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.

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

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

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

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

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

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 $\text{var}(u_i)$, $\text{var}\left(\hat{\beta}_1\right)$ 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.

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.

$$\hat{\beta}_2 = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sum (X-\bar{X})^2} = \frac{20.58}{211.00} = 0.0975$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \qquad \bar{Y} = 2.8278 \qquad \bar{X} = 21.5556$$

$$2.8278 = \hat{\beta}_1 + 0.0975(21.5556)$$

$\hat{\beta}_1 = 0.7261$; if $X = 0$, $Y = 0.7261$

$\hat{\beta}_2 = 0.0975$ if x increase by 1 unit, Y will increase by 0.0975

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

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$ESS = \sum(\hat{Y} - \bar{Y})^2$$

$$RSS = \sum \hat{u}^2 = \sum (Y_i - \hat{Y})^2 = 0.5781$$

$$TSS = \sum (Y_i - \bar{Y})^2 = 2.5844$$

$$R^2 = 1 - \frac{0.5781}{2.5844}$$

$= 0.7763$ The total variation in Y explain by the regression 77.63%

$$Var(\hat{u}) = \hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-2} = \frac{0.5781}{16} = 0.0361$$

$$Var(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \hat{\sigma}^2 = \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} \cdot \hat{\sigma}^2 = \frac{9{,}620}{18(211)} \cdot 0.0361 = 0.0914$$

$$Var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum x_i^2} = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2} = \frac{0.0361}{211} = 0.0002$$

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

point estimate

$$\hat{Y} = 0.7261 + 0.0975(30) = 3.6511$$

individual estimate

$$Var(Y_0 - \hat{Y}_0) = \hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-2} = \frac{0.5781}{16} = 0.0361$$

$$= 0.0361\left[1 + \frac{1}{18} + \frac{(30 - 21.5556)^2}{211}\right] = 0.0503$$

$$SE(Y_0 - \hat{Y}_0) = \sqrt{0.0503}$$

$$\hat{Y} - t_{\frac{\alpha}{2}} SE(Y_0 - \hat{Y}_0) \leq \beta_1 + \beta_2(30) \leq \hat{Y} + t_{\frac{\alpha}{2}} SE(Y_0 - \hat{Y}_0)$$

We 95% confidence that $X_i = 30$ Y is on $(3.1756, 4.1266)$

$$3.6511 - 2.120(\sqrt{0.0503}) \leq \beta_1 + \beta_2(30) \leq 3.6511 + 2.120(\sqrt{0.0503})$$

$$3.1756 \leq E(Y|X=30) \leq 4.1266$$

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

$\text{var}(u_i) = 0.0361$

$\text{var}(\hat{\beta}_1) = \dfrac{\sum X_i^2}{n\sum x_i^2} \cdot \sigma^2 = \dfrac{\sum X_i^2}{n\sum (X_i - \bar{X})^2} \cdot \sigma^2 = \dfrac{9{,}620}{18(211)} \cdot 0.0361 = 0.0914$

$\text{Var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2} = \dfrac{\sigma^2}{\sum (X_i - \bar{X})^2} = \dfrac{0.0361}{211} = 0.0002$

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

0.01

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

$0.0975 - 1.746(\sqrt{0.0002}) \leq \hat{\beta}_2 \leq \hat{\beta}_2 + 1.746(\sqrt{0.0002})$

