Assignment 1
Assigned on Apr 9th, 2022. Due date Saturday, May 7th, 2022 before midnight.

Tha Taritta Suppasadid
6304641571

Instructions

(1) Please read the instruction carefully.
(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.

Question 1. (12 points) Economic model of Crime.

1.a) Based on the regression results provided, write out the estimated coefficients in the form of regression equation (1.1). Interpret the estimated coefficients associated with $avgsen$. Based on Model (1.1), test whether the average sentence served from prior convictions has an impact on the number of arrests in the current year (1986). Show your work. (Use α = 0.05)

1.b) What is the overall significance of the regression from Model (1.1) and Model (1.2)? What test do you use? (Use α = 0.01)

1.c) If we are interested in testing whether "ethnic background and legal income" has an impact on the number of arrests in the current year (1986), what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use α = 0.05)

Estimate the model (1.1) reports in the Table 1.1

$$narr86_i = \beta_1 + \beta_2 pcnv_i + \beta_3 avgsen_i + \beta_4 tottime_i + \beta_5 ptime86_i + \beta_6 qemp86_i + u_i \quad \textbf{(1.1)}$$

Table 1.1

Source	SS	df	MS			
				Number of obs	=	2,725
				F(5, 2719)	=	24.29
Model	85.9532425	5	17.1906485	Prob > F	=	0.0000
Residual	1924.39391	2,719	.707757967	R-squared	=	0.0428
				Adj R-squared	=	0.0410
Total	2010.34716	2,724	.738012906	Root MSE	=	.84128

narr86	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
pcnv	-.1512246	.040855	Omitted for the purpose of this exam		
avgsen	-.0070487	.0124122			
tottime	.0120953	.0095768			
ptime86	-.0392585	.0089166			
qemp86	-.1030909	.0103972			
_cons	.7060607	.0331524			

Estimate the model (1.2) reports in the Table 1.2

$$narr86_i = \beta_1 + \beta_2 pcnv_i + \beta_3 avgsen_i + \beta_4 tottime_i + \beta_5 ptime86_i + \beta_6 qemp86_i + \beta_4 inc86_i + \beta_5 black_i + \beta_6 hispan_i + u_i \quad \textbf{(1.2)}$$

where

$narr86_i$	= the number of arrests in the current year (1986)
$pcnv_i$	= the proportion of prior arrests that led to a conviction
$avgsen_i$	= the average sentence served from prior convictions (in months)
$tottime_i$	= months spent in prison since age 18 prior to 1986
$ptime86_i$	= months spent in prison in 1986
$qemp86_i$	= the number of quarters that the man was legally employed in 1986
$inc86_i$	= legal income, 1986, (hundred dollars)
$black_i$	= 1 if black ethnic background
$hispan_i$	= 1 if Hispanic ethnic background

Table 1.2

Source	SS	df	MS			
				Number of obs	=	2,725
				F(8, 2716)	=	26.47
Model	145.390104	8	18.173763	Prob > F	=	0.0000
Residual	1864.95705	2,716	.686655763	R-squared	=	0.0723
				Adj R-squared	=	0.0696
Total	2010.34716	2,724	.738012906	Root MSE	=	.82865

narr86	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
pcnv	-.1332344	.0403502	Omitted for the purpose of this exam		
avgsen	-.0113177	.0122401			
tottime	.0120224	.0094352			
ptime86	-.0408417	.008812			
qemp86	-.0505398	.0144397			
inc86	-.0014887	.0003406			
black	.3265035	.0454156			
hispan	.1939144	.0397113			
_cons	.5686855	.0360461			

Question 2. (12 points) Dummy variables and interaction terms.

Using the Thailand labor force survey (LFS) in quarter 2 of 2019 and 2020, employees log of wage is modeled as follows. (Number of observations is 97,878 in total)

$$\ln wage_i = \beta_1 + \beta_2 civil_i + \beta_3 year_i + \beta_4 civil_i \cdot year_i + u_i$$

where $\ln wage_i$ = natural logarithmic scale of monthly wage

$civil_i$ = 1; civil servant and state employee

= 0; otherwise

$year_i$ = 1; year 2020

= 0; otherwise (2019)

This model is also known as Difference-in-Differences (DiD) and its intention is to capture the effect of COVID-19 since March of 2020 on different types of employment. During the pandemic, we assume that civil servant and state employee's wage is not reduced (control group) while others', namely employees in private firms or freelance, etc., is suspected to be reduced (treatment group). The estimation result is shown below with standard errors in parentheses. Answer the following questions.

$$\ln \widehat{wage}_i = \underset{(0.0035)}{9.1748} + \underset{(0.0072)}{0.587}\, civil_i - \underset{(0.005)}{0.0336}\, year_i + \underset{(0.0102)}{0.0444}\, civil_i \cdot year_i + u_i$$

2.a) Test all the parameters individually if each of them is significantly different from zero of not.

