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

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.

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

$$\hat{B}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{20.58}{211} = 0.0975 \text{ Ans}$$

$$\hat{B}_1 = \bar{Y} - \hat{B}_2 \bar{X}$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{50.90}{18} = 2.8278$$

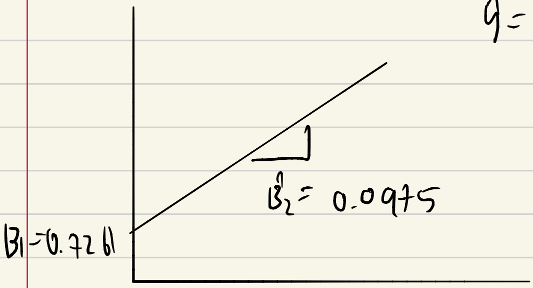
$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{388}{18} = 21.5556$$

$$= \hat{B}_1 = 2.8278 - 0.0975(21.5556) = 0.7261 \text{ Ans}$$

If $X_i = 0$, $\hat{Y}_i = 0.7261$

If X_i increase by 1 unit $\hat{Y}$ will increase by 0.0975 unit on average.

$\hat{y} = 0.7261 + 0.0975X$



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

fromula $= R^2 = 1 - \frac{RSS}{TSS}$, $TSS = \Sigma (y - \bar{y})^2$, $RSS = \Sigma (\hat{y} - y)^2$

$= 1 - \frac{0.5781}{2.5849} = 0.7727$ Ans

Meaning $R^2 = 0.7727$ therefore varible X can define or explain 77.27 percent of variation in Y

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 meani

formula

C). $\hat{Y} = \hat{B}_1 + \hat{B}_2 X_i = \hat{y} = 0.7261 + 0.0975X_i$

$= 0.7261 + 0.0975X_i$, plug $X_i = 30$

$= 0.7261 + 0.0975(30)$

$= 3.6511$ Ans

meaning If $X_i = 30$, $\hat{Y}$. value is expected to be 3.6511

d).

formula degree of freedom $k=2$

$\text{var}(u_i) = \frac{\sum \hat{u}_i^2}{n-k} = \frac{0.5741}{16-2} = 0.0361$

$\text{var}(\hat{B}_1) = \frac{\sum x_i^2}{n \sum x_i^2} \sigma^2 = \frac{388}{16(211)} \cdot 0.0361 = 0.0036$

$\text{var}(\hat{B}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{0.0361}{211} = 0.00017 \approx 0.0002$

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

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.

e) $t = \frac{1-90}{2} = 0.05$

formula

$se(\hat{B_2}) = \sqrt{var(\hat{B}_2)}$

$\hat{B}_2 \pm [t_{\alpha/2, n-2}] \cdot se(\hat{B}_2)$

$= \sqrt{0.0002} = 0.0141$

$T_{0.05} = 1.746$ for $\alpha = 0.1$

$Pr(0.0975 - (1.746 \cdot 0.0141) \leq B_2 \leq 0.0975 + (1.746 \cdot 0.0141)) = 0.90$ Any

Interval

$Pr(0.0767 \leq B_2 \leq 0.1183) = 0.90$

f)

① $H_0: B_2 = 0$ $H_a: B_2 \neq 0$

② $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

$se(\hat{B}_2) = 0.0141$

③ $t_{cal} = \frac{0.0975 - 0}{0.0141} = 6.9149$

④ t critical fall in reject Ho are, we reject Ho/

Reject null hypothesis, 95% $B_2 \neq 0$

reject Ho

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	Number of obs	=	27,886
Residual	11565.0627	27,884	.414756231	F(1, 27884)	=	186.96
Total	11642.6072	27,885	.417522223	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 $\bar{X}$.