$0.0728 \leq \hat{\beta}_2 \leq 0.1222$

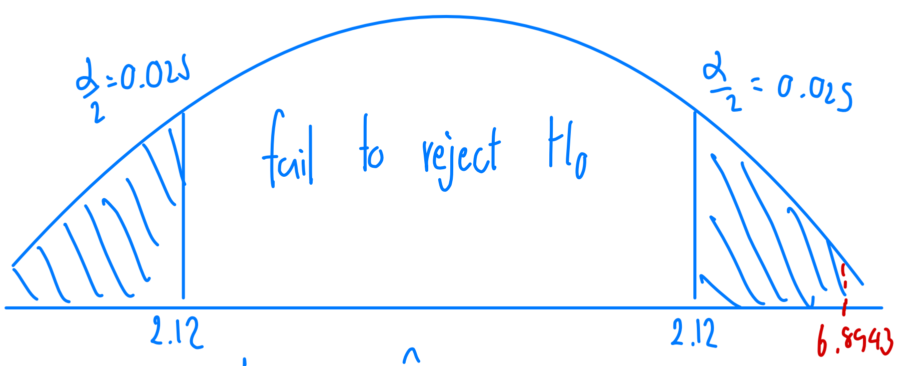
We have 90 % confidence for β_2 is on $(0.0728, 0.1222)$

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

$H_0 : \beta_2 = 0$ (not dif)

$H_1 : \beta_2 \neq 0$ (diff)

$\alpha = 0.05$



$t_{cal} = \dfrac{\hat{\beta}_2 - \beta_2}{se\ \beta_2} = \dfrac{0.0975 - 0}{\sqrt{0.0002}} = 6.8943$

$\therefore$ tcal fall in to reject H_0 region, we 95 % confidence that the slope is difference from zero.

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

$$outp_i = \beta_1 + \beta_2 age_i + u_i \rightarrow \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

where $K \begin{cases} outp_i \text{ is how many times person i has visited hospital in 2015, from 0 to 7 times} \\ age_i \text{ is how old is person i, from 0 to 97 years.} \end{cases}$

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.

TSS = ESS + RSS

Source	SS	K-1 df	MS			
Model	ESS 77.5444409	1	$\frac{ESS}{K-1}$ 77.5444409	Number of obs	=	27,886
Residual	RSS 11565.0627	n-K 27,884	.414756231 $\frac{RSS}{n-K}$	F(1, 27884)	=	186.96 $\frac{MSE}{MSS}$
				Prob > F	=	0.0000
				R-squared	=	0.0067
				Adj R-squared	=	0.0066
Total	TSS 11642.6072	n-1 27,885	.417522223 or $\frac{TSS}{n-1}$	Root MSE	=	.64402

outp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
age	β_2 .0031338	SE_{β_2} .0002292	Omitted		.0026846	.003583
_cons	β_1 .4279898	SE_{β_1} .0140339			.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) Test if both parameters are significantly different from zero or not. Use $\alpha = 0.05$.

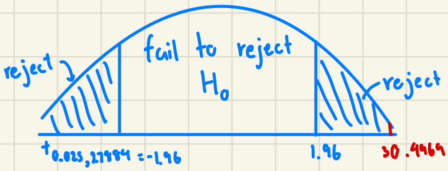
$H_0 : \beta_1 = 0$ $\quad \alpha = 0.05$

$H_1 : \beta_1 \neq 0$ $\quad \frac{\alpha}{2} = 0.025$

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

$$= \frac{0.4279898 - 0}{0.0140339}$$

$= 30.4969$ fall in to reject H_0 region.



t cal fall in reject H_0 region, β_1 different from zero at $\alpha = 0.05$ (95% confidence) #

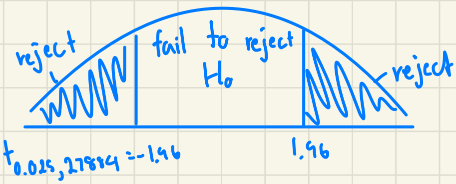
$H_0 : \beta_1 = 0$ $\quad \alpha = 0.05$

$H_1 : \beta_2 \neq 0$ $\quad \frac{\alpha}{2} = 0.025$

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

$$= \frac{0.0031338 - 0}{0.0002292}$$

$= 13.6728$ fall in to reject H_0 region



t cal fall in reject H_0 region, β_2 different from zero at $\alpha = 0.05$ (95% confidence) #

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

$\hat{\beta}_2 = 0.0031338$: Age of person increase by 1 year, Time to visit to hospital will increase 0.0031338 time.

