$

$$= P(-7.1092 < \beta_2 < 11.094) = 95\%$$

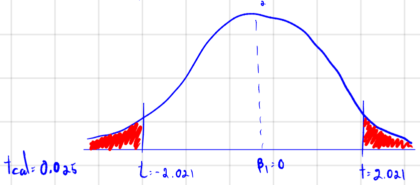
f) $H_0: \beta_1 = 0$ - null hypothesis

$H_1: \beta_1 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_1-\beta_1}{\sigma\hat{\beta}_1} = \frac{-102.2622}{4.5035} = -22.73$$

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

Upper bound $= t_{\frac{\alpha}{2}} = 2.021$



when t cal fall in rejection area we can reject null hypothesis

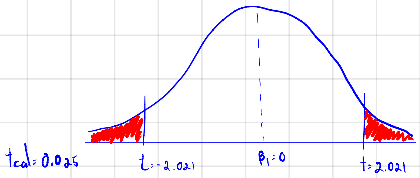
$H_0: \beta_2 = 0$ - null hypothesis

$H_1: \beta_2 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_2-\beta_2}{\sigma\hat{\beta}_2} = \frac{1.9924}{0.5099} = 39.0743$$

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

Upper bound $= t_{\frac{\alpha}{2}} = 2.021$



when t cal fall in rejection area we can reject null hypothesis

Assignment 1
Assigned on Sep 14th, 2021. Due date Sep 27th, 2021 before midnight.

---

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) $Y_i = \beta_1 + \beta_2 X_i$

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) $\beta_1 = 7.658082$

b) wage increase $\beta_2 = 0.0318017$ expect to increase 3.18017%

****

c) $\ln \hat{wage} = 7.658082 + (0.0318017)\beta_2$ if work 0 hour, wage = 7.658082 baht

se = (0.1256392)(0.003312) wage increase 3.18017% per hour increase

$\ln \hat{wage}$ day = $7.658082 + (0.0318017)(24)\beta_2$ if work 0 day, wage = 7.658082 baht

(0.1256392)(0.003312)(24) wage increase 76.3240% per day increase

a) (2 points) If we have only one data point, can we create a sample regression function? Why?

No because SRF is a linear function so it need at least two data

(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.

β_2 is a slope of PRF so it tell relationship o X and Y
example if β_2 is 0.5 X increase by 1 Y is increase by 0.5
of if β_2 is 0 PRF is horizontal

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.

$\beta_2 \neq 0$ this can be clarified that X is relate to Y

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

When we find estimator, if it come with a point which single n there will be error because when we draw sample a data can be vary. So to cover a range of data we must be estimate in the interval