Metasith Pajareya 6304640623

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

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.

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{319{,}943.18}{364{,}023.3} = \underline{\underline{0.8789}}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x} = 69.1478 - 0.8789 \cdot 86.0826 = \underline{\underline{-6.51}}$$

$\hat{\beta}_2$ and $\hat{\beta}_1$ is the estimators which create the SRF with the smallest value of u_i possible. The SRF function is $y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i$,
meaning $\hat{\beta}_1$ is an estimator of β_1, which is the y-axis intercept.
$\hat{\beta}_2$ is an estimator of β_2 which is the slope of SRF.

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

$$R^2 = 1 - \frac{\sum u_i^2}{\sum (y_i - \bar{y})} = 1 - \frac{2{,}610.9211}{94{,}525.1748} = \underline{\underline{0.973}}$$

R^2 measures the goodness of fit between a SRF to data points y_i
If $R^2 = \frac{ESS}{TSS} = 97.3\%$, then 97.3% of the variation in Y_i can be explained by the model.

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

$$\hat{y}_i = -6.5 + 0.879(x_i)$$
$$= -6.5 + 0.879(60)$$
$$\hat{y}_i = \underline{\underline{46.24}}$$

$\hat{y}_i = 46.24$ means that we estimate that the average value of multiple Y_i when $x_i = 60$ would be 46.24, and SRF would pass through.

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

$$\sigma^2 = \operatorname{var}(U_i) = \frac{\sum \hat{u}_i^2}{n-k} = \frac{2610.9211}{46-2} = \underline{\underline{59.3391}}$$

$$\sigma^2_{\hat{\beta}_1} \quad \operatorname{var}(\hat{\beta}_1) = \frac{\sigma^2}{n} = \frac{59.3391}{46} = \underline{\underline{1.29}}$$

$$\sigma^2_{\hat{\beta}_2} \quad \operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{59.3391}{364{,}023.3} = \underline{\underline{1.63 \times 10^{-4}}}$$

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

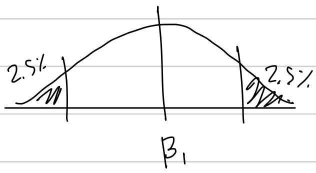
$$df = n - k = 46 - 2 = 44$$

$$\alpha = 0.05$$

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

$$t_{0.025} = 2.021$$

$df \approx 40$



upper $\longrightarrow$ $0.879 + 2.021 \times \sqrt{1.63 \times 10^{-4}}$

$= 0.879 + 0.0258$

$= \underline{\underline{0.9048}}$

lower $\longrightarrow$ $0.879 - 0.0258 = 0.8532$

Ans: $\longrightarrow$ $0.8532 \leq \beta_2 \leq 0.9048$

We are 95% certain that β_2 would be between 0.9048 and 0.8532

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.

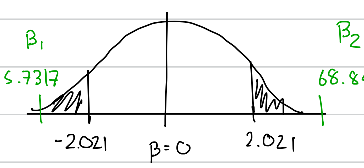
Null hypothesis $\beta_1 = 0$, $\beta_2 = 0$ Alternative Hypothesis $\beta_1 \neq 0$, $\beta_1 \neq 0$

For β_1 $\longrightarrow$ $t_{cal} = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = -\dfrac{6.51 - 0}{\sqrt{1.29}} = -5.7317$

For β_2 $\longrightarrow$ $t_{cal} = \dfrac{0.8789 - 0}{\sqrt{1.63 \times 10^{-4}}} = 68.84$

$t_{\frac{\alpha}{2}}$ upper $= 2.021$

$t_{\frac{\alpha}{2}}$ lower $= -2.021$



$\longrightarrow$ Both in rejection area, reject H_0

Ans: We can say sure that β_1 & β_2 is not 0 95 out of 100 times we take the 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?

No, because a SRF shows a relationship between multiple sets of multiple independent variable (Y) and multiple dependent variable (X), but with one data point, a function that shows a relationship between multiple data points cannot be drawn.

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, although a significant β_2, or $\beta_2 \neq 0$, means that X and Y are related. But there are cases that the relationship between X and Y is just a coincidence and not a causal one. For instance, even if income is related to consumption, we can't conclude that an increase income causes an increase in consumption without additional information.

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.

Let's say the level of significant is 95%, β_2 is significantly different from 0 means that according to the result, we cannot reject the null hypothesis of $\beta_2 = 0$. Simply put, we cannot say for sure that β_2 is not 0 as out of 100 times when we sample. Thus, we can conclude that the 2 variables within the SRF is related

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

Interval estimation provides extra information about the estimation. It contains not just the point where the parameter is likely to be, but also the accuracy of the estimation, by pointing out at what range the parameter would be based on the significant level.

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.

y

x

```
      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)

$$\ln \widehat{wage} = 7.658082 + 0.0318017 \underbrace{hours}_{=0}$$

$$\ln \widehat{wage} = 7.658082$$

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

$$\widehat{wage} = \underline{\underline{2117.6918}}$$

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

$$\ln \widehat{wage} = 7.658082 + 0.0318017 \text{ hours}$$

$$\frac{d \ln \widehat{wage}}{d \text{ hour}} = 0.0318017$$

$$\frac{d \widehat{wage} / wage}{d \text{ hour}} =$$

$$\frac{d \widehat{wage}}{wage} \times 100 = 0.0318017 \times 100 \; d \text{ hour}$$

$$\Delta\% \text{ wage} = 3.18017 \; d \text{ hour}$$

Ans: $\underline{\underline{\text{wage increase } 3.18017\%}}$

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.

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
main_hr	.0318017	.003312	9.60	0.000	.0252844	.0383189
Ans:	0.7632408	0.079488	no change	no change	0.6055	0.9209

$$\ln \hat{y} = \hat{\beta}_1 + \underset{\times 24}{\hat{\beta}_2} x \qquad \underset{\times 24}{SE_{\beta_2}}$$

new $\sigma_{\hat{\beta}_2} = 0.079488$ new $\hat{\beta}_2 = 0.7632408$

$df = 308 - 2$

$= 306$

└ use $df = 100$

confidence interval $= \hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$

upper $= 0.7632408 + 1.984\,(0.079488) = 0.9209$

lower $= 0.7632408 - 1.984\,(0.079488) = 0.6055$