Assignment 1
Assigned on Feb 8th, 2022. Due date Saturday, Feb 26th, 2022 before midnight.

Tha Taritta Suppasadid 6304641571

Instructions

(1) Please read the instruction carefully. Also take this habit with you into the exam room.
(2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at lease a paragraph for your analysis, or a graph when necessary, even when the question does not indicate so.
(3) Handing and submitting assignments are only available via BE Moodle.

Answering the questions and preparing answer sheets

(1) Answers are to be handwritten, in either digital or analog form, in a blank canvas or any clean paper. Make sure that your handwriting is clearly visible and readable.
(2) There is no need to rewrite the question. Just indicate the question number clearly for each of the answer, such as 1.a).
(3) Default decimal point is 4.
(4) Choose precise wordings, especially when you want to interpret the meaning of a test, confidence interval, or coefficients.
(5) When done, for the digital case, collage all the pages into a single PDF file. For those who write on sheets of paper, take photo of all pages then convert all of them into a single PDF file as well.
(6) Name your PDF file as StudentID_YourNickname, such as 640123456_Bo.

Submitting your answers

(1) Make sure your file does not exceed 10MB. This is the maximum file size for BE Moodle upload.
(2) Login to BE Moodle, head into the course, then the assignment topic.
(3) Choose your file to submit. Done. There will be timestamp for your upload date and time, so please make sure to not submit later than that.

Assignment 1
Assigned on Feb 8th, 2022. Due date Saturday, Feb 26th, 2022 before midnight.

For all questions, answer up to 4 decimal places

Question 1. (15 points) Given this information

$$n = 18 \qquad \sum_{i=1}^{n} X_i = 388.00 \qquad \sum_{i=1}^{n} Y_i = 50.90$$

$$\sum_{i=1}^{n} (X_i)^2 = 9{,}620.00 \qquad \sum_{i=1}^{n} X_i Y_i = 1{,}254.90$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 211.00 \qquad \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 2.5844$$

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 20.58 \qquad \sum_{i=1}^{n} \hat{u}_i^{\,2} = 0.5781$$

Use the above sample information to answer all the following questions. Show explicitly all formulas and calculations.

a) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID\left(0, \sigma^2\right)$, **find the estimators** of β_1 and β_2 with OLS method. Interpret the intercept and slope coefficients.

b) Compute the value of R^2 and explain its meaning.

c) If $X_i = 30$, estimate the value of $\hat{Y}_i$ and explain its meaning.

d) Calculate the estimators of $\text{var}\left(u_i\right)$, $\text{var}\left(\hat{\beta}_1\right)$ and $\text{var}\left(\hat{\beta}_2\right)$.

e) What are the 90-percent confident intervals for β_2 ? Interpret the meaning.

f) Test the hypothesis whether the slope coefficients are different from zero at 0.05 level of significance.

Question 2. Using the 2015 Health and Welfare Survey from the National Statistical Office, a simple linear regression is modeled as follows,

$$outp_i = \beta_1 + \beta_2 age_i + u_i$$

where $outp_i$ is how many times person i has visited hospital in 2015, from 0 to 7 times

age_i is how old is person i, from 0 to 97 years.

We assume that both $outp_i$ and age_i are continuous, the estimation results in the following table. Answer the following questions and show your work.

```
      Source |       SS           df       MS      Number of obs   =    27,886
-------------+----------------------------------   F(1, 27884)     =    186.96
       Model |  77.5444409         1  77.5444409   Prob > F        =    0.0000
    Residual |  11565.0627    27,884  .414756231   R-squared       =    0.0067
-------------+----------------------------------   Adj R-squared   =    0.0066
       Total |  11642.6072    27,885  .417522223   Root MSE        =    .64402

------------------------------------------------------------------------------
        outp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0031338   .0002292                      .0026846     .003583
       _cons |   .4279898   .0140339      Omitted         .4004828    .4554969
------------------------------------------------------------------------------
```

a) Test if both parameters are significantly different from zero or not. Use $\alpha = 0.05$.

b) Interpret the meaning of $\hat{\beta}_2$. Does the sign of $\hat{\beta}_2$ make economic sense? Explain.

c) If $outp_i$ is turned into natural logarithmic scale (ln), how would you reinterpret the relationship between $\hat{\beta}_2$ and $\widehat{outp}_i$, assumed that the given coefficient given in the table above can be used to interpret this new functional form.

d) If age_i variable is divided by 10, how does it affect both the coefficients, standard errors, and confidence intervals? Answer the changes of both the constant and slope (if there is).

e) Find the confidence interval of mean prediction at the age of 50 years old, given that $var(\hat{Y}_0) = 0.00002$ and $\alpha = 0.01$.

Question 3. Discuss in a short paragraph why the confidence interval for both the mean prediction and individual prediction get larger as the X_0 is further away from $\bar{X}$.