In my opinion its make sense because If we get older, We have more chance to go to the hospital. So its mean we have to pay more when we old its reason why insurance company can have profit from this situation by using fear from people.

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.

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

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

β_1 : If age = 0 year, time person to visit hospital = ln output p_i time

β_2 : If age increase 1 year, time to visit hospital increase outp$_i$ $100\hat{\beta}_2$ %.

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).

$w_1 = 10$

$w_2 = \frac{1}{10}$

$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2 = \left(\frac{1}{\frac{1}{10}}\right)\hat{\beta}_2$

$\hat{\beta}_2^* = 10\hat{\beta}_2$

$\hat{\beta}_2$ change to $10\hat{\beta}_2$

$\hat{\beta}_1^* = w_1\hat{\beta}_1 \rightarrow \hat{\beta}_1^* = 1\hat{\beta}_1$

$\hat{\beta}_1$ not change

$Var\,\hat{\beta}_1^* = w_1^2 Var\,\hat{\beta}_1$

$= (1)^2 Var\,\hat{\beta}_1$

$SE\,\hat{\beta}_1^* = SE\,\hat{\beta}_1$

Standard error of $\hat{\beta}_1$ not change

$Var\,\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)^2 Var\,\hat{\beta}_2$

$= (10)^2 Var\,\hat{\beta}_2$

$Var\,\hat{\beta}_2^* = 100\,Var\,\hat{\beta}_2$

$SE\,\hat{\beta}_2^* = 10\,SE\,\hat{\beta}_2$

$\therefore$ Standard error of $\hat{\beta}_2^*$ change to $10\,SE\,\hat{\beta}_2$

C.I. β_1 : $\hat{\beta}_1 - t_{\frac{\alpha}{2}} SE\, \hat{\beta}_1 < \beta_1 < \hat{\beta}_1 + t_{\frac{\alpha}{2}} SE\, \beta_1$

Becuse $\hat{\beta}_1$ and SE $\hat{\beta}_1$ are not change so confidence interval will not change.

C.I. β_2 : $\hat{\beta}_2 - t_{\frac{\alpha}{2}} SE\, \hat{\beta}_2 < \beta_2 < \hat{\beta}_2 + t_{\frac{\alpha}{2}} SE\, \hat{\beta}_2$

Because $\hat{\beta}_2$ and SE $\hat{\beta}_2$ are change so confidence interval will change to $10\hat{\beta}_2 - t_{\frac{\alpha}{2}} 10\, SE\, \hat{\beta}_2 < \beta_2 < 10\hat{\beta}_2 + t_{\frac{\alpha}{2}} 10\, SE\, \hat{\beta}_2$

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

$$\hat{Y} - t_{\frac{\alpha}{2}} \sqrt{var\, \hat{Y}_0} \leq E(Y|X=50) \leq \hat{Y} + t_{\frac{\alpha}{2}} \sqrt{var\, \hat{Y}_0}$$

$$\hat{Y} - t_{\frac{\alpha}{2}} Se\, \hat{Y}_0 \leq E(Y|X=50) \leq \hat{Y} + t_{\frac{\alpha}{2}} SE\, \hat{Y}_0$$

$\hat{Y} = \beta_1 + \beta_2 X_i$

$\hat{Y} = 0.4279898 + 0.0031338 X_i$

$X_i = 50$ $\hat{Y} = 0.4279898 + 0.0031338(50) = 0.5846798$ $\frac{\alpha}{2} = 0.005$ $t_{0.005} = 2.576$

$$0.5846798 - (2.576)\sqrt{0.00002} \leq E(Y|X=50) \leq 0.5846798 + (2.576)\sqrt{0.00002}$$

$$0.5731596 \leq E(Y|X=50) \leq 0.5962$$

Age of 50 years old, Time to visit hospital average on [0.5731596, 0.5962]

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 $\bar{X}$.

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

$$var(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

from Variance of both when x_0 far from $\bar{x}$ effect to var w larger

So standard error larger, final confidence interval larger.