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.

EE325 Introductory Econometrics **Section 2** Weerawat Phattarasukkumjorn
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

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

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

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

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

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

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

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 20.58 \qquad \sum_{i=1}^{n} \hat{u}_i^{\,2} = 0.5781$$ 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) $X = \frac{\sum X_i}{n} = \frac{388}{18} , 21.5556$ $\qquad \hat{Y} = \frac{50.9}{18} , 2.8278$

$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{20.58}{211} , 0.0975$

$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$; $2.8278 = \hat{\beta}_1 + 0.0975(21.5556)$

$\hat{\beta}_1 = 2.8278 - 2.1017 , 0.7261$

$\hat{\beta}_1 = 0.7261$; if x equal to 0, y will equal to 0.7261.

$\hat{\beta}_2 = 0.0975$; if x increase 1 unit, y increase 0.0975 unit.

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

$= 1 - \frac{0.5781}{2.5844} , 0.7763$

∴ the total variation in Y explain by the regression 77.63%.

c) $\hat{Y} = 0.7261 + 0.0975(X)$

$= 0.7261 + 0.0975(30)$

$= 3.6511$

individual estimate

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

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

$var(Y_0 - \hat{Y}) = 0.0361\left[1 + \frac{1}{18} + \frac{(30 - 21.5556)^2}{211}\right]$

$= 0.0503$

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

$\hat{Y} - t_{\frac{\alpha}{2}}[se(Y_0 - \hat{Y})] < \beta_1 + \beta_2 X_i < \hat{Y} + t_{\frac{\alpha}{2}}[se(Y_0 - \hat{Y})]$

$3.6511 - 2.12(\sqrt{0.0503}) \leqslant \beta_1 + \beta_2 X_i \leqslant 3.6511 + 2.12(\sqrt{0.0503})$

$3.1756 \leqslant Y_0 | X_0 = 30 \leqslant 4.1266$

we 95% confidence that $X_i = 30$, Y is on $[3.1756, 4.1266]$

d) $var(u) = \sigma^2, \frac{\sum \hat{u}^2}{n-2} = \frac{0.5781}{16}, \quad 0.0361$

$var(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2 = \frac{9620}{18(211)} \times 0.0361 = 0.0914$

$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{0.0361}{211} = 0.0002$

e) $\hat{\beta}_2 - t_{\frac{\alpha}{2}}(se\,\hat{\beta}_2) \leqslant \beta_2 \leqslant \hat{\beta}_2 + t_{\frac{\alpha}{2}}(se\,\hat{\beta}_2)$

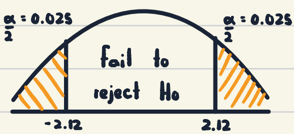
$0.0975 - 1.746(\sqrt{0.0002}) \leqslant \beta_2 \leqslant 0.0975 + 1.746(\sqrt{0.0002})$

$0.0728 \leqslant \beta_2 \leqslant 0.1222$

$\therefore$ we have 90% confidence, β_2 is on $[0.0728, 0.1212]$

f) $H_0 : \beta_2 = 0 \qquad \alpha = 0.05$

$H_1 : \beta_2 \neq 0$



$t\ cal = \frac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = \frac{0.0975 - 0}{\sqrt{0.0002}}$

$= 6.8943$

$\therefore$ t cal fall in reject Ho region, we 95% confidence that the slope is different from zero.

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

Anova

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

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

k=2 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.

k = number of variable

Source	SS	df	MS
			ESS/k-1
Model	ESS 77.5444409	k-1 1	77.5444409
Residual	RSS 11565.0627	27,884	.414756231
		n-k	RSS/n-k
Total	TSS 11642.6072	27,885	.417522223
		n-1	TSS/n-1

Number of obs	=	27,886
F(1, 27884)	=	186.96 MSE/MSS
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]
		se $\hat{\beta}_2$				
age	$\hat{\beta}_2$ .0031338	.0002292	Omitted		.0026846	.003583
_cons	$\hat{\beta}_1$ .4279898	.0140339			.4004828	.4554969
		se $\hat{\beta}_1$				

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 $\bar{X}$.

