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.

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 $\text{var}(u_i)$, $\text{var}\left(\hat{\beta}_1\right)$ and $\text{var}\left(\hat{\beta}_2\right)$.

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.

β_2

$$n = 18 \qquad \sum_{i=1}^{n} X_i = 388.00 \qquad \sum_{i=1}^{n} Y_i = 50.90$$

$$\sum_{i=1}^{n} (X_i)^2 = 9{,}620.00 \qquad \sum_{i=1}^{n} X_i Y_i = 1{,}254.90$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 211.00 \qquad \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 2.5844$$

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 20.58 \qquad \sum_{i=1}^{n} \hat{u}_i^{\,2} = 0.5781$$

a)

$$\beta_2 = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

$$= \frac{20.58}{(211)} = 0.0975$$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{388}{18}$$

$$= 21.5555$$

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

$$= 2.8278$$

$$\beta_1 = \bar{y} - \beta_2 \bar{x}$$

$$= 2.8278 - 0.0975(21.5555)$$

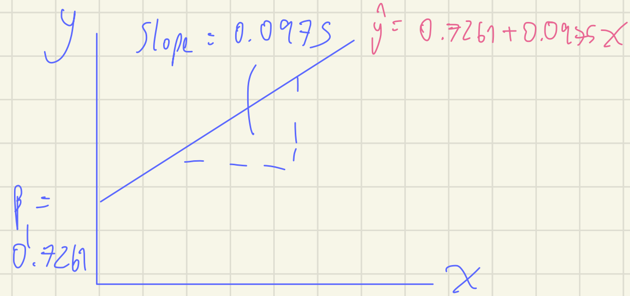
$$= 0.7261$$

$$\therefore \hat{y} = 0.7261 + 0.0975X$$

$$\hat{y} = \overset{\beta_1}{0.7261} + \overset{\beta_2}{0.0975}x$$

$\beta_2 = 0.0975$; if x increases by 1 unit, $\hat{y}$ will increase by 0.0975 unit (in the same direction)

$\beta_1 = 0.7261$; if $x = 0$, $\hat{y} = 0.7261$



b) $R^2 = \dfrac{ESS}{TSS}$ or $1 - \dfrac{RSS}{TSS}$

$ESS = (\hat{y} - \bar{y})^2$ $\quad TSS = \Sigma(y - \bar{y})^2$ $\quad RSS = \Sigma(\hat{y} - y)^2$ or Σu^2

$$\therefore R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{0.5781}{2.5844} = 0.7763$$

The variation of y can be explained by x = 77.63%

c) $\hat{y} = 0.7261 + 0.0975x$

$x_i = 30$

$\hat{y} = 0.7261 + 0.0975(30)$

$= 3.6511$

$\therefore$ When $x_i = 30$, the value of Y will be 3.6511 on average.

d) $Var(\hat{u}) = \dfrac{\sum \hat{u}^2}{n-2}$ $\quad Var(\beta_2) = \dfrac{\sigma^2}{\sum(x-\bar{x})^2}$

$var(\beta_1) = \dfrac{\sigma^2 \cdot \sum x^2}{n \cdot \sum(x-\bar{x})^2}$

$Var(\hat{u}) = \dfrac{0.5781}{18-2} = 0.0361$ $\quad Var(\beta_2) = \dfrac{0.0361}{211}$
(σ^2)

$= 0.00017$

$var(\beta_1) = \dfrac{0.0361 \cdot 388}{18(211)} = 0.0036$

e) Confident Interval β

$\alpha = 0.1$

$\alpha/2 = \frac{0.1}{2} = 0.05$

$$= \hat{\beta}_2 \pm \left[t_{\alpha/2,\, n-2} \right] \cdot se(\hat{\beta}_2)$$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

$t = 1.746$

$se(\beta) = \sqrt{var}$

$se(\beta_2) = \sqrt{0.00017}$

$= 0.0130$

Confident interval

$= 0.0975 \pm 1.746 \cdot 0.0130$

$= 0.0748,\ 0.1202$

$\therefore$ According to the regression, the true β_2 is between 0.0748 and 0.1202 at 90% significant level.

f) Hypothesis testing

Step 1 set hypothesis

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

Step 2 Significant level = 0.05

$\alpha/2 = \frac{0.05}{2} = 0.025$

Step 3 t-cal

$$t = \frac{\hat{\beta}_2 - \beta_2}{Se(\hat{\beta}_2)}$$

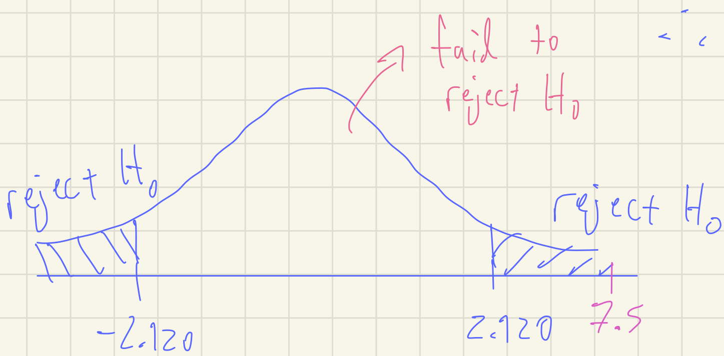
$$t = \frac{0.0975 - \overset{H_0}{0}}{0.0130}$$

$$= 7.5$$

Step 4

t - critical = 2.120

Step 5



$\therefore$ Since t-cal falls in reject H_0 area, we reject H_0 or $\beta_2 \neq 0$ at 95% confident level.

