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

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.

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 $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$.

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
Model	77.5444409	1	77.5444409
Residual	11565.0627	27,884	.414756231
Total	11642.6072	27,885	.417522223

Number of obs	=	27,886
F(1, 27884)	=	186.96
Prob > F	=	0.0000
R-squared	=	0.0067
Adj R-squared	=	0.0066
Root MSE	=	.64402

outp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
age	.0031338	.0002292	Omitted		.0026846	.003583
_cons	.4279898	.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 $\bar{X}$.

I. a) $\hat{\beta}_2 = \frac{\Sigma(x-\bar{x})(y_i-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{20.58}{211}$

$= 0.0975$

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

$2.8298 = \hat{\beta}_1 + 0.0975(21.5556)$

$0.7261 = \hat{\beta}_1$

$\beta_1 = 0.7261$: If x is equal to zero, Y will be 0.7261

$\beta_2 = 0.0975$: If x increases by 1 unit, Y will increase 0.0975 unit

b) $R^2 = \frac{ESS}{TSS} = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y_i-\bar{y})^2}$

$= \frac{0.5781}{2.5844}$

$= 0.7763$ The total variation of Y is explained by the regression 77.63%.

c) Point prediction : $\hat{Y} = 0.7261 + 0.0975(30)$

$= 3.6511$

Individual prediction : $Var(Y_0-\hat{Y}_0) = 6^2\left[1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\Sigma X_i^2}\right]$

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

$= 0.0503$

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

$\hat{Y}_0 - t_{\alpha/2}\cdot Se(Y_0-\hat{Y}_0) \leq \beta_1+\beta_2 X_i \leq \hat{Y}_0 + t_{\alpha/2}\cdot Se(Y_0-\hat{Y}_0)$

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

$3.1756 \leq \beta_1+\beta_2 X_i \leq 4.1266$

I'm 95% confident that when X_i is 30, y will be on (3.1756, 4.1266)

d) $Var(\hat{u}) = 6^2 = \frac{\Sigma\hat{u}^2}{n-2}$

$= \frac{0.5781}{16}$

$= 0.0361$

$Var(\hat{\beta}_1) = \frac{\Sigma X_i^2}{n\Sigma x_i^2}$

$= \frac{\Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2}\cdot 6^2$

$= \frac{9{,}620}{18(211)}\cdot 0.0361$

$= 0.0914$

$Var(\hat{\beta}_2) = \frac{6^2}{\Sigma x_i^2}$

$= \frac{6^2}{\Sigma(X_i-\bar{X})^2}$

$= \frac{0.0361}{211}$

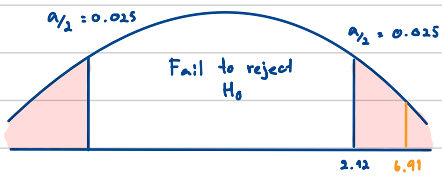
$= 0.0002$

e) $\hat{\beta}_2 - t_{\alpha/2}\, Se\,\hat{\beta}_2 \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\, Se\,\hat{\beta}_2$

$0.0975 - 1.746(\sqrt{0.0002}) \leq \beta_2 \leq 0.0975 + 1.746(\sqrt{0.0002})$

$0.0728 \leq \beta_2 \leq 0.1222$

I'm 90% confident that β_2 is on (0.0728, 0.1222) f)



$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$a = 0.05$

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

$$= \frac{0.0975 - 0}{\sqrt{0.0141}}$$

$$= 6.9149$$

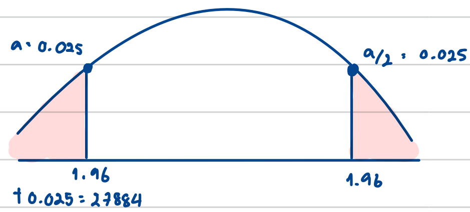
∴ tcal falls in "reject H0" region, in which, I'm 95% confident that the slope is different from zero.

II.

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$a = 0.05$



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

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

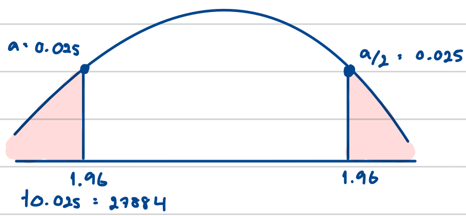
$$= 30.4969$$

∴ tcal falls in rejected region, β_1 is different from zero at 95% confidence level

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$a = 0.05$



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

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

$$= 13.6728$$

∴ tcal falls in rejected region, β_2 is different from zero at 95% confidence level

b) $\hat{\beta}_2 = 0.0031338$

If an age of a person increases by 1 year, time to visit hospital will increase 0.0031338 time.

A person becoming older will need more health care services. So people will go to the hospital more often.

This means that more demand and health care expenses, both goods and services, are expected to rise.

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

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

β_1 : If age is equal to zero, time to visit hospital will be $\ln outp_i$

β_2 : If age increases for 1 year, time to visit hospital increases for $outp_i$ $100\hat{\beta}_2$ %.

d) w_1 = change in Y

w_2 = change in X

$w_1 = 1$

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

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

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

$\hat{\beta}_2$ changes $10\hat{\beta}_2$

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

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

$\hat{\beta}_1$ doesn't change

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

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

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

$Se\,\hat{\beta}_1 = Se\,\hat{\beta}_1$

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

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

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

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

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

$\therefore$ Standard error of $\hat{\beta}_2$ changes to $10 Se\hat{\beta}_2$

Confident interval of β_1 $\quad \hat{\beta}_1 - t_{a/2} Se\hat{\beta}_1 < \beta_1 < \hat{\beta}_1 + t_{a/2} Se\hat{\beta}_1$

$\hat{\beta}_1$ and $Se\hat{\beta}_1$ doesn't change, so, confidence interval of β_1 doesn't change

β_2 $\quad \hat{\beta}_2 - t_{a/2} Se\hat{\beta}_2 < \beta_2 < \hat{\beta}_2 + t_{a/2} Se\hat{\beta}_2$

$\hat{\beta}_2$ and $Se\hat{\beta}_2$ changes, so, confidence interval changes $10\hat{\beta}_2$ $10Se\hat{\beta}_2$ / $10\hat{\beta}_2$ $10Se\hat{\beta}_2$

e) $\hat{Y} - t_{a/2}\sqrt{Var\,\hat{Y}_0} \leq E(Y \mid X = 50) \leq \hat{Y} + t_{a/2}\sqrt{Var\,\hat{Y}_0}$

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

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

$X_i = 50$ $\quad \hat{Y} = 0.4279898 + 0.0031338\,(50)$

$= 0.5846798$

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

$a = 0.01$ $\quad 0.5731596 \leq E(Y \mid X = 50) \leq 0.5962$

$a/2 = 0.005$

A 50-year-old person has an average time to visit hospital on (0.5731596, 0.5962)

III.

$$Var(y_0 - \hat{y}_0) = E(y_0 - \hat{y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2} \right]$$

$$Var(\hat{y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2} \right]$$

from variance of x_0 being far from $\bar{x}$, the effect to variance larger so standard error will be larger with larger confidence interval.