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

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{20.58}{211} = 0.0975$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{50.90}{18} = 2.8278 \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{388}{18} = 21.5556$$

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

$\therefore$ when $x = 0$, $y = 0.7261$ and when x increase by 1 unit, y increase by 0.0978

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

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{0.5781}{2.5844} = 0.7763$$

R^2 mean x variable explain 77.63 percent of variation in y

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

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i$$
$$= 0.7261 + 0.0975(30)$$
$$= 3.6511$$

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

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

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum(X_i-\bar{X})^2}\hat{\sigma}^2 = \frac{9,620}{18(211)}(0.0361) = 0.0914$$

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

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

$\alpha = 0.1$, then $t_{0.05} = 1.746$

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

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

$$\Pr(0.0729 \le \beta_2 \le 0.1221) = 0.90$$

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

$H_0: \beta_2 = 0$

$H_a: \beta_2 \neq 0$

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

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

$\therefore$ We can reject the null hypothesis (H_0) since at 95 percent, $\beta_2 \neq 0$

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		
				Number of obs =	27,886
				F(1, 27884) =	186.96
Model	77.5444409	1	77.5444409	Prob > F =	0.0000
Residual	11565.0627	27,884	.414756231	R-squared =	0.0067
				Adj R-squared =	0.0066
Total	11642.6072	27,885	.417522223	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 $\alpha = 0.05$.

$H_0 : \beta_1 = 0 \qquad H_a : \beta_1 \neq 0$

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

$H_0 : \beta_2 = 0 \qquad H_a : \beta_2 \neq 0$

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

$\therefore$ We can reject null hypothesis of both test since at 95 percent, β_1 and $\beta_2 \neq 0$

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

When people age increase 1 year, the visit in hospital per year will increase 0.0031 times. The sign of $\hat{\beta}_2$ make economic sense because as people become older, they will prefer more medical service

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.

If $outp_i$ turned to natural logarithmic scale $\ln \widehat{outp_i} = \hat{\beta}_1 + \hat{\beta}_2 age$, as age increase, the number of time people visited the hospital will increase by $\hat{\beta}_2 \times 100 = 0.3134$ percent

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).

If age_i variable is divided by 10, the coefficients, standard error and confidence interval will scale up by 10 while others remain constant.

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

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

$se(\hat{Y}_0) = \sqrt{var(\hat{Y}_0)} = \sqrt{0.00002} = 0.0045$

$\alpha = 0.01$, then $t_{\frac{\alpha}{2}} = t_{0.05} = 2.576$

$Pr(0.583 - 2.576(0.0045) \leq Y_0 \leq 0.583 + 2.576(0.0045)) = 1 - 0.01$

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

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 $\bar{X}$.

We use $var(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (x_i - \bar{x})^2} \right]$ to calculated the confidence interval of mean prediction. As X_0 is further away from $\bar{X}$, the variance become larger. $\bar{x}$ is central tendency to x_i, so the further the $\bar{x}$, the less data of x_i.