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

Instructions

Rachata Anuwanawong 6304641266

(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 $\operatorname{var}(u_i)$, $\operatorname{var}\left(\hat{\beta}_1\right)$ and $\operatorname{var}\left(\hat{\beta}_2\right)$.

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.

mean Y   mean X

$$\hat\beta_1 = \frac{\sum_{i=1}^n Y_i}{n} - \frac{\sum_{i=1}^n \hat\beta_2 X_i}{n}$$

$$\hat\beta_1 = \bar Y - \hat\beta_2 \bar X$$

$$Y_i = \hat\beta_1 + \hat\beta_2 X_i + \hat u_i$$
$$\hat u_i = Y_i - (\hat\beta_1 + \hat\beta_2 X_i)$$

Set objective function

$$\min_{\hat\beta_1, \hat\beta_2} \sum \hat u_i^2 = \sum [Y_i - \hat\beta_1 - \hat\beta_2 X_i]^2$$

solve for $\hat\beta_1$ F.O.N.C.

$$\frac{\partial \sum \hat u_i^2}{\partial \hat\beta_1^2} = \frac{\partial}{\partial \hat\beta_1}\left[\sum (Y_i - \hat\beta_1 - \hat\beta_2 X_i)^2\right] = 0$$

$$\frac{\partial -\hat\beta_1}{\partial \hat\beta_1} = -1$$

$$= -2\sum (Y_i - \hat\beta_1 - \hat\beta_2 X_i) = 0$$

$-n\hat\beta_1$

$$\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat\beta_1 - \sum_{i=1}^n \hat\beta_2 X_i = 0$$

$$\sum_{i=1}^n Y_i - n\hat\beta_1 - \sum_{i=1}^n \hat\beta_2 X_i = 0$$

$$\sum_{i=1}^n Y_i - \hat\beta_2 \sum_{i=1}^n X_i = n\hat\beta_1$$

$n = 18$

$\sum_{i=1}^n X_i = 388$

$\sum_{i=1}^n Y_i = 50.9$

$$50.9 - \hat\beta_2(388) = 18\hat\beta_1$$

$$\therefore \hat\beta_1 = 2.8278 - 21.5556\hat\beta_2 \qquad \hat\beta_2 = 0.0975$$

$$\hat\beta_1 = 2.8278 - 21.5556(0.0975)$$
$$= 0.7261$$

Ans. $\hat\beta_1 = 0.7261$ (interception at $X_i = 0$)

$\hat\beta_2 = 0.0975$ (slope as X_i increase by 1, Y_i increase by 0.0975)

solve for $\hat\beta_2$

F.O.N.C. $$\frac{\partial \sum \hat u_i^2}{\partial \hat\beta_2^2} = \frac{\partial}{\partial \hat\beta_2}\left[\sum (Y_i - \hat\beta_1 - \hat\beta_2 X_i)^2\right] = 0$$

$$\frac{\partial -\hat\beta_2 X_i}{\partial \hat\beta_2} = -X_i$$

$$= -2\sum X_i (Y_i - \hat\beta_1 - \hat\beta_2 X_i) = 0$$
$$= \sum_{i=1}^n X_i (Y_i - \hat\beta_1 - \hat\beta_2 X_i) = 0$$
$$= \sum_{i=1}^n X_i (Y_i - [\bar Y - \hat\beta_2 \bar X] - \hat\beta_2 X_i) = 0$$
$$= \sum_{i=1}^n X_i (Y_i - \bar Y + \hat\beta_2 \bar X - \hat\beta_2 X_i) = 0$$
$$= \sum_{i=1}^n X_i (Y_i - \bar Y - \hat\beta_2 (X_i - \bar X)) = 0$$
$$= \sum_{i=1}^n X_i (Y_i - \bar Y) - \hat\beta_2 \sum_{i=1}^n X_i (X_i - \bar X) = 0$$
$$\sum_{i=1}^n X_i (Y_i - \bar Y) = \hat\beta_2 \sum_{i=1}^n X_i (X_i - \bar X)$$

$$\therefore \hat\beta_2 = \frac{\sum_{i=1}^n X_i (Y_i - \bar Y)}{\sum_{i=1}^n X_i (X_i - \bar X)}$$

$$\sum_{i=1}^n (X_i - \bar X)(Y_i - \bar Y) = 20.58$$
$$\sum_{i=1}^n (X_i - \bar X) = 211$$

$$\hat\beta_2 = \frac{\sum_{i=1}^n (X_i - \bar X)(Y_i - \bar Y)}{\sum_{i=1}^n (X_i - \bar X)(X_i - \bar X)} = \frac{20.58}{211}$$

$$\hat\beta_2 = 0.0975$$

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

$$r^2 = \frac{ESS}{TSS}$$

TSS = ESS + RSS

$$\sum (Y_i - \bar Y)^2 = \sum (\hat Y_i - \bar Y)^2 + \sum (Y_i - \hat Y_i)^2$$

$$\frac{\sum (Y_i - \bar Y)^2}{\sum (Y_i - \bar Y)^2} = \frac{\sum (\hat Y_i - \bar Y)^2}{\sum (Y_i - \bar Y)^2} + \frac{\sum (Y_i - \hat Y_i)^2}{\sum (Y_i - \bar Y)^2}$$

