6304640839
Nuttarika

Assignment ①

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

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

a) $Y_i = \beta_1 + \beta_2 X_i + u_i, \; u_i \sim NIID(0, \sigma^2)$

estimators $= \beta_1$ & β_2

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

$$\hat{\beta}_2 = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

$$\hat{\beta}_2 = \frac{319{,}943.18}{364{,}023.30} = 0.8789085$$

≈ 0.8789 ※

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 69.1478 - (0.8789085)(86.0826)$

$= -6.5102$ ※

$\hat{\beta}_2$ represents the estimation of the slope of the graph while $\hat{\beta}_1$ represents the estimation of the Y intercept

∴ The Y-intercept estimated to be -6.5012 and estimated slope is 0.8789

c) $X_i = 60$

$\hat{Y}_i = ?$

Y_i or $\hat{Y} = \bar{Y}$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

$-6.5102 = \bar{Y} - (0.8789085)(60)$

$\bar{Y} = -6.5102 + [(0.8789085)(60)]$

$\bar{Y} = 46.22431$

≈ 46.2243 ※

∴ when one of point changes to 60 ($X_i = 60$), the estimated point Y_i changes to 46.2243 as well.

b $r^2 = 1 - \frac{RSS}{TSS}$ $\quad \therefore 0 \leq r^2 \leq 1$

$= 1 - \frac{\Sigma \hat{u}^2}{\Sigma(Y_i - \bar{Y})^2}$

$= 1 - \frac{2{,}610.9211}{94{,}525.1748}$

$= 0.97237856$

≈ 0.9724 ※

r^2 tells the measurement of 'goodness of fit' of the lifted regression line comparing to an estimator.

∴ X_i explains about 97.24% of variation in Y_i

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

$$var(u_i) = \frac{\Sigma \hat{u}_i^2}{n-k}$$

$$= \frac{2,610.9211}{46-2}$$

$$= 59.3391159$$

$$\approx 59.3391 ※$$

$$var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1}$$

$$= \frac{\Sigma X_i^2}{n \Sigma x_i^2} \sigma^2$$

$$= \frac{364,023.30}{46(364,023.30)} (59.3391)$$

$$= 1.28998$$

$$\approx 1.2900 ※$$

$$var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2}$$

$$= \frac{\sigma^2}{\Sigma x_i^2}$$

$$= \frac{59.3391159}{364,023.30}$$

$$= 0.000163009$$

$$\approx 0.0002 ※$$

e) CI = 95%

$\alpha = 0.05$

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

$$= 0.8789085 \pm t_{0.025} \cdot 0.000163009$$

$$= 0.8789085 \pm (2.021)(0.000163009)$$

$$= 0.8789085 \pm 0.000329441189$$

$$= (0.878579 , 0.8792379)$$

∴ There are 95% confident intervals that $\hat{\beta}_2$ would lies between 0.878579 and 0.8792379

f) $H_0 : \beta_2 = 0$ $\alpha = 0.05$

$H_a : \beta_2 \neq 0$

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

$$= \frac{0.8789085 - 0}{\sqrt{\frac{59.3391159}{364,023.30}}}$$

$$= 68.83950203$$

$$\approx 68.8395 ※$$

lower bound: $\hat{\beta}_2 - t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$

$$= 0.8789085 - 2.021 \cdot \sqrt{0.000163009}$$

$$\approx 0.8531 ※$$

upper bound $= \hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$

$$= 0.8789085 + 2.021 \cdot \sqrt{0.000163009}$$

$$\approx 0.9047 ※$$

$$P[0.8531 \leq \beta_2 \leq 0.9047]$$

ส่วน β_1

$H_0 : \beta_1 = 0$

$H_a : \beta_1 \neq 0$

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

$$= \frac{-6.5102 - 0}{\sqrt{1.28998}}$$

$$= -5.7319559 \approx -5.7320 ※$$

lower bound: $\hat{\beta}_1 - t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_1}$

$$= -6.5102 - 2.021 \cdot \sqrt{1.28998}$$

$$\approx -8.8056 ※$$

upper bound: $\hat{\beta}_1 + t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_1}$

