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.

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

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. ทำไมค. หลาย make sense

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 $\bar{X}$.

Sirephessorn Noochoochai
63046414 15

Question 1. (15 points) Given this information

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

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

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

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

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

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

$$\hat{\beta}_2 = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = \frac{20.58}{211} = 0.0975 \ \#$$

If X_i change by 1 unit, Y_i will change 0.0975 unit in the same direction (because $\hat{\beta}_2$ is positive)

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \frac{\sum_{i=1}^{n} Y_i}{n} - \hat{\beta}_2 \frac{\sum_{i=1}^{n} X_i}{n} = \frac{50.9}{18} - (0.0975)\left(\frac{388}{18}\right) = 2.8278 - 2.1017 = 0.7261 \ \#$$

$\hat{\beta}_1$ represents the intercept, therefore, If X_i is zero, Y will equel to 0.7261 unit.

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

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

$$= 1 - \frac{\Sigma(Y_i - \hat{Y})^2}{\Sigma(Y_i - \bar{Y})^2}$$

$$= 1 - \frac{0.5781}{2.5844}$$

$$= 1 - 0.2237$$

$$= 0.7763 \#$$

r^2 is a measurement of goodness of fit. It means that the value of r^2 (which is 0.7763) is closer to 1. So a SRF line can fit very well with the data point.

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

The regression model : $\hat{Y}_i = 0.7261 + 0.0975X_i$

if $X_i = 30$; $\hat{Y}_i = 0.7261 + 0.0975(30)$

$$= 0.7261 + 2.925$$

$$= 3.6511 \#$$

If X_i equal to 30, the prediction of Y_i is 3.6511.

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

$$\text{var}(u_i) = \frac{\Sigma \hat{u}_i^2}{n-k} = \frac{RSS}{n-k} = \frac{0.5781}{18-2} = 0.0361 \# \leftarrow \hat{\sigma}^2$$

$$\text{var}(\hat{\beta}_1) = \frac{\Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2}\hat{\sigma}^2 = \frac{9620(0.0361)}{18(211)} = 0.0914 \#$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\Sigma(X_i - \bar{X})^2} = \frac{0.0361}{211} = 0.0002 \#$$

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

$\alpha = 0.1$, $n-k = 18-2 = 16$, $\hat{\beta}_2 = 0.0975$

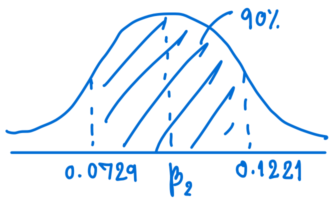
$t_{\frac{\alpha}{2}} = t_{\frac{0.1}{2}(16)} = t_{0.05(16)} = 1.746$

$se_{\hat{\beta}_2} = \sqrt{var(\hat{\beta}_2)} = \sqrt{0.0002} = 0.0141$

upper limit : $\hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_2} = 0.0975 + [(1.746)(0.0141)] = 0.1221$

lower limit : $\hat{\beta}_2 - t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_2} = 0.0975 - [(1.746)(0.0141)] = 0.0729$

$P[0.0729 \le \beta_2 \le 0.1221] = 0.9$



∴ From this graph, it means that β_2 which is the true value have value fell between 0.0729 and 0.1221

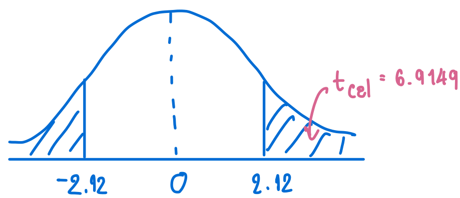
f) Test the hypothesis whether the slope coefficients are different from zero at 0.05 level of significance.

$H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

$\alpha = 0.05$

$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}} = \frac{0.0975 - 0}{0.0141} = 6.9149$



upper bound : $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}(16)} = t_{0.025(16)} = 2.120$

lower bound . $t_{\frac{\alpha}{2}} = -2.120$

∴ We can reject the null hypothesis, 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.

$n-k = 27{,}884$

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]
$\hat{\beta}_2$	age	.0031338	.0002292	Omitted		.0026846	.003583
$\hat{\beta}_1$	_cons	.4279898	.0140339			.4004828	.4554969

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

Test β_1

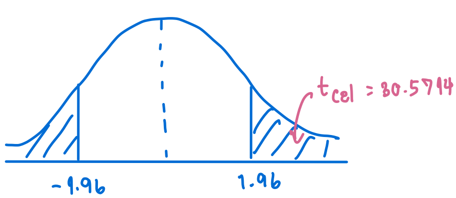
$H_0 : \beta_1 = 0$

$H_a : \beta_1 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_1 - \beta_1}{se\,\hat{\beta}_1} = \frac{0.4280 - 0}{0.0140} = 30.5714$$

upper bound : $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}(27{,}884)} = 1.960$

lower bound . $t_{\frac{\alpha}{2}} = -1.960$



∴ we can reject H_0, intercept parameter is not zero at 0.05 level of significance.

