

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 least 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

$$\begin{aligned}
 n &= 18 & \sum_{i=1}^n X_i &= 388.00 & \sum_{i=1}^n Y_i &= 50.90 \\
 \sum_{i=1}^n (X_i)^2 &= 9,620.00 & \sum_{i=1}^n X_i Y_i &= 1,254.90 \\
 \sum_{i=1}^n (X_i - \bar{X})^2 &= 211.00 & \sum_{i=1}^n (Y_i - \bar{Y})^2 &= 2.5844 \\
 \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= 20.58 & \sum_{i=1}^n \hat{u}_i^2 &= 0.5781
 \end{aligned}$$

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

- 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.
- Compute the value of R^2 and explain its meaning.
- If $X_i = 30$, estimate the value of \hat{Y}_i and explain its meaning.
- Calculate the estimators of $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$.
- What are the 90-percent confident intervals for β_2 ? Interpret the meaning.
- Test the hypothesis whether the slope coefficients are different from zero at 0.05 level of significance.

$$1. a) \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{20.58}{211.00} = 0.0975$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{50.90}{18} = 2.8278$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{388.00}{18} = 21.5556$$

$$\hat{\beta}_1 = 2.8278 - 0.0975(21.5556) = 0.7258$$

$\therefore y_i = 0.7258 + 0.0975 X_i$ meaning that the intercept is 0.7258 and the slope of this function is 0.0975

$$b) r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{0.5781}{2.5844} = 0.7763$$

r^2 is a measurement of goodness of fit of regression line compared to an estimator \bar{y} . In this case, 77.63% of y is explained by the regression model.

$$c) y_i = 0.7258 + 0.0975 X_i = 0.7258 + 0.0975(30)$$

$$y_i = 3.6508 \quad \therefore \text{When } x_i = 30 \text{ average of } y_i \text{ will be } 3.6508$$

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

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

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

$$e) 90\% \text{ Confidentinterval} \Rightarrow \alpha = 0.1 \quad n-k = 18-2 = 16$$

$$\rightarrow 1 - \alpha = 90\% \text{ confidentinterval}$$

$$P \left[\hat{\beta}_2 - \left(\frac{t_{\frac{\alpha}{2}}}{2} \cdot \text{se} \hat{\beta}_2 \right) \leq \beta_2 \leq \hat{\beta}_2 + \left(\frac{t_{\frac{\alpha}{2}}}{2} \cdot \text{se} \hat{\beta}_2 \right) \right] = 1 - \alpha$$

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

$$\rightarrow \frac{t_{\frac{\alpha}{2}}}{2} = 1.746$$

$$P \left[0.0975 - (1.746 \cdot 0.0141) \leq \beta_2 \leq 0.0975 + (1.746 \cdot 0.0141) \right] = 0.9$$

$$P \left[0.0729 \leq \beta_2 \leq 0.1221 \right] = 0.9$$

Which means that 90 out of 100 times, this confident interval will contain the true β_2

F)

step 1 : state a hypothesis

$H_0 : \beta_2 = 0$ - Null hypothesis

$H_a : \beta_2 \neq 0$ - Alternative hypothesis

step 2 : Pick an $\alpha = 0.05$

Step 3 : calculate test statistics

$$t_{\text{cal}} = \frac{\hat{\beta}_2 - \beta_2}{\text{se}\hat{\beta}_2} = \frac{0.0975 - 0}{0.0141} = 6.9145$$

step 4 : state the decision test

upper bound : $t_{\frac{\alpha}{2}} = 2.12$

lower bound : $t_{\frac{\alpha}{2}} = -2.12$

Step 5 : Conclude

$\therefore t_{\text{cal}}$ is beyond the boundary of test value CI, we can reject the null hypothesis, at the significance level of 95%.

Assignment 1

Assigned on Feb 8th, 2022. Due date Saturday, Feb 26th, 2022 before midnight.

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
Model	77.5444409	1	77.5444409	F(1, 27884)	=	186.96
Residual	11565.0627	27,884	.414756231	Prob > F	=	0.0000
Total	11642.6072	27,885	.417522223	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			.0026846 .003583
_cons	.4279898	.0140339			.4004828 .4554969

- Test if both parameters are significantly different from zero or not. Use $\alpha = 0.05$.
- Interpret the meaning of $\hat{\beta}_2$. Does the sign of $\hat{\beta}_2$ make economic sense? Explain.
- 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 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).
- 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} .

$$2. a) t = \frac{\hat{\beta}_2 - \beta_2}{S.E.}$$

$$= \frac{.0031 - 0}{.0002}$$

$$= 13.672$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{S.E.}$$

$$= \frac{.428 - 0}{.014}$$

$$= 30.4968$$

→ Critical t-value for $\alpha = 0.05$ and degree of freedom infinite is 1.96

$$- \beta_2 + \frac{t_{\alpha}}{2} \cdot se_{\beta} = .0031338 + 1.96(.0140339) = .0306$$

$$- \beta_2 + \frac{t_{\alpha}}{2} \cdot se_{\beta} = .0031338 - (1.96(.0140339)) = -.0322$$

∴ We can reject null hypothesis for both case which means that we can make sure that 95 percent interval for β_1 and β_2 are not zero

b) People will increase the visited time per year on average of 0.0031 times when they age increase 1 more year. This relation make a positive sense for economic because people tend to use more medical services when they old.

$$c) \ln \hat{out}_i = \hat{\beta}_1 + \hat{\beta}_2 \text{ age}_i$$

we can interpreted that if age increase by one year, we expect that people's visiting hospital will increase by 0.3134 percent

d) IF age_i variable is divided by 10

$$\begin{aligned} \text{coefficient} &= \frac{.003138}{10} = .031338 \\ \text{standard Error} &= \frac{.0002292}{10} = .002292 \\ \text{Confident interval} &= \frac{.0026846}{10} = .026846 \\ &= \frac{.003583}{10} = .03583 \end{aligned}$$

and other values are remain constant

$$E) (\text{var})_{\hat{y}_0} = 0.00002, \alpha = 0.01$$

$$- \hat{y} = \beta_1 + \beta_2 (X_i) = 0.428 + 0.0031 (50) = 0.583$$

$$- se(\hat{y}_0) = \sqrt{(\text{var})_{\hat{y}_0}} = \sqrt{0.0002} = 0.0045$$

$$- P[0.583 - (2.576 \times .0045) \leq y_0 \leq 0.583 + (2.576 \times .0045)] = 1 - .01$$

$$P(0.5714 \leq y_0 \leq 0.6945) = 0.99$$

3. To find the constant between the mean prediction and individual prediction we can use $\text{var}(\hat{y}) = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$.

from the equation, if $x_0 - \bar{x}$ is large it means x_0 is far away from \bar{x} which make the variance large. And large variance would lead to a large standard error. Therefore, in order to cover most data as much as we can to make the best error, the confidence interval must be large.