Assignment 1

Penguin
6304640821

1. (15 points) Given this information,

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

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.

a) $\hat{\beta}_2 = \dfrac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} = \dfrac{46{,}131.6183}{23{,}153.3861} = 1.9924$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = 69.1478 - (1.9924)(86.0826) = -102.3632$

- Plugging the estimators into the equation, we have

  $\hat{Y}_i = -102.3632 + 1.9924X_i$

  The intercept of this model is at -102.3632 and the slope is 1.9924.

  When X_i increases by 1 unit, $\hat{Y}_i$ increase by 1.9924.

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

b) $r^2 = 1 - \dfrac{RSS}{TSS} = 1 - \dfrac{\Sigma\hat{u}_i^2}{\Sigma(Y_i-\bar{Y})^2} = 1 - \dfrac{2{,}610.9211}{94{,}525.1748} = 0.9723$

- r^2 value of 0.9723 suggests that X_i explain about 97.23 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.

c) We can replace $X_i = 60$ to find $\hat{Y}_i$, therefore

$\hat{Y}_i = -102.3632 + (1.9924)(60) = 17.1808$

- When $X_i = 60$. $\hat{Y}_i$ is 17.1808

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

d) $\text{Var}(u_i) = \dfrac{\Sigma\hat{u}_i^2}{n-k} = \dfrac{2{,}610.9211}{46-2} = 59.3391$

$\text{Var}(\hat{\beta}_1) = \dfrac{\Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2}\hat{\sigma}^2 = \dfrac{364{,}023.30}{(46)(23{,}153.3861)}(59.3391) = 20.2814$

$\text{Var}(\hat{\beta}_2) = \dfrac{\hat{\sigma}^2}{\Sigma x_i^2} = \dfrac{59.3391}{23{,}153.3861} = 0.0026$

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

① $1-\alpha = 95\%$; $\alpha = 0.05$ (d.f. = 44)

② $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.0025} = 2.021$

③ $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \delta_{\hat{\beta}_2}$ → upper : $1.9924 + (2.021 \times 2.5629) = 7.1720$

→ lower : $1.9924 + (2.021 \times 2.5629) = -3.1872$

$P\left[7.1720 \leq \hat{\beta}_2 \leq -3.1872\right] = 0.95$

At confident level 95 percent, 95% of data will fall into −3.1872 to 7.1720 interval.

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.

$\delta_{\hat{\beta}_1} = \sqrt{var(\delta_{\hat{\beta}_1})} = \sqrt{20.2814} = 4.5035$

β_1) ① $H_0 = \beta_1 = 0$

$\beta_1 \neq 0$

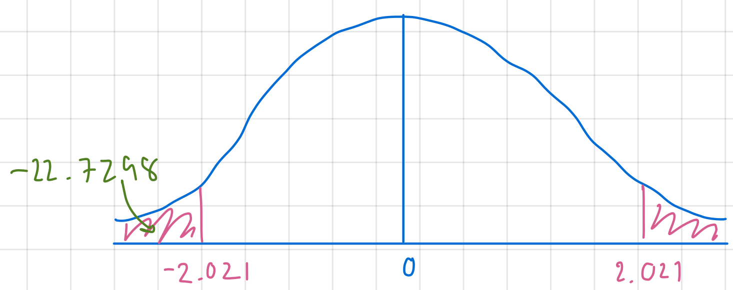
② $t_{calc} = \dfrac{\hat{\beta}_1 - \beta_1}{\delta_{\hat{\beta}_1}} = \dfrac{-102.3632 - 0}{4.5035} = -22.7298$

③ Decision rule

d.f. $= n-k = 46-2 = 44$

$\alpha = 0.05$

$t_{\frac{\alpha}{2}} = t_{0.025} = \pm 2.021$



∴ $t_{calc} < -2.021$, can reject null hypothesis, at the significance level 95%.

We are sure that β_1, is not zero 95 out of 100 times when we sample.

$\delta_{\hat{\beta}_2} = \sqrt{var(\delta_{\hat{\beta}_2})} = \sqrt{0.0026} = 0.051$

β_2) ① $H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

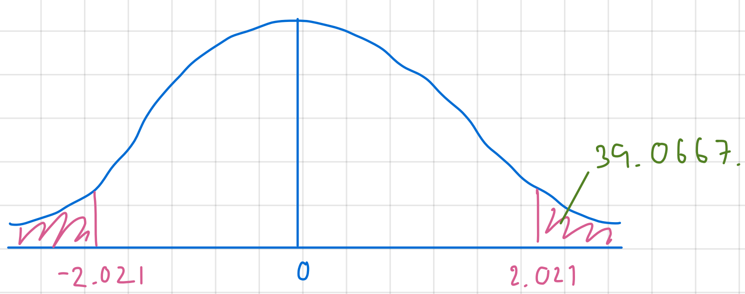
② $t_{calc} = \dfrac{\beta_2 - \beta_2}{\delta_{\hat{\beta}_2}} = \dfrac{1.9924 - 0}{0.051} = 39.0667$

③ Decision rule.

d.f. $= 46-2 = 44$

$\alpha = 0.05$

$t_{\frac{\alpha}{2}} = t_{0.025} = \pm 2.021$



∴ $t_{calc} > 2.021$, can reject the null hypothesis, at the significance level of 95%.

We are sure that β_2 is not zero 95 out of 100 times when we sample.

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?

Ans: No, if we have only one data point, we can't show colleration between data point.
∴ One data point is not sufficient for creating a sample regression function.

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.

Ans: Yes, when β_2 is not zero, x and y are said to be related.
To illustrate, $Y_i = 13.25 + 2.78X_i$; $\beta_2 = 2.78$
When X_i increase 1 unit, Y_i will increase 2.78 unit accordingly.

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.

Ans: When β_2 is significantly different from zero, it can interpret that β_2 is different from zero at particular confidence interval, and implies we expect that changes in X are associated with change in Y.

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

Ans: Interval estimation can express specific range with any confident level but point estimate express only one point on data dispersion which narrower than interval estimation. So, interval estimation represent more accurate data.

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)

$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \rightarrow \hat{Y}_i = 7.658082 + 0.318017 X_i$

when $X_i = 0$, $\hat{Y}_i = 7.658082$

∴ A person who works 0 hour a week, will have 7.658082 nominal wage on average. #

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

$\hat{Y}_i = 7.658082 + 0.318017 X_i$

when $X_i = 1$, $\hat{Y}_i = 7.976099$

The difference of nominal wage is $7.976099 - 7.658082 = 0.318017$

or from $\hat{\beta}_2 = 0.318017$, it can interpret that

when the worker works an hour more, we expect nominal wage will increase by 0.318017 #

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.

Assumption:

Assume that each worker work 8 hours per day.

hour : $\widehat{wage} = 7.658082 + 0.318017(\text{main\_hr})$

$se = (0.003312) \quad (0.1256392)$

day : $\widehat{wage} = 7.658082 + 0.318017(8)(\text{main\_day})$

$se = (0.003312) \quad (0.1256392)(8)$

$\therefore \widehat{wage} = 7.658082 + 2.544136(\text{main\_day})$

$se = (0.00312) \quad (1.0051136)$