Napon Charuwattanasakul
6304640680

1)

a) $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{319{,}943.18}{364{,}023.30} = 0.8789$ # $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 69.1478 - (0.8789 \times 86.0826) = -6.5101$ #

$\therefore \hat{Y}_i = -6.5101 + 0.8789 x_i$

The intercept of the model is -6.5101 and the slope is 0.8789.

b) $R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})} = 1 - \frac{2{,}610.9211}{94{,}525.1748} = 0.9724$ #

R^2 is a measure for the goodness of fit and 97.24% of the variation in Y_i can be explained by the model.

c) $\hat{Y}_i = -6.5101 + 0.8789(60) = 46.2239$ #

It means that when $x_i = 60$, the average of $\hat{Y}_i$ will be 46.2239.

d) $\sigma^2 = \frac{\sum \hat{u}_i^2}{n-k} = \frac{\sum (Y_i - \hat{Y})^2}{n-k} = \frac{2610.9211}{46-2} = 59.3391$ #

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

$\sigma^2_{\hat{\beta}_2} = var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{59.3391}{364{,}023.30} = 1.63 \times 10^{-4}$ #

e) $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}} \cdot \sigma_{\hat{\beta}_2} = 0.8789 \pm 0.681 \times 1.63 \times 10^{-4}$

$0.8789 - 2.021 \times \sqrt{1.63 \times 10^{-4}} \leq \beta_2 \leq 0.8789 + 2.021 \times \sqrt{1.63 \times 10^{-4}}$

$0.8789 - 0.0258 \leq \beta_2 \leq 0.8789 + 0.0258$

$0.8530 \leq \beta_2 \leq 0.9047$ #

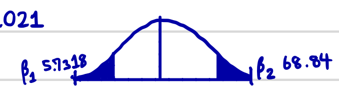
$\therefore$ It's 95% certain that β_2 would be between 0.8532 and 0.9048.

f) Null hypothesis : $\beta_1 = 0, \beta_2 = 0$ / Alternative hypothesis: $\beta_1 \neq 0, \beta_2 \neq 0$

$\beta_1 : t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = \frac{-6.5101 - 0}{\sqrt{1.29}} = -5.7318$

$\beta_2 : t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} = \frac{0.8789 - 0}{\sqrt{1.63 \times 10^{-4}}} = 68.84$

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



We can be 95% certain that β_1 and β_2 is not 0.

Napon Charuwattanasakul
6304640680

2)

a) No, we can't because an SRF represents a relationship between several sets of multiple independent variables Y and multiple dependent variables X, but a function showing a relationship between multiple data points cannot be created with only one data point.

b) No, they aren't because β_2 shows that X and Y are related but in some cases, the relationship between X and Y is just coincidence. For example, even though income is related to consumption, we can't assume that an increase in income causes an increase in consumption.

c) If the level of significant is 95% and β_2 is different from 0, we can't say for sure that β_2 is not 0 and we can't 100% conclude that β_2 is 0 as well.

d) Interval estimation shows extra information about the estimation. It has the accuracy of the estimation based on significant level. So, it's more accurate than estimate.

3)

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

$\ln \widehat{wage} = 7.658082$

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

$\widehat{wage} = 2117.6918$ #

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

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

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

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

$\therefore$ wage increase 3.18017% #

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

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

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

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