---

a) 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. Interpret the intercept and slope coefficients.

- $\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{20.58}{211.00} = 0.0975$

- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \frac{50.90}{18} - (0.0975 \cdot \frac{388.00}{18})$

$= 2.8278 - 2.1017 = 0.7261$

intercept $= \hat{\beta}_1 = 0.7261$ slope $= \hat{\beta}_2 = 0.0975$ #

b) Compute the value of R^2 and explain its meaning.

- $R^2 = 1 - \frac{\sum_{i=1}^{n}\hat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \left(\frac{0.5781}{2.5844}\right)$

$= 1 - 0.2234 = 0.7763$

R^2 is a measure of goodness of fit, which means that 77.63% of Y is explained by this regression model. #

c) If $X_i = 30$, estimate the value of $\hat{Y}_i$ and explain its meaning.

$E(\hat{Y}_i | X_i = 30)$: $\hat{Y}_i = 0.7261 + 0.0975(30) = 3.6511$

If $X_i = 30$, the average $\hat{Y}_i$ will be 3.6511 #

d) Calculate the estimators of $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$.

- $\text{var}(U_i) = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{U}_i^2}{n-k} = \frac{0.5781}{18-2} = 0.0361$ #

- $\text{var}(\hat{\beta}_1) = \hat{\sigma}^2_{\hat{\beta}_1} = \frac{\sum_{i=1}^{n}(X_i)^2}{n\sum_{i=1}^{n}(X_i - \bar{X})^2}\hat{\sigma}^2 = \frac{9620.00}{18 \cdot 211.00}(0.0361)$

$= 2.5329(0.0361) = 0.0914$ #

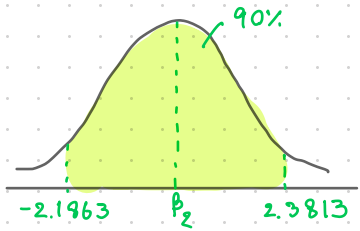
- $\text{var}(\hat{\beta}_2) = \hat{\sigma}^2_{\hat{\beta}_2} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{0.0361}{211.00} = 1.7109$ #

e) What are the 90-percent confident intervals for β_2 ? Interpret the meaning.

① finding 90% CI $\Rightarrow$ $\alpha = 0.1$

② $df = n - K = 18 - 2 = 16$

$t_{\frac{\alpha}{2}} = t_{0.05} = 1.746$



③ Upper limit : $\hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_2} = 0.0975 + (1.746 \cdot \sqrt{1.7109})$

$= 0.0975 + 2.2838 = 2.3813$

lowwer limit : $\hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_2} = 0.0975 - (1.746 \cdot \sqrt{1.7109})$

$= 0.0975 - 2.2838 = -2.1963$

- $P[-2.1963 \leq \hat{\beta}_2 \leq 2.3813] = 0.90$ #

f) Test the hypothesis whether the slope coefficients are different from zero at 0.05 level of significance.

① $H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

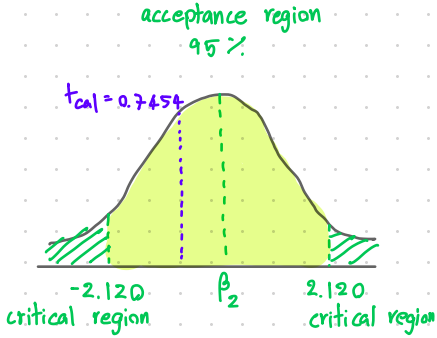
② pick α : $\alpha = 0.05$ $(df = n - k = 18 - 2 = 16)$

③ $t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} = \frac{0.0975 - 0}{\sqrt{1.7109}}$

$= \frac{0.0975}{1.3080} = 0.7454$

④ upper bound : $t_{\frac{\alpha}{2}} = t_{0.025(16)} = 2.120$

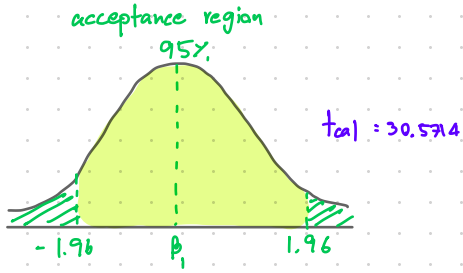
lowwer bound : $t_{\frac{\alpha}{2}} = -2.120$



$\therefore$ Accept H_0 : We are sure that slope coefficients (β_2) are different from zero 95% of the time we sample #

a) Test if both parameters are significantly different from zero or not. Use $\alpha = 0.05$.

