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

Q1

A) $\hat{\beta}_2 = \frac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - (\sum x_i)^2}$ $= \frac{18(1254.9) - (388)(50.9)}{18(9620) - 150544}$

$= \frac{2839}{22616} = 0.1255$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x}$ $= \frac{509}{180} - 2.7052 = 0.1226$

$\therefore$ Estimator of $\hat{\beta}_2$ is 0.1255 and $\hat{\beta}_1$ is 0.1226

$\hat{\beta}_1$ is when $x = 0$, on average y is 0.1226

$\hat{\beta}_2$ is when x rise by 1, on average y will go up by 0.1255

B) TSS : $\sum y_i^2 = \sum (y_i - \bar{y})^2$ $= 2.5844$

ESS : TSS - RSS $= 2.5844 - 0.5781$

$= 2.0063$

RSS : $\sum \hat{u}_i^2 = \sum (y_i - \hat{y})^2$ $= 0.5781$

$R^2 = \frac{ESS}{TSS}$ $= \frac{2.0063}{2.5844} = 0.7763$

$\therefore$ The R^2 value of 0.7763 mean approximately 77.63% of the variation in y explaind by x

C) $\hat{y} = 0.1226 + 0.1255 x_i$

$= 0.1226 + 0.1255(30)$

$= 3.8876$

$\therefore$ when x is 30, on average y is approximately 3.8876

D) $Var(u) = E(u - E(u))^2$

$= E(u_i - 0)^2$

$= E(u)^2$

$= \sigma^2$ or $var(u_i) = \hat{\sigma}^2$

$Var(u) = \hat{\sigma}^2 = \frac{\sum (y - \hat{y})^2}{n-2} = \frac{0.5781}{16} = 0.0361$

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

≈ 0.00017

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

$= 0.0915$

$\therefore$ $var(u) = 0.0361$, $var(\hat{\beta}_2) = 0.00017$, $var(\hat{\beta}_1) = 0.0915$

E) 90% confidence

$\beta_2 \pm t_{\frac{\alpha}{2}, n-2} \cdot SE(\hat{\beta}_2)$

$0.1255 \pm t_{0.05, 16} \cdot 0.0131$

$0.1255 \pm 1.746 \cdot 0.0731$

$(0.1026, 0.1484)$

$SE(\hat{\beta}_2) = \sqrt{var(\hat{\beta}_2)}$

$= 0.0131$

We are 90% confident that mean of y increase by between (0.1026, 0.1484) for each extra x

F) $\alpha = 0.05$

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

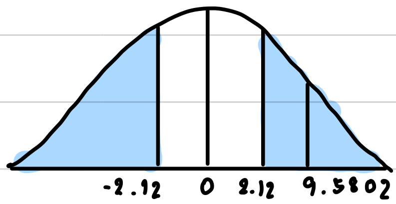
$t = \dfrac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} = \dfrac{0.1255 - 0}{0.0131}$

$= 9.5802$

$t_{\frac{\alpha}{2}, n-2} = t_{0.025, 16}$

$= 2.12$

reject H_0



∴ The null hypothesis is rejected

There is enough evidence to say that $\beta_2 \neq 0$

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 $\bar{X}$.

Q_2

A) $\hat{y} = 0.4279898 + 0.0031338x$

$\alpha = 0.05$

$H_0 : \beta_1 = 0$

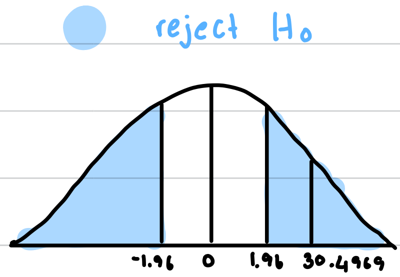
$H_1 : \beta_1 \neq 0$

$t = \dfrac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \dfrac{0.4279898 - 0}{0.0140339}$

$= 30.4969$

$t_{\frac{\alpha}{2}, n-2} = t_{0.025, 27,884}$

$= 1.96$



$\therefore$ The null hypothesis is rejected

There is enough evidence to say that $\beta_1 \neq 0$

$H_0 : \beta_2 = 0$

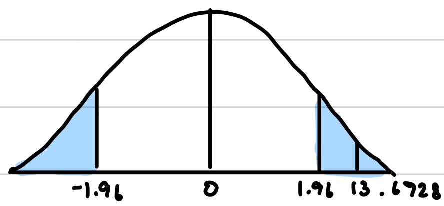
$H_1 : \beta_2 \neq 0$

$t = \dfrac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} = \dfrac{0.0031338 - 0}{0.0002292}$

$= 13.6728$

$t_{\frac{\alpha}{2}, n-2} = t_{0.025, 27,884}$

$= 1.96$



$\therefore$ The null hypothesis is rejected

There is enough evidence to say that $\beta_2 \neq 0$

B) $\therefore$ $\hat{\beta}_2$ is about 0.003, suggest in that age of a person goes up by 1 year, on average, number of time person visit hospital will go up about 0.003 times

$\therefore$ Yes, if the age increase, the person have to visit hospital more frequency

C) In outp. $= \beta_1 + \beta_2$ age $: + u :$

$\ln \hat{y} = 0.4279898 + 0.0031338x$

$\beta_2 = 0.0037338$

It mean that an increase in age of 1 year, on average level to about 0.31% number of time person visit hospital increase

D) $\hat{y} = 0.427988 + 0.0031338x$

$W_1 = 1$

$W_2 = \frac{1}{10}$

$\hat{\beta}_1^+ = W_1 \hat{\beta}_1 = 1 \cdot 0.4279898$
$= 0.4279898$

$\hat{\beta}_2^+ = \left(\frac{W_1}{W_2}\right)\hat{\beta}_2 = 10 \cdot 0.0031338$
$= 0.031338$

$SE(\hat{\beta}_1^+) = W_1 \cdot SE(\hat{\beta}_1) = 1 \cdot 0.140339$
$= 0.140339$

$SE(\hat{\beta}_2^+) = \left(\frac{W_1}{W_2}\right) SE(\hat{\beta}_2) = 10 \cdot 0.0002292$
$= 0.002292$

Confidence interval at 95%

$\beta_1 = \hat{\beta}_1^+ \pm t_{\frac{\alpha}{2}} \cdot SE(\hat{\beta}_1^+)$
$= 0.4279898 \pm t_{0.025, 29,886} \cdot 0.140339$
$= (0.1529, 0.70305)$

$\beta_2^+ = \hat{\beta}_2^+ \pm t_{\frac{\alpha}{2}} \cdot SE(\hat{\beta}_2^+)$
$= 0.031338 \pm 1.96\,(0.002292)$
$= (0.0268, 0.0358)$

E) $\hat{y} = 0.4279898 + 0.0031338x$
$= 0.4279898 + 0.0031338(50)$
$= 0.5847$

$Var(\hat{y}_0) = 0.00002$

$SE(\hat{y}_0) = 0.0045$

$t_{\frac{\alpha}{2}, n-2} = t_{0.005, 29,884}$
$= 2.576$

$E(Y_0 \mid X_0) = \hat{y} \pm t_{\frac{\alpha}{2}, n-2} \cdot SE(\hat{y}_0)$
$= 0.5847 \pm 2.576\,(0.0045)$
$= (0.5731, 0.5964)$

At $\alpha = 0.01$ the confidence interval of mean prediction at age of 50 years old is between 0.5731 and 0.5964

Q3

According to both mean prediction and individual prediction variance formular if we set x more further away from $\bar{x}$, the both variance is getting larger, therefore confidence inteval for there is wider than when we set x near $\bar{x}$