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.

Assignment 1
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: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID(0, \sigma^2)$, **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.

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 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.

Rearrange equation: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$

objective f^n :

$$\min_{\hat{\beta}_1, \hat{\beta}_2} \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

solve $\hat{\beta}_1$: $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = \frac{\partial}{\partial \hat{\beta}_1}\left[\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2\right]$

$= -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$

$\sum Y_i - \sum \hat{\beta}_1 - \hat{\beta}_2 \sum X_i = 0$

$\sum Y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum X_i = 0$

$\hat{\beta}_1 = \frac{\sum Y_i}{n} - \frac{\hat{\beta}_2 \sum X_i}{n}$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$ — ①

solve $\hat{\beta}_2$: $\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = \frac{\partial}{\partial \hat{\beta}_2}\left[\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2\right]$

$= -2 X_i \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$

$\sum X_i (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$

plug ① into $\sum X_i \left(Y_i - (\bar{Y} - \hat{\beta}_2 \bar{X}) - \hat{\beta}_2 X_i\right) = 0$

$\sum X_i \left[Y_i - \bar{Y} - \hat{\beta}_2 (X_i - \bar{X})\right] = 0$

$\hat{\beta}_2 \sum X_i (X_i - \bar{X}) = \sum X_i (Y_i - \bar{Y})$

$\hat{\beta}_2 = \frac{\sum X_i (Y_i - \bar{Y})}{\sum X_i (X_i - \bar{X})}$ — ②

From ② rearrange so that

$\hat{\beta}_2 = \frac{\sum X_i (Y_i - \bar{Y})}{\sum X_i (X_i - \bar{X})}$

$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$

$\hat{\beta}_2 = \frac{20.58}{211} = 0.0975$ #

Sub $\hat{\beta}_2 = 0.0975$ into $\hat{\beta}_1 = \frac{\sum Y_i}{n} - \frac{\hat{\beta}_2 \sum X_i}{n}$

$\hat{\beta}_1 = \frac{50.9}{18} - (0.0975)\frac{388}{18}$

$= 0.7253$ #

The intercept is $\hat{\beta}_1 = 0.7253$ while the slope is $\hat{\beta}_2 = 0.0975$

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

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{0.5781}{2.5844}$$

$$= 0.7763 \ \#$$

R^2 tells us the measurement of 'goodness of fit' of the fitted regression line.

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

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$\hat{Y}_i = 0.7253 + 0.0975(30)$$

$$\hat{Y}_i = 3.6503 \ \#$$

when $X_i = 30$, average of $\hat{Y}_i = 3.6503$ #

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

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

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

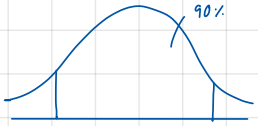
$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2} = \frac{0.0361}{211} = 1.7109 \times 10^{-4} \ \#$$

warisara 6304641795

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

d.f = n-k
= 18-2 = 16

CI = 1-α 0.9 = 1-α α = 0.1 #



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

= 1.746 #

$P\left[\hat{\beta}_2 - \delta \le \beta_2 \le \hat{\beta}_2 + \delta\right] = 1-\alpha$

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

$P\left[0.0975 - (1.746)(0.0131) \le \beta_2 \le 0.0975 + (1.746)(0.0131) = 1-\alpha\right.$

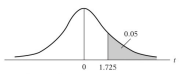
$P\left[0.0746 \le \beta_2 \le 0.1204\right] = 0.9$ #

$se(\hat{\beta}_2) = \sqrt{var(\hat{\beta}_2)} = \sqrt{1.7109 \times 10^{-9}}$

$= 0.0131$ #

∴ which means that 90 out of 100 times, β_2 is between 0.0746 and 0.1204 #

Example
$Pr(t > 2.086) = 0.025$
$Pr(t > 1.725) = 0.05$ for df = 20
$Pr(|t| > 1.725) = 0.10$



Pr / df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

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

Testing β_2 : set up null and alternative hypothesis

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

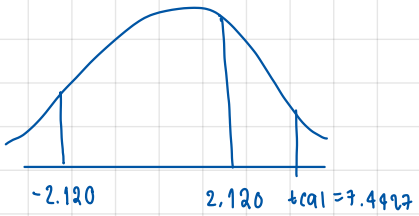
H_a : $\beta_2 \neq 0$ (claim)

compute t_{cal}

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = \frac{0.0975 - 0}{0.0131} = 7.4427$$

Find critical value

$t_{\frac{\alpha}{2}} = \pm 2.120$



Conclusion:

t_{cal} is beyond the critical value, therefore we can reject null hypothesis and it is sure that β_2 is not zero 95 out of 100 times.