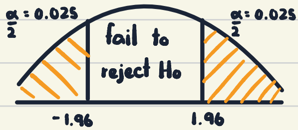
----

a) $H_0 : \beta_1 = 0$ $\alpha = 0.05$

$H_1 : \beta_1 \neq 0$

$$t\ cal = \frac{\hat{\beta}_1 - \beta_1}{se\ \hat{\beta}_1}$$

$$= \frac{.4279898 - 0}{.0140339}$$



$= 30.4969$

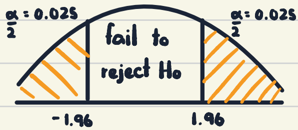
$\therefore$ t cal fall in reject Ho, β_1 different from 0 at $\alpha = 0.05$

$H_0 : \beta_2 = 0$ $\alpha = 0.05$

$H_1 : \beta_2 \neq 0$

$$t\ cal = \frac{\hat{\beta}_2 - \beta_2}{se\ \hat{\beta}_2}$$

$$= \frac{.0031338 - 0}{.0002292}$$



$= 13.6728$

$\therefore$ t cal fall in reject Ho region, we have 95 % confidence β_2 different from 0.

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

I believed that it make sense because when you get older, the demand of hospital service from the more sensitive physique.

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

β_1 : If age of person increase 1 year, time to visit hospital equal to $\ln outp_i$ time.

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

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2$$
$$\hat{\beta}_1^* = w_1\hat{\beta}_1$$
$$\hat{\sigma}^{*2} = w_1^2\hat{\sigma}^2$$
$$\text{var}(\hat{\beta}_1^*) = w_1^2\,\text{var}(\hat{\beta}_1)$$
$$\text{var}(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2\text{var}(\hat{\beta}_2)$$
not change
$$r_{xy}^2 = r_{x^*y^*}^2$$

d) w_1 : change in Y $= 1$

w_2 : change in X $= 1/10$

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

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

$$\text{var}\,\hat{\beta}_1^* = w_1^2\,\text{var}\,\hat{\beta}_1$$
$$= (1)^2\,\text{var}\,\hat{\beta}_1$$

$$\text{se}\,\hat{\beta}_1^* = \text{se}\,\hat{\beta}_1$$

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

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

$\hat{\beta}_1$ not change

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

$$\text{var}\,\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)^2\text{var}\,\hat{\beta}_2$$
$$= (10)^2\,\text{var}\,\hat{\beta}_2$$
$$\text{var}\,\hat{\beta}_2^* = 100\,\text{var}\,\hat{\beta}_2$$
$$\text{se}\,\hat{\beta}_2^* = 10\,\text{se}\,\hat{\beta}_2$$

standard error of $\hat{\beta}_2$ change to $10\,\text{se}\,\hat{\beta}_2$

C.I. β_1 : $\hat{\beta}_1 - t_{\frac{\alpha}{2}}(\text{se}\,\hat{\beta}_1) \leqslant \beta_1 \leqslant \hat{\beta}_1 + t_{\frac{\alpha}{2}}(\text{se}\,\hat{\beta}_1)$

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

C.I. β_2 : $\hat{\beta}_2 - t_{\frac{\alpha}{2}}(\text{se}\,\hat{\beta}_2) \leqslant \beta_2 \leqslant \hat{\beta}_2 + t_{\frac{\alpha}{2}}(\text{se}\,\hat{\beta}_2)$

Because $\hat{\beta}_1$ and se $\hat{\beta}_1$ are change so confidence interval will change to $10\hat{\beta}_2 - t_{\frac{\alpha}{2}}(10\,\text{se}\,\hat{\beta}_2) \leqslant \beta_2 \leqslant 10\hat{\beta}_2 + t_{\frac{\alpha}{2}}(10\,\text{se}\,\hat{\beta}_2)$.

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

$$\hat{Y} - t_{\frac{\alpha}{2}}\,\text{se}\,\hat{Y}_0 \leqslant E(Y|X=50) \leqslant \hat{Y} + t_{\frac{\alpha}{2}}\,\text{se}\,\hat{Y}_0$$

$$\hat{Y} = \beta_1 + \beta_2 X_i$$
$$= 0.4279898 + 0.0031338X_i$$

$\alpha = 0.01$

$X_i = 50$; $\hat{Y} = 0.4279898 + 0.0031338(50) = 0.5846798$

$\frac{\alpha}{2} = 0.005$, $t_{\frac{\alpha}{2}} = 2.576$

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

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

$\therefore$ Age of 50 year old, time to visit hospital average on $[0.5731596,\ 0.5962]$.

3. From variance of both, when X_0 far from $\bar{x}$ effect to larger variance, standard error is larger. Last, confident interval is larger.

Mean Prediction

$$\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2} se(\hat{Y}_0) \le \beta_1 + \beta_2 X_0 \le \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2} se(\hat{Y}_0)$$

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

Individual Prediction

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

$$\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2} se(\hat{Y}_0) \le \beta_1 + \beta_2 X_0 \le \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2} se(\hat{Y}_0)$$