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: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID(0, \sigma^2)$, 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 $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$

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

$\hat{\beta}_2 = \dfrac{\sum(x-\bar{x})(Y-\bar{Y})}{\sum(x-\bar{x})^2}$ $\quad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{x}$

$\hat{\beta}_2 = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x_i)^2}$

$\hat{\beta}_2 = \dfrac{46{,}137.6783}{23{,}153.3867} = 7.9924$ $\quad \hat{\beta}_1 = 69.1478 - 7.9924(86.0826) = 69.1478 - 171.511 = -102.3632$

$\hat{Y}_i = -102.3632 + 7.9924 X_i$

∴ Then, our SRF is $\hat{Y}_i = -102.3632 + 7.9924X_i$, meaning that the intercept is -102.3632 and slope of the function is $7.9924X_i$ #

b) $r^2 = \dfrac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2} = \dfrac{(\sum(x-\bar{x})(y-\bar{y}))^2}{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}$

$= \dfrac{(46{,}137.6783)^2}{23{,}153.3861 \times 94{,}525.1748}$

$= \dfrac{2{,}128{,}126{,}206.9769}{2{,}188{,}577{,}968.3744} = 0.9723 = 97.23\%$

∴ r^2 is the coefficience of determination. the implication is to measure the fitness of the proportion of the total variation in y explained by the regression model. In this case 97.23% of y is explained by the regression model #

C. $\hat{Y}_i = -102.3632 + 7.9924(60)$ ∴ Replacing X_i with 60

$= -102.3632 + 179.544$ the answer we get is $\hat{Y}_i = 77.1808$

$= 77.1808$ when $X_i = 60$, the average of Y_i will be 77.1808 #

d) $\hat{\sigma}^2 : Var(u_i) = \dfrac{\sum \hat{u}_i^2}{n-k} = \dfrac{\sum(Y_i - \hat{Y})^2}{n-k} = \dfrac{2{,}610.9211}{46-2} = 59.3391$ #

$Var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \dfrac{\sum X^2}{n\sum(X_i - \bar{X})^2}\sigma^2 = \dfrac{364{,}023.30(59.3391)}{46(23{,}153.3867)} = 20.2780$ #

$var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \dfrac{\sigma^2}{\sum(x-\bar{x})^2} = \dfrac{59.3391}{23{,}153.3867} = 0.00256$ #

6204647697 Korupat Udomchoksakul

e). $\alpha : 0.05$

$t_{\frac{\alpha}{2}, n-k}$ DF: $n-k : 46-2 : 44$ $t_{\frac{\alpha}{2}, 44} : 2.0154$

$$\hat{\beta}_2 - (t_{\frac{\alpha}{2}, n-k} \cdot \sigma_{\hat{\beta}_2}) \leq \beta_2 \leq \hat{\beta}_2 + (t_{\frac{\alpha}{2}, n-k} \cdot \sigma_{\hat{\beta}_2})$$

$1.9924 \pm (2.0154)(\sqrt{0.00256})$

$1.9924 \pm (2.0154)(0.0506)$

1.9924 ± 0.10198

$1.89042 \leq \beta_2 \leq 2.09438$

∴ We can estimate that, 95 out of 100 times, β_2 value will be between 1.89042 and 2.09438

f) $H_0 : \beta_1 = 0$

$H_a : \beta_1 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}}$$

$$= \frac{-102.3632 - 0}{\sqrt{20.2780}}$$

$$= \frac{-102.3632}{4.5031}$$

$$= -22.7317$$

$H_0 : \beta_2 = 0$

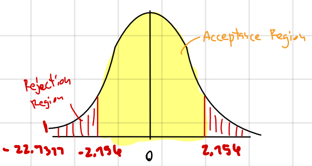
$H_a : \beta_2 \neq 0$

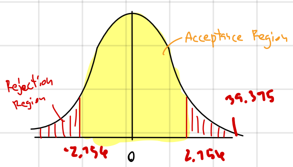
$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}$$

$$t_{cal} = \frac{1.9924 - 0}{0.0506}$$

$t_{cal} = 39.375$

$t_{\frac{\alpha}{2}, 44} = \pm 2.154$





Both t_{cal} is beyond the critical value, therefore, we can reject the null hypothesis. In other words, we can say for sure that both β_1 and β_2 is not equal to 0 95 out of 100.

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?

a) If there is only one data point, we cannot create a sample regression function, ^cause 2 data point is needed for a sample regression function.

b) It is sufficient since β_2 is the slope of population regression function
If $\beta_2 = 0$ we can claim that the population regression function is a holizontal line
and $\beta_2 = 2$, it implies that when x increase by 1, y increase by 2

c). The reason why we usually test against zero is that we want to make sure that β_2 is not zero
In other words, when β_2 is not zero, X and Y are said to be related

d.) An estimate of a population parameter given by a single number is called a point estimate of the parameter
An estimate of a population parameter given by 2 numbers between which the parameter may be considered to lie is called interval estimate of the parameter

point estimation gives us a particular value as an estimate of the population parameter.
Interval estimation gives us a range of value which is likely to contain the population parameter.
Interval estimation is more preferable because it gives the range of value of population parameter that is more accurate than that point estimation provides.

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.

a) $lwage = 7.658082 + 0.0318017 x_i$

$x = 0$; $\ln wage = 7.658082 + 0.0318017(0)$

$\ln wage = 7.658082$

$wage = e^{7.658082} = 2,117.692$

∴ Assume that a person works 0 hour a week, he will get initial wage for 7.658082 baht

b) $\ln \hat{wage} = 7.658082 + 0.0318017(1)$

$\ln \hat{wage} = 7.689883$

$wage = e^{7.689883} = 2,186.119$

$wage_{x=1} - wage_{x=0} = 2186.119 - 2,117.642 = 68.427$

∴ If a person work one more hour, he will get additional average wage 68.427 baht.

c) $\ln \hat{wage} = 7.658082 + \frac{0.0318017}{24}(24 x_i)$

$se = (0.1256392) \quad \left(\frac{0.003312}{24}\right)$

x_i = days worked (24 hours) ; $\ln \hat{wage} = 7.658082 + 0.001325 (x_i)$

$se = (0.1256392) \quad (0.000138)$

∴ If we scale this x up from 1 hour to a day (24 hours), It means that if a person works for a day (24 hours), then, the scale of working time will be multiplied by 24. The scale of wage must follow the unit of working time. In other words, the depreciation is 24 times less compared to the original data.

Therefore, the coefficient $\hat{\beta}_2$ and its standard error must be divided by 24