Example
$Pr(t > 2.086) = 0.025$
$Pr(t > 1.725) = 0.05$ for df = 20
$Pr(|t| > 1.725) = 0.10$

Pr / df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

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                      .0026846    .003583
       _cons |   .4279898   .0140339                      .4004828    .4554969
------------------------------------------------------------------------------
```

Omitted

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 $\bar{X}$.

---

warisara 6304641795

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

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

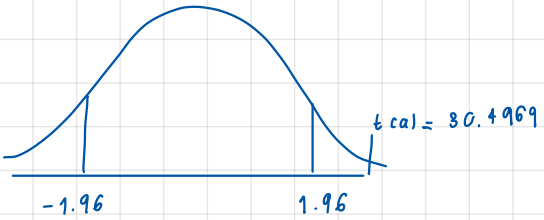
compute t_{cal}:

$$\frac{\hat{\beta}_1 - \beta_1}{se\,\hat{\beta}_1} = \frac{0.4279898 - 0}{0.0140339} = 30.4969 \ \#$$

critical value:

$$t_{\frac{\alpha}{2}} = t_{0.025} = \pm 1.96 \ \#$$

d.f = 27,884

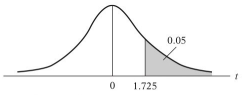


Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df = 20

$\Pr(|t| > 1.725) = 0.10$



Pr / df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

conclusion:

t_{cal} is beyond the critical value, therefore we can reject null hypothesis and it is sure that β_1 is not zero 95 out of 100.

$H_0: \beta_2 = 0$

$\beta_2 \neq 0$

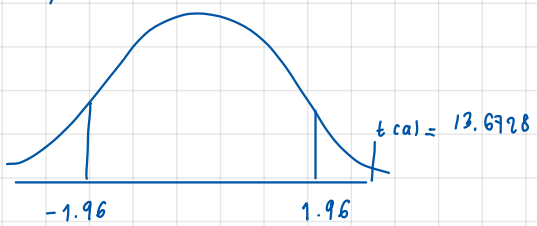
compute tcal

$$\frac{\hat{\beta}_2 - \beta_2}{Se\,\hat{\beta}_2} = \frac{0.0031338 - 0}{0.0002292} = 13.6728 \ \#$$

critical value:

$$t_{\frac{\alpha}{2}} = t_{0.025} = \pm 1.96 \ \#$$

d.f = 27,884



conclusion:

t_{cal} is beyond the critical value, therefore we can reject null hypothesis and it is sure that β_2 is not zero 95 out of 100.

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

$\hat{\beta}_2$ is the slope of the equation so when age_i change by 1 year, the times the person has visited increase by 0.0031338.
The sign of $\hat{\beta}_2$ is positive which makes economic sense because there is a positive correlation between age and how many times the person has visited the hospital. The reason is because as you age, there are more health problems.

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.

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

$$\frac{d \ln \widehat{outp}_i}{d age} = 0.0031338 \ \#$$

$$\ln(y) = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

$$\frac{d \ln y}{dx} = \hat{\beta}_2$$

$$\frac{\frac{dy}{y}}{dx} = \hat{\beta}_2$$

$$\hat{\beta}_2 = \frac{dy}{y} \cdot dx$$

An increase in one year of age is associate with output increase by $\hat{\beta}_2 \cdot 100 = 0.3134$ percent visiting the hospital.

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).

scaling age_i affects only the unit of age in coefficient, standard error and confidence interval. constant stays the same at 0.4179898 times while slope which is the coefficient of age changes from 0.0031338 years to 0.00031338 years.

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

$$Pr\left[\hat{Y}_0 - \left(t_{\frac{\alpha}{2}} \cdot Se\hat{Y}_0\right) \leq Y_0 \in \hat{Y}_0 + \left(t_{\frac{\alpha}{2}} \cdot Se\hat{Y}_0\right)\right] = 1-\alpha$$

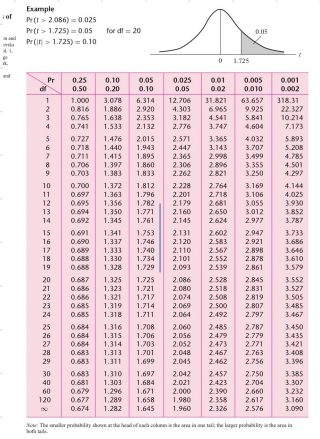
$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$\hat{Y}_i = 0.4279898 + 0.0031338\, X_i$$

so : $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ where $X_0 = 50$

$$\hat{Y}_0 = 0.4279898 + 0.0031338(50)$$

$$\hat{Y}_0 = 0.5847 \ \#$$



Example

$\Pr(t > 2.086) = 0.025$
$\Pr(t > 1.725) = 0.05$ for df = 20
$\Pr(|t| > 1.725) = 0.10$

Pr / df	0.25 / 0.50	0.10 / 0.20	0.05 / 0.10	0.025 / 0.05	0.01 / 0.02	0.005 / 0.010	0.001 / 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

$t_{\frac{\alpha}{2}} = t_{0.005}$ d.f. = 27,884

$= 2.576$ #

$$Pr\left[0.5847 - (2.576)(\sqrt{0.00002}) \leq Y_0 \leq 0.5847 + (2.576)(\sqrt{0.00002})\right] = 1 - 0.01$$

$$Pr\left[0.5732 \leq Y_0 \leq 0.5962\right] = 0.99$$

∴ which means that 99 out of 100 times, Y_0 is between 0.5732 and 0.5962 #

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 $\bar{X}$.

Mean prediction provides the mean estimation where $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ X_0 represents a value of interest. Individual prediction estimates the variance around Y_0, and focuses on the forcasting error (fe). $fe = \hat{Y}_0 - Y_0$. If $X_0 - \bar{X}$ is very different. var will be bigger, Se will be bigger so confidence interval higher.