$$1 = r^2 + \frac{\sum \hat u_i^2}{\sum_{i=1}^n (Y_i - \bar Y)^2}$$

$$1 = r^2 + \frac{0.5781}{2.5844}$$

$$r^2 = 1 - 0.2237$$

$$\therefore r^2 = 0.7763$$

$\sum (Y_i - \hat Y)^2 = \sum \hat u_i^2$

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

$\sum_{i=1}^n \hat u_i^2 = 0.5781$

r^2 is a measurement between the regression line and $\bar Y$

which $0 < r^2 < 1$ as r^2 closer to 1 → the data fits the line well

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

$\hat\beta_1 = 0.7261$

$\hat\beta_2 = 0.0975$

$\hat Y_i = \hat\beta_1 + \hat\beta_2 X_i$

$\hat Y_i = 0.7261 + 0.0975 X_i$

at $X_i = 30$, $\hat Y_i = 0.7261 + 0.0975(30)$

$\hat Y_i = 0.7261 + 2.925$

$\hat Y_i = 3.6511$

At $X_i = 30$, the data on regression line is 3.6511 but the actual data is nearly 3.6511 (not exact number)

d) Calculate the estimators of $\text{var}(u_i)$, $\text{var}(\hat\beta_1)$ and $\text{var}(\hat\beta_2)$.

$$var(u_i) = \hat\sigma^2 = \frac{\sum u_i^2}{n-k} = \frac{0.5781}{18-2} = 0.0361$$

$\sum_{i=1}^n \hat u_i^2 = 0.5781$, $n = 18$

$$var(\hat\beta_1) = \hat\sigma^2_{\hat\beta_1} = \frac{\sum X_i^2}{n\sum (X_i - \bar X)^2} \cdot \hat\sigma^2 = \frac{9{,}620}{18(211)} \cdot 0.0361 = 0.0914$$

$\sum_{i=1}^n X_i^2 = 9{,}620$, $\sum_{i=1}^n (X_i - \bar X) = 211$

$$var(\hat\beta_2) = \hat\sigma^2_{\hat\beta_2} = \frac{\hat\sigma^2}{\sum (X_i - \bar X)^2} = \frac{0.0361}{211} = 0.00017109$$

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

$1-\alpha = 0.9 \rightarrow \alpha = 0.1$, degree of freedom = $n-k = 18-2 = 16$, $se_{\hat{\beta}_2} = \sqrt{var(\hat{\beta}_2)} = \sqrt{0.00017109} = 0.0131$

$\therefore t_{16,0.1/2} = t_{16,0.05} = 1.746$, $\hat{\beta}_2 = 0.0975$

$$P[\hat{\beta}_2 - t_{\alpha/2}(se_{\hat{\beta}_2}) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}(se_{\hat{\beta}_2})] = 1-\alpha$$

$$P[0.0975 - (1.746)(0.0131) \leq \beta_2 \leq 0.0975 + (1.746)(0.0131)] = 0.9$$

$$P[0.0975 - 0.0229 \leq \beta_2 \leq 0.0975 + 0.0229] = 0.9$$

$$P[0.0746 \leq \beta_2 \leq 0.1204] = 0.9$$

At the 90% confidence when we calculate β_2 the answer of β_2 will between range of [0.0746, 0.1204]

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

two-tails test

$\hat{\beta}_2 = 0.0975$

$se_{\hat{\beta}_2} = 0.0131$

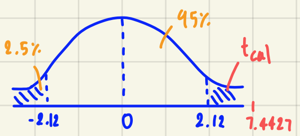
① $H_0 \rightarrow \beta_2 = 0$

$H_1 \rightarrow \beta_2 \neq 0$

② $\alpha = 0.05$

③ Calculate $t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} = \frac{0.0975 - 0}{0.0131} = 7.4427$

④ $t_{16,\alpha/2} = t_{16,0.025} = 2.12$



⑤ Conclusion : Since t_{cal} is on the rejection region, we can reject the null-hypothesis (H_0) → conslusion is H_1

So, at the 95% confidence , the β_2 is not zero

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

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

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	df. 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]
$\hat{\beta}_2$ age	.0031338	$SE_{\hat{\beta}_2}$ .0002292	Omitted		.0026846	.003583
$\hat{\beta}_1$ _cons	.4279898	$SE_{\hat{\beta}_1}$ .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 $\alpha = 0.05$.

