$n = 46$ | $\sum_{i=1}^{n} X_i = 3{,}959.80$ | $\sum_{i=1}^{n} Y_i = 3{,}180.80$

$\bar{X} = 86.0826$ | $\bar{Y} = 69.1478$

$\sum_{i=1}^{n}(X_i)^2 = 364{,}023.30$ | $\sum_{i=1}^{n} X_i Y_i = 319{,}943.18$

$\sum_{i=1}^{n}(X_i - \bar{X})^2 = 23{,}153.3861$ | $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 94{,}525.1748$

$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = 46{,}131.6183$ | $\sum_{i=1}^{n} \hat{u}_i^{\,2} = 2{,}610.9211$

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.

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

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = 69.1478 - (\underbrace{1.9924 \times 86.0826}_{171.5110}) = -102.3632$$

> plugging the estimators into the equation, we have

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

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

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

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

$$r^2 = 1 - \frac{\Sigma \hat{u}_i^2}{\Sigma(Y_i - \bar{Y})^2} = 1 - \underbrace{\frac{2{,}610.9211}{94{,}525.1748}}_{0.0276} = 0.9724$$

> 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.

$$\hat{Y}_i = -102.3632 + \underbrace{1.9924(60)}_{119.544} = 17.1808$$

> When $X_i = 60$, $\hat{Y}_i$ is 17.1808

$n = 46$ $\quad$ $\sum_{i=1}^{n} X_i = 3{,}959.80$ $\quad$ $\sum_{i=1}^{n} Y_i = 3{,}180.80$

$\bar{X} = 86.0826$ $\quad$ $\bar{Y} = 69.1478$

$\sum_{i=1}^{n} (X_i)^2 = 364{,}023.30$ $\quad$ $\sum_{i=1}^{n} X_i Y_i = 319{,}943.18$

$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 23{,}153.3861$ $\quad$ $\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 94{,}525.1748$

$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 46{,}131.6183$ $\quad$ $\sum_{i=1}^{n} \hat{u}_i^{\,2} = 2{,}610.9211$

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

$$\mathrm{var}(u_i) = \mathrm{var}(u_i|X_i) = \sigma^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{2{,}610.9211}{46-2} = 59.3391$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 = \frac{364{,}023.30}{46(23{,}153.3861)}(59.3391) = 20.2814$$

$x_i = (X_i - \bar{X})$

$$\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{59.3391}{23{,}153.3861} = 0.0026$$

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

> $P[\hat{\beta}_2 - \sigma \le \beta_2 \le \hat{\beta}_2 + \sigma] = 1-\alpha$

> $P\left[-t_{\frac{\alpha}{2}} \le \dfrac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \le t_{\frac{\alpha}{2}}\right] = 1-\alpha$ $\quad t_{\frac{\alpha}{2}} = t_{\frac{.05}{2}} = t_{.025}$

> $P[\hat{\beta}_2 - (t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}) \le \beta_2 \le \hat{\beta}_2 + (t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2})] = 1-\alpha$

> $P[1.9924 - (2.021 \times 0.0506) \le \beta_2 \le 1.9924 + (2.021 \times 0.0506)] = 0.95$

> $P[1.8901 \le \beta_2 \le 2.0947] = .95$

We can make sure that 95% of β_2 is between 1.8901 and 2.0947.

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.

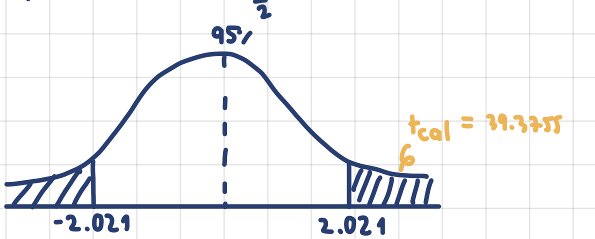
step 1 > $H_0: \beta_2 = 0$ → Null Hypothesis

$H_a: \beta_2 \ne 0$ → Alternative Hypothesis

step 2 > $t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} = \dfrac{1.9924 - 0}{0.0506} = 39.3755$

step 3 > Lower bound : $t_{\frac{\alpha}{2}} = -2.021$

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



step4 > We are sure that β_2 is not zero 95 out of 100 times when we sample or

We can reject the null hypothesis,

at the significance level of 95%

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.

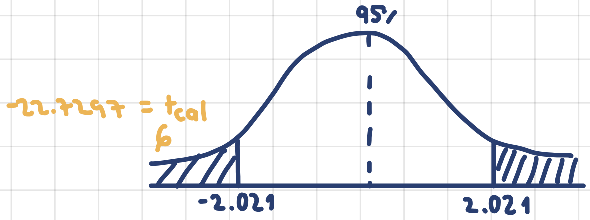
step 1 > $H_0 : \beta_1 = 0$ → Null Hypothesis

$H_a : \beta_1 \neq 0$ → Alternative Hypothesis

step 2 > $t_{cal} = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma \hat{\beta}_1} = \dfrac{-102.3632 - 0}{4.5035} = -22.7297$

step 3 > Lower bound : $t_{\frac{\alpha}{2}} = -2.021$

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



step4 > We are sure that β_1 is not zero 95 out of 100 times when we sample OR

We can reject the null hypothesis, at the significance level of 95%.

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?

No, because in order to create the slope of the line or even the intercept, we have to get the mean from the set of data so that we can estimate the line and it becomes SRF.

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.

Yes, β_2 is sufficient for us to believe that X and Y are causally related since X represents the independent variable and Y as dependent variable, described as regression of Y on X. And β_2 is the estimated coefficients of the explanatory variables indicating a change on response variable. For example, If the β_2 is positive, for every 1-unit increase in the predictor variable, the outcome variable will increase by the β_2 value.

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 we found that β_2 is significantly different from zero, wo would suggest to reject the null hypothesis or we are sure that β_2 is not zero when we sample

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

The advantage of an interval estimation over point estimate is that since a confidence level can be specified. For a specific confidence interval, the larger the sample size, the smaller the margin of error will be.

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)

$$\ln Y = \beta_1 + \beta_2 X_i$$

$$= 7.6581 + 0.0318(0)$$

$$= 7.6581$$

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

$$\ln \widehat{wage}_i = 7.6581 + 0.0318\, main\_hr$$

> $\frac{d \ln \widehat{wage}_i}{d\, main\_hr} = 0.0318 \Rightarrow \frac{d\widehat{wage}_i}{wage} = 0.0318\, d\, main\_hr$

> multiply both sides with 100

> $\% \Delta \widehat{wage}_i = 0.0318\, d\, main\_hr \times 100 = 3.18\%$

> If hours work per week increase by 1 percent, we expect wages to increase by 3.18%..

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. days work per week, lwage in Thai Baht

> $\ln \widehat{wage}_i = 7.6581 + 0.0318(24)\, main\_hr = 7.6581 + 0.7632\, main\_hr$

> Interpret meaning of β_2 : work increases by 1 day, we expect wage to increase by 0.7632

Interpret meaning of β_1 : work is equal to zero day, we expect wage to be 7.6581