Test β_2

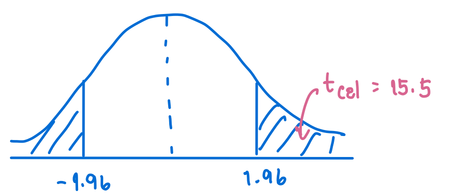
$H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = \frac{0.0031 - 0}{0.0002} = 15.5$$

upper bound : $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}(27{,}884)} = 1.960$

lower bound . $t_{\frac{\alpha}{2}} = -1.960$



∴ we can reject H_0, slope parameter is not zero at 0.05 level of significance.

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

$\hat{\beta}_2 = 0.0031$ is significantly different from zero at $\alpha = 0.05$

$\hat{\beta}_2$ represents the slope in equation, it means if age change by 1 year, the number of time visited hospital change by 0.0031 times in the same direction

Therefore, the sign of $\hat{\beta}_2$ make economic sense because when people get older, we tend to visit hospital more due to health problems.

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.

$$\ln \widehat{outp}_i = 0.4279 + 0.0031\, age_i$$

$$\frac{d \ln \widehat{outp}_i}{d\, age_i} = 0.0031$$

If age increase by 1 unit (1 year), it is associated with the number of time visited hospital increase of 0.31%

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).

$w_1 = 1, \quad w_2 = \frac{1}{10}$

$\hat{\beta}_1^* = w_1 \hat{\beta}_1 = (1)(0.4280) = 0.4280$

$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right) \hat{\beta}_2 = \left(\frac{1}{1/10}\right)(0.0031) = 0.031$

$se(\hat{\beta}_1^*) = w_1 se_{\hat{\beta}_1} = (1)(0.0140) = 0.0140$

$se(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right) se_{\hat{\beta}_2} = \left(\frac{1}{1/10}\right)(0.0002) = 0.002$

CI of $\hat{\beta}_1$ at $\alpha = 0.05$

upper limit : $\hat{\beta}_1 + t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_1} = 0.4280 + [(1.96)(0.014)] = 0.4554$

lower limit : $\hat{\beta}_1 - t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_1} = 0.4280 - [(1.96)(0.014)] = 0.4005$

$0.4005 \leq \hat{\beta}_1^* \leq 0.4554$

CI of $\hat{\beta}_2$ at $\alpha = 0.05$

upper limit : $\hat{\beta}_2 + t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_2} = 0.031 + \left[(1.96)(0.002)\right] = 0.0349$

lower limit : $\hat{\beta}_2 - t_{\frac{\alpha}{2}} \cdot se_{\hat{\beta}_2} = 0.031 - \left[(1.96)(0.002)\right] = 0.0271$

$0.0271 \leq \hat{\beta}_2^* \leq 0.0349$

$\therefore$ $\hat{\beta}_1$, $se_{\hat{\beta}_1}$ and CI of $\hat{\beta}_1$ remain the same but $\hat{\beta}_2$, $se_{\hat{\beta}_2}$ and CI of $\hat{\beta}_2$ change by $\hat{\beta}_2^*$ and $se^*_{\hat{\beta}_2}$ are bigger and CI of $\hat{\beta}_2^*$ is narrower.

- when age is zero, the intercept is 0.4280 times.
- when age increses by 1 year, the number of time visited hospital increases by 0.031 times.

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

$x_0 = age_0 = 50$ $\quad\quad t_{\frac{\alpha}{2}} = 2.576$

$\widehat{outp}_i = 0.4280 + 0.0031\, age_i$

$\widehat{outp}_0 = 0.4280 + 0.0031(50) = 0.583$

$se_{\widehat{outp}_0} = \sqrt{var(\widehat{outp}_0)} = \sqrt{0.00002} = 0.0044$

$Pr\left[0.583 - (2.576 \cdot 0.0044) \leq \widehat{outp}_0 \leq 0.583 + (2.576 \cdot 0.0044)\right] = 1 - 0.01$

$Pr\left[0.5717 \leq \widehat{outp}_0 \leq 0.5943\right] = 0.99$ #

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 $\bar{X}$.

The confidence interval for the mean prediction and individual prediction get larger as the x_0 is further away from $\bar{x}$ either to left or right due to the variance that is larger from formular that we look at. Moreover, the variance from mean prediction is narrower than from individual prediction. So if we want to find larger confident interval, we can use variance of forecasting error.