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	Number of obs	=	27,886
Residual	11565.0627	27,884	.414756231	F(1, 27884)	=	186.96
Total	11642.6072	27,885	.417522223	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 $\bar{X}$.

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

$$Outp = \beta_1 + \beta_2 age + u$$

$$\beta_2 = 0.0031338$$

$$\beta_1 = 0.4279898$$

Regression :

$$Outp = 0.4279898 + 0.0031338\ age$$

$$Se(\beta_1) = 0.0140339 \quad Se(\beta_2) = 0.0002292$$

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

test β_1

Step 1 set hypothesis

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Step 2

test β_2

Step 1

$H_0 : \beta_1 = 0$

$H_1 : \beta_2 \neq 0$

Significant level = 0.05

H_1 is two tails

$\therefore$ Significant level$/2$ = 0.025

Step 3

$$t\text{-test} = \frac{\hat{\beta}_1 - \beta_1}{Se(\beta_1)}$$

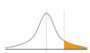
$$t\text{-test} = \frac{0.42798 - 0}{0.014}$$

$$= 30.57$$

$df = 27886 - 2 = 27884$

Step 4

Table 2 T-Distribution Table



$P(t > 2.23) = .025$ d.f. = 10

df	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
	Tail probability p											
1	1.000	1.376	1.963	3.078	6.314	12.706	15.895	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291

t-critical = 1.960

Step 2

H_1 is two tails

$\therefore$ Significant level$/2$ = 0.025

Step 3

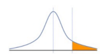
$$t\text{-test} = \frac{\hat{\beta}_2 - \beta_1}{Se(\beta_2)}$$

$$= \frac{0.0031338 - 0}{0.0002292}$$

$$= 13.6728$$

Step 4

Table 2 T-Distribution Table



$P(t > 2.23) = .025$ d.f. = 10

df	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
	Tail probability p											
1	1.000	1.376	1.963	3.078	6.314	12.706	15.895	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291

t-critical = 1.960

Step 5

$\therefore$ Reject H_0

Step 6

β_1 is different from zero at 0.05 significant level

Step 5

Step 6

β_2 is different from zero at 0.05 significant level

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

$$Outp = 0.4278 + 0.0031\,age$$

if age increases by 1 year, outp will increase by 0.0031. Which make economic sense because as the ages increase, people will visit hospital more due to their health condition.

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.

$$\ln outp = 0.4278 + 0.0031 \, age$$

if age increase by 1 year, outp will increase by $0.0031 \times 100\% = 0.31\%$

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).

$$Oupt = \beta_1 + \beta_2 \, age$$

$$y = \beta_1 + \beta_2 x + v$$

$$\hat{\beta}_2 = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

$$X_{new} = x_i/10$$

$$\hat{\beta}_{2\,new} = \frac{\sum[x_{new} - \bar{x}_{new}][y - \bar{y}]}{\sum[x_{new} - \bar{x}]^2}$$

$$= \frac{\sum\left[\frac{x}{10} - \frac{\bar{x}}{10}\right][y - \bar{y}]}{\sum\left[\frac{x}{10} - \frac{\bar{x}}{10}\right]^2}$$

$$= \frac{\cancel{\frac{1}{10}}\sum[x - \bar{x}][y - \bar{y}]}{\left[\frac{1}{10}\right]^{\cancel{2}}\sum[x - \bar{x}]^2}$$

$$\hat{\beta}_{2\,new} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\left[\frac{1}{10}\right]\sum(x - \bar{x})^2}$$

$$= \frac{1}{0.1}\left[\cancelto{\beta_{2\,old}}{\frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}}\right]$$

$$\hat{\beta}_{2\,new} = \frac{\hat{\beta}_{2\,old}}{0.1} = \frac{0.0031}{0.1} = 0.031$$

$$\hat{\beta}_{1\,new} = \bar{y} - \beta_{2\,new}\,\bar{x}$$

$$\hat{\beta}_{1\,new} = \bar{y} - \left[\frac{\beta_2}{0.1}\right]\left[\frac{\bar{x}}{10}\right]$$

$$\hat{\beta}_{1\,new} = \bar{y} - \beta_2\bar{x}$$

$$\hat{\beta}_{1\,new} = \hat{\beta}_{1\,old} = 0.4272 \;\#$$

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

$$\alpha/2 = 0.005$$

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

$$Se(\hat{Y}_0) = \sqrt{Var(\hat{Y}_0)} = 0.0045$$

Confidence interval (mean prediction):

$$\hat{Y} \pm t_{\alpha/2,\, n-k} \cdot Se(\hat{Y}_0)$$

$$= 0.583 \pm 2.576 \cdot 0.0045$$

$$= 0.5714,\ 0.5946$$

$\therefore$ The confidence interval of mean prediction at age of 50 years old is between 0.5714 and 0.5946 at 99% confidence level.

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 $\bar{X}$.

$$\hat{Y} \pm t_{\alpha/2, n-k} \sqrt{\sigma^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma x_i^2}\right)}$$

- If $X_0 - \bar{X}$ in individual prediction get larger, it means that the standard error (Se) will also get larger too because the information is getting wider so it is possible for error to be occured.