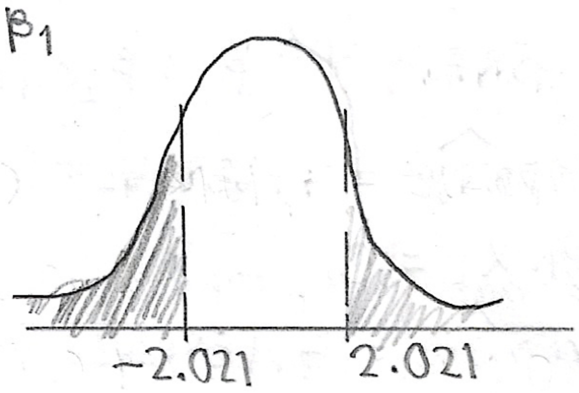
$$= -6.5102 + 2.021 \cdot \sqrt{1.28998}$$

$$= -4.214803041$$

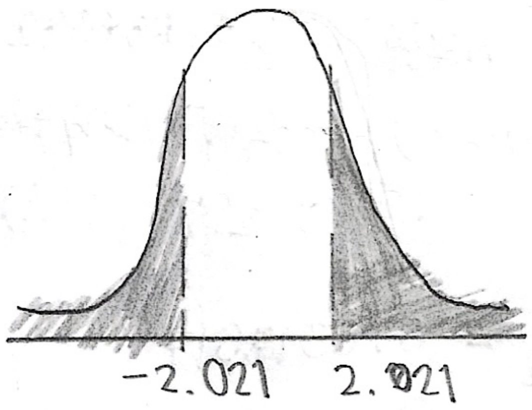
$$\approx -4.2148 ※$$

∴ The value of $t_{cal\,\beta_1}$ and $t_{cal\,\beta_2}$ both beyond the boundaries between -2.021 to 2.021, we can ensure that at 0.05 level of significant, β_1 & β_2 is not zero 95 out of 100 times when we sample.

For β_1



For β_2



(2)

a) No as the only one data point does not make up the sample. In addition, the sample regression function shows the relationship between many sets of dependent variables and independent variables. Thus, the only one data point cannot shows the relatiship between the independent and dependent variables.

b). Yes, the significant β_2 shows the slope of graph with variables X and Y. β_2 shows how much Y would increase when X changes. For example $\beta_2 = 5$, it suggests that when x increases by 1 Y would increased by 5 times of X.

$$(\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i , \quad \hat{Y} = 0 + 5(1) = 5 \qquad \hat{Y} = 0 + 5(2) = 10)$$

c) The result of the hypothesis roughly suggests about the how the 2 variables related within the SRF. For example, the level of significant is 90%, β_2 is significantly different from 0. According from the result, we cannot reject the null hypothesis which is $\beta_2 = 0$. Therefore, we cannot say for sure that β_2 is not 0 out of 100 times we sample. Hence, the 2 variables within the SRF is related.

d) The interval estimation shows wider range of information It covers more points and lower chance of missing any points unlike point estimate which could miss out point that have some distance away. Therefore, it suggests the range that the parameter would be based on the significant level.

③ a) SRF : $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i$

$\widehat{\ln wage} = 7.658082 + 0.0318017 X_i$

when $X_i = 0$

$\widehat{\ln wage} = 7.658082$

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

$= 2117.691798$

≈ 2117.6918 #

b) $\hat{\beta}_2 = 0.0318017$

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

% change $= 0.0318017(100)$

$= 3.18017\ \%$

∴ If a person works an hour more, the wage would increase around 3.1802 %

c) 1 day = 24 hours

$\ln \widehat{wage} = \hat{\beta}_1 + \hat{\beta}_2 X_i$

SE_{β_2} ก็ต้องเปลี่ยน

change the unit to day by × 24

new $\hat{\beta}_2 = 0.0318017(24)$

$\approx 0.7632408 \approx 0.7632$ *

new $SE_{\hat{\beta}_2} = 0.003312(24)$

$= 0.079488 \approx 0.0795$ *

หา confindent interval ใหม่

$df = n - k$

$= 308 - 2$

$= 306$ (use df = 100 in the table)

t table upper & lower boundary

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

upper boundary $= \hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$

$= 0.7632408 + 1.984(0.079488)$

$= 0.92094499$

≈ 0.9209 *

lower boundary : $\hat{\beta}_2 - t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2}$

$= 0.7632408 - 1.984(0.079488)$

$= 0.605536608$

≈ 0.6055 *

∴ The 95 % confident intervals that the $\hat{\beta}_2$ lies is between 0.6055 to 0.9209