Test B_1

$\hat{B}_1 = 0.4279898$ $\quad se_{\hat{B}_1} = 0.140339$

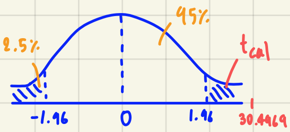
① $H_0 \rightarrow B_1 = 0$

$H_1 \rightarrow B_1 \neq 0$

② $\alpha = 0.05$

③ $t_{cal} = \dfrac{\hat{B}_1 - B_1}{se_{\hat{B}_1}} = \dfrac{0.0427989 - 0}{0.140339} = 30.4969$

④ $t_{27884, 0.025} = 1.96$



⑤ Conclusion Since t_{cal} is on the rejection region, we can reject the null-hypothesis $(H_0) \rightarrow$ conslusion is H_1

So, at the 95% confidence, person i visit the hospital is not zero

Test B_1

$\hat{B}_2 = 0.0031338$ $\quad se_{\hat{B}_2} = 0.0002242$

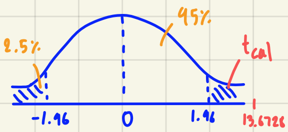
① $H_0 \rightarrow B_2 = 0$

$H_1 \rightarrow B_2 \neq 0$

② $\alpha = 0.05$

③ $t_{cal} = \dfrac{\hat{B}_2 - B_2}{se_{\hat{B}_2}} = \dfrac{0.0031338 - 0}{0.0002242} = 13.6728$

④ $t_{27884, 0.025} = 1.96$



⑤ Conclusion Since t_{cal} is on the rejection region, we can reject the null-hypothesis $(H_0) \rightarrow$ conslusion is H_1

So, at the 95% confidence, age of person i is not zero

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

In mathematic B_2 is slope of function. In this case B_2 interpret age of people who visit the hospital. Therefore B_2 is always positive sign due to age can't be negative. Generally as age increase, people will visit hospital more and more which very make sense.

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.

Log- lin Model

Let say $outp_i = Y_i$ and $age_i = X_i$

$\ln(Y_i) = \hat{B}_1 + \hat{B}_2 X_i$

$\dfrac{d \ln(Y_i)}{dX_i} = 0 + \hat{B}_2$

$\dfrac{dY/Y}{dX_i} = \hat{B}_2$

$\therefore \text{slope} = \dfrac{dY}{dX_i} = Y\hat{B}_2 \xrightarrow{\text{elasticity}} \dfrac{dY}{dX_i}\dfrac{X_i}{Y} = \cancel{Y}\hat{B}_2\dfrac{X_i}{\cancel{Y}} = \hat{B}_2 X_i$

Elasticity concept

As age of person i (X_i) increase by 1%, visiting times (Y_i) will increase by $\hat{B}_2$ %.

X_i

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).

- Constant $\hat{B}_1$: There were not change because of Y_i doesn't change which is at $X = 0$ the intercept is the same.
- Slope $\hat{B}_2$: There will be change for $\hat{B}_2$, $se_{\hat{B}_2}$, and $CI_{\hat{B}_2}$ because of the change in variable X_i divided by 10 which means the scale on x-axis will smaller than before. So the slope will steeper due to X_i decreased.

old data $\xrightarrow{\times 10}$ new data

new $\rightarrow$ $\hat{B}_2 = 0.031338$, $se_{\hat{B}_2} = 0.002292$, and $CI: 0.026946 \leq B_2 \leq 0.03583$

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

$se_{\hat{Y}_0} = \sqrt{var(\hat{Y}_0)} = \sqrt{0.00002} = 0.0044$

$\hat{\beta}_2 = 0.0031338$

$\hat{\beta}_1 = 0.4279898$

$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

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

Let $X_i = 50 \rightarrow \hat{Y}_{50} = 0.4279898 + 0.15669$

$= 0.5846798$

$P[\hat{Y}_{50} - t_{27884,0.005} \cdot se_{\hat{Y}_0} \leq Y_{50} \leq \hat{Y}_{50} + t_{27884,0.005} \cdot se_{\hat{Y}_0}] = 1 - 0.01$

$P[0.5846798 - 2.576 \cdot 0.0044 \leq Y_{50} \leq 0.5846798 + 2.576 \cdot 0.0044] = 0.99$

$P[0.5846798 - 0.0113344 \leq Y_{50} \leq 0.5846798 + 0.0113344] = 0.99$

$P[0.5733 \leq Y_{50} \leq 0.5960]$

Ans. Confident interval is on $[0.5733, 0.5960]$

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 $\bar{X}$.

$se = \sqrt{var(\_\_)}$

Mean prediction

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

$$\Pr\left[\hat{Y}_0 - (t_{\frac{\alpha}{2}} \cdot \mathrm{se}_{\hat{Y}_0}) \leq Y_0 \leq \hat{Y}_0 + (t_{\frac{\alpha}{2}} \cdot \mathrm{se}_{\hat{Y}_0})\right] = 1 - \alpha$$

Individual prediction

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

$$\Pr\left[\hat{Y}_0 - (t_{\frac{\alpha}{2}} \cdot \mathrm{se}_{fe}) \leq Y_0 \leq \hat{Y}_0 + (t_{\frac{\alpha}{2}} \cdot \mathrm{se}_{fe})\right] = 1 - \alpha$$

As we can see from the variance equation which calculated from the $X_0 - \bar{X}$.
When X_0 increases that makes the gap between $X_0 - \bar{X}$ was larger, the result is variance increased.
Due to the variance increased associated with the standard error increased because it calculated from the variance, as variance increase resulting to standard error increased.
Lastly the standard error associated with the confident interval boundary of both upper and lower bound, as standard error increased the gap of boundaries will larger at the same confident interval.