- $\hat{\beta}_1 = 0.4280$



① $H_0 : \beta_1 = 0$

$H_f : \beta_1 \neq 0$

② $\alpha = 0.05$ $\quad df = \infty$

③ $t_{cal} : \dfrac{\hat{\beta}_1 - \beta_1}{se_{\hat{\beta}_1}} = \dfrac{0.4280 - 0}{0.0140}$

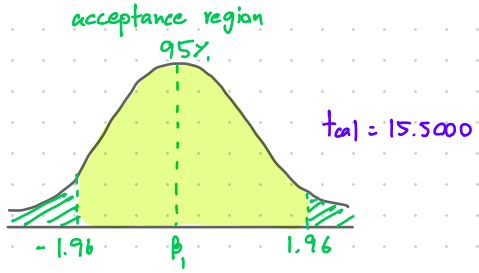
$= 30.5714$

④ Upper bound : $t_{\frac{\alpha}{2}} = t_{0.025(\infty)} = 1.96$

lowwer bound : $t_{\frac{\alpha}{2}} = -1.96$

∴ We are sure that $\beta_1 \neq 0$ 95% of the time we sample #

- $\hat{\beta}_2 = 0.0031$



① $H_0 : \beta_2 = 0$

$H_f : \beta_2 \neq 0$

② $\alpha = 0.05$ $\quad df = \infty$

③ $t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} = \dfrac{0.0031 - 0}{0.0002}$

$= 15.5000$

④ Upper bound : $t_{\frac{\alpha}{2}} = t_{0.025(\infty)} = 1.96$

lowwer bound : $t_{\frac{\alpha}{2}} = 1.96$

∴ We are sure that $\beta_2 \neq 0$ 95% of the time we sample #

b) Interpret the meaning of $\hat{\beta}_2$. Does the sign of $\hat{\beta}_2$ make economic sense? Explain.

The meaning of $\hat{\beta}_2 = 0.0031$ is when people grow older 1 more year, it can be expected that people will increase time visit the hospital 0.0031 times per year. The positive of $\hat{\beta}_2$ makes economic sense because when people get older, they tend to need medical services more #

c) If $outp_i$ is turned into natural logarithmic scale (ln), how would you reinterpret the relationship between $\hat{\beta}_2$ and $\widehat{\ln outp}_i$, assumed that the given coefficient given in the table above can be used to interpret this new functional form.

$$\widehat{\ln outp}_i = \hat{\beta}_1 + \hat{\beta}_2 age_i$$

$\widehat{\ln outp}_i$	1% increase	$\hat{\beta}_2$ = percent
100	1 time	0.0031 times
10000	100 times	0.3134 times

∴ If people's age increase by 1 year, it can be expected that time that people visit hospital will increase by 0.3134 % #

d) If age_i variable is divided by 10, how does it affect both the coefficients, standard errors, and confidence intervals? Answer the changes of both the constant and slope (if there is).

If age_i variable is divided by 10, the coefficients, standard errors, and confidence intervals will be multiplied by 10

Coefficient : 0.0031338 >> 0.031338

SE : 0.0002292 >> 0.002292

CI : 0.0026846 , 0.003583 >> 0.026846, 0.03583

While value of the constant remain the same. #

e) Find the confidence interval of mean prediction at the age of 50 years old, given that $var(\hat{Y}_0) = 0.00002$ and $\alpha = 0.01$.

$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 (X_i) = 0.4280 + 0.0031 (50) = 0.5830$

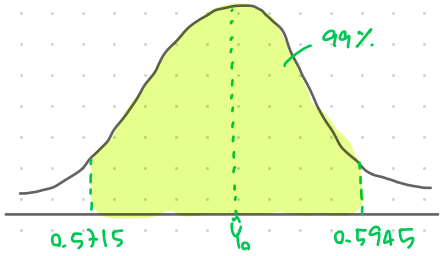
① $\alpha = 0.01 \quad df = \infty$

② $t_{\frac{\alpha}{2}} = t_{0.005(\infty)} = 2.576$

③ Upper limit : $\hat{Y}_0 + t_{\frac{\alpha}{2}} \cdot SE_{\hat{Y}_0} = 0.5830 + (2.576 \cdot \sqrt{0.00002}) = 0.5830 + 0.0115 = 0.5945$ (4.4721)

lowwer limit : $\hat{Y}_0 - t_{\frac{\alpha}{2}} \cdot SE_{\hat{Y}_0} = 0.5830 - (2.579 \cdot \sqrt{0.00002}) = 0.5830 - 0.0115 = 0.5715$

$\therefore \Pr [0.5715 \leq \hat{Y}_0 \leq 0.5945] = 0.99$ #



Question 3. Discuss in a short paragraph why the confidence interval for both the mean prediction and individual prediction get larger as the X_0 is further away from $\bar{X}$.

Using variance that used for calculating confidence interval for the mean prediction is

$$var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X - \bar{X})^2} \right]$$

From the equation, there is $(X_0 - \bar{X})^2$. Therefore, the further away from the empirical data mean for X-values, the larger the quantity will be. #