----

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

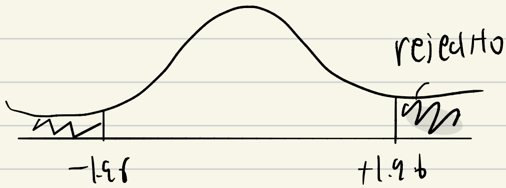
b) Interpret the meaning of $\hat{\beta}_2$. Does the sign of $\hat{\beta}_2$ make economic sense? Explain.

a) (1) $H_0: B_1 = 0$; $Ha: B_1 \neq 0$

(2) $\frac{\alpha}{2} = 0.05$, (3) $t_{test} = \frac{\hat{B} - B_1}{se(B_1)}$

(4) $t_{test} = \frac{0.0427996 - 0}{0.0140339} = 30.49685 \approx 30.4969$

t-critical $= 1.960$



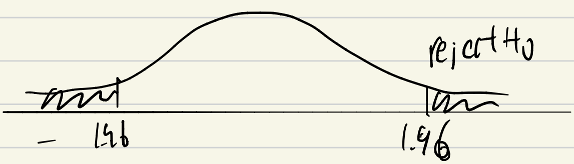
reject Ho , $B_1 \neq 0$ at 0.05

(1) $H: B_2 = 0$, $Ha: B_2 \neq 0$

(2) $\frac{\alpha}{2} = 0.025$

(3) $t_{test} = \hat{B}_2 - B_1$

(4) $= \frac{0.0011348 - 0}{0.0002292} = 13.6728$ t= critical $= 1.96$



$B_2 \neq 0$ at 0.05
reject Ho

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.

b) An average of visit per year increase by 0.0031 time. If people get older by one year.

As people are aging therefore we are likly to rely on medical care, thus this make economical sense.

$\widehat{out}_i = 0.4278 + 0.0031\, age_i$

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.

C

$$outp = 0.4278 + 0.0031\, age_i$$

$$= \ln outp = 0.4278 + 0.0031\, age.$$

Age increase by 1 year, so with outp will increase by

$= \text{coefficient} \times 100\%$

$= 0.0031 \times 100\% = 0.31\%$ Ans

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 $\alpha = 0.01$.

d. Ans. coefficient, SE and CI will scaled by 10.

New coefficient   SE   CI

= 0.031318   = 0.002292   .026846   .03583

Proof

$$\hat{\beta}_{2\,new} = \frac{\sum [X_{new} - \bar{X}_{new}][y - \bar{y}]}{\sum [X_{new} - \bar{X}]^2} \qquad X_{new} = \frac{X_i}{10}$$

$$= \frac{\sum [\frac{X}{\cancel{10}} - \frac{\bar{X}}{\cancel{10}}][y - \bar{y}]}{\sum [\frac{X}{10} - \frac{\bar{X}}{10}]^{\cancel{2}}}$$

$$= \frac{\sum (X - \bar{X})(y - \bar{y})}{[\frac{1}{10}] \sum (X - \bar{X})^2} \quad \text{— } B_2 \text{ old}$$

$$= \frac{0.0031}{0.1} = 0.031$$

same for other

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 $\alpha = 0.01$.

e. $\hat{Y}_0 = 0.128 + 0.0091(50) = 0.583$ Ans $\quad \frac{1}{2} = 0.05$

formula

$se(\hat{Y}_0) = \sqrt{var(\hat{Y}_0)} = 0.0045$ Ans

confidence interval

$\hat{Y} \pm t_{\frac{\alpha}{2}, n-k} \cdot se(\hat{Y})$

$= 0.583 \pm 2.576 \cdot 0.0045$

$= Pr(0.5714 \leq Y_0 \leq 0.5946) = 0.99$

Confidence interval of mean prediction of 50 years old

$= Pr(0.5714 \leq Y_0 \leq 0.5946)$ at 99% confidence level.

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 $\bar{X}$.

$$\hat{y} \pm t_{\frac{\alpha}{2}, n-k} \sqrt{\sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2}\right)}$$

$\bar{x}$ is central tendency of x_i and if $x_0 - \bar{x}$ in individual prediction get larger, also the standard error will get larger. because the information is getting more disperse and error may occured.