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.

Pimthnida
6304647084

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

$\bar{Y} = \frac{\Sigma_{i=1}^{n} Y_i}{n} = \frac{50.9}{18} = 2.8278$

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

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 2.8278 - 0.0975(21.5556) = 0.7261$

$\therefore$ The coefficient when $X = 0$, we expect that $\hat{Y} = 0.7261$
and if X_i increase by 1 unit, we can expect
that $\hat{Y}$ will increase on average by 0.0975

b) $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\Sigma_{i=1}^{n} \hat{u}_i^2}{\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{0.5781}{2.5849} = 0.7763$

$R^2 = 0.7763$ $\therefore$ X variable can explain 77.63% of variation in Y.

c) From the SRF, $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i = 0.7261 + 0.0975 X_i$
$X_i = 30$
$E(\hat{Y} \mid X_i = 30) = 0.7261 + 0.0975(30) = 3.6511$

d) $var(u_i) = \hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{n-K} = \frac{0.5781}{18-2} = 0.0361$

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

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

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

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

$Pr(0.0975 - (1.746 \cdot 0.0141) \leq \beta_2 \leq 0.0975 + (1.746 \cdot 0.0141)) = 1 - 0.1$

$Pr(0.0729 \leq \beta_2 \leq 0.1221) = 0.9$

Pimthida
6304641084

f) 1 $H_0 : \beta_2 = 0$ 2 $\alpha = 0.05$

$H_a : \beta_2 \neq 0$

3 From e, $se(\hat{\beta}_2) = 0.0141$

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

4 Upper bound : $t_{\alpha/2} = t_{0.025}(16) = 2.12$

Lower bound : $-t_{\alpha/2} = -t_{0.025}(16) = -2.12$

5 t_{cal} lied beyond CI $(6.9141 > 2.12)$

so, we can reject the null hypothesis

This shows that 95% of the time, 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	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 $\bar{X}$.

Pimtnida
6304641084

② a) ① $H_0: \beta_1 = 0$ ② $\alpha = 0.05$

$H_a: \beta_1 \neq 0$

③ $t_{cal} = \frac{0.4279898 - 0}{0.0140339} = 30.4969$

④ upper bound : $t_{\alpha/2} = t_{0.025}(27,884) = 1.96$

Lower bound : $-t_{\alpha/2} = -1.96$

⑤ t_{cal} > upper bound ($30.4969 > 1.96$) so, we can reject null hypothesis.

This means t + 95% of the time, β_1 is not zero.

① $H_0: \beta_2 = 0$ ② $\alpha = 0.05$

$H_a: \beta_2 \neq 0$

③ $t_{cal} = \frac{0.0031338 - 0}{0.0002292} = 13.6728$

④ upper bound : $t_{\alpha/2} = t_{0.025}(27,884) = 1.96$

Lower bound : $-t_{\alpha/2} = -1.96$

⑤ t_{cal} > upper bound ($13.6728 > 1.96$) so, we can reject null hypothesis

This means that 95% of the time, β_2 is not zero

∴ we can reject null hypothesis for both tests, so we can make sure that 95 percent, β_1, β_2 are not zero

b) when people age 1 more year, we can expect that visit per year will increase on average of 0.0031 times. ($\hat{\beta}_2 = 0.0031338$)

This positive sign makes economic sense because of people age, we tend to have to visit hospital.

c) If outp$_i$ is turned into natural logarithmic scale, or

$$\ln \widehat{outp}_i = \hat{\beta}_1 + \hat{\beta}_2 age_i$$

then, the interpretation of $\hat{\beta}_2$ becomes

if age$_i$ increases by 1 year, we expect that people's visiting hospital will increase by $\hat{\beta}_2 \times 100 = 0.3134\%$.

Pimthida
6304641084

d) The coefficient, standard errors and confidence interval will scale up by 10 so, coefficient => 0.031338

standard errors => 0.002292

confidence interval => 0.02846 and 0.03983

e) $\widehat{outp}_i = \beta_1 + \beta_2 age_i$

$\hat{Y}_0 = 0.428 + 0.0031(50) = 0.583$

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

$t_{0.005} = 2.576$

$Pr(0.583 - (2.576 \times 0.0045) \leqslant Y_0 \leqslant 0.583 + (2.576 \times 0.0045)) = 0.99$

$pr(0.5714 \leqslant Y_0 \leqslant 0.5946) = 0.99$

3 $var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$

From the equation, if $X_0 - \bar{X}$ gets larger, or X_0 is farther away from $\bar{X}$ the variance get larger leading to larger $se(\hat{Y}_0)$

In other words, $\bar{X}$ is the centrdl tendency of X_i, which means that the further away from $\bar{X}$, the less information of X_i is available.

The confidence interval must be larger to cope with more unknown.