2.b) How much on average does a civil servant and state employee earns more or less than the others disregarding the year?

2.c) How much on average does the pandemic affect wage overall?

2.d) Are the control group and the treatment group better-off or worse-off during the pandemic. Discuss each group separately, show your work and explain with economic reasons according to the intention of this model.

Question 3. (8 points) Multicollinearity.

As cheese ages, several chemical processes take place that determine the taste of the final product. The data given pertain to concentrations of various chemicals in a sample of 30 mature cheddar cheeses and subjective measure of taste for each sample.

Estimate the model (3.1) reports in the Table 3.1

$$Taste = \beta_0 + \beta_1 acetic + \beta_2 h2s + \beta_3 lactic + u \qquad \textbf{(3.1)}$$

Where $Taste$ = Measures of taste for each sample

$acetic$ = The natural logarithm of concentration of acetic

$h2s$ = The natural logarithm of concentration of hydrogen sulfide

$lactic$ = Lactic

Table 3.1

Source	SS	df	MS
Model	5020.64468	3	1673.54823
Residual	2642.24237	26	101.624706
Total	7662.88705	29	264.237485

Number of obs	=	30
F(3, 26)	=	16.47
Prob > F	=	0.0000
R-squared	=	0.6552
Adj R-squared	=	0.6154
Root MSE	=	10.081

taste	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
acetic	1.538645	3.000501	Omitted for the purpose of this exam		
h2s	3.915242	1.153106			
lactic	18.80235	8.342614			
_cons	-34.13491	15.67628			

	acetic	h2s	lactic
acetic	1.0000		
h2s	0.2700	1.0000	
lactic	0.3607	0.6448	1.0000

Variable	VIF	1/VIF
lactic	1.83	0.546648
h2s	1.72	0.582609
acetic	1.15	0.867477
Mean VIF	1.57	

3.a) Is there evidence of multicollinearity in the data? How do you know? Explain your answers in detail and state the critical value for hypothesis testing to receive full points.

3.b) What is the property of BLUE? If there is the multicollinearity problem, is the OLS estimators still retain the property of BLUE? If not, which properties are violated?

Question 4. (8 points) Heteroscedasticity.

The data on U.S. inflation rates (%) and unemployment rates (%), 1948-2006

Estimate the model (4.1) reports in the Table 4.1

$$Inf_t = \beta_1 + \beta_2 unem_t + u_t \quad \textbf{(4.1)}$$

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

Table 4.1

Source	SS	df	MS
Model	32.3284496	1	32.3284496
Residual	478.096987	57	8.38766644
Total	510.425437	58	8.80043856

Number of obs	=	59
F(1, 57)	=	3.85
Prob > F	=	0.0545
R-squared	=	0.0633
Adj R-squared	=	0.0469
Root MSE	=	2.8961

inf	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
unem	.5054734	.2574699	Omitted for the purpose of this exam		
_cons	1.010847	1.491583			

White's general test statistic: 1.0266 Chi-sq (2)

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

chi2(1) = 1.12

Answer the following questions.

4.a) Interpret the intercept and slope coefficients.

4.b) According to the test statistics given after Table 4.1 below, is there any sufficient evidence to conclude that there is heteroscedasticity problem? Show your work on the hypothesis testing. (Use $\alpha = 0.05$)

4.c) Given your test results in a), do the OLS estimators still retain the property of BLUE? If not, which properties are violated?

****

1.a) Based on the regression results provided, write out the estimated coefficients in the form of regression equation (1.1). Interpret the estimated coefficients associated with *avgsen*. Based on Model (1.1), test whether the average sentence served from prior convictions has an impact on the number of arrests in the current year (1986). Show your work. (Use α = 0.05)

from $narr86_i = \beta_1 + \beta_2 pcnv_i + \beta_3 avgsen_i + \beta_4 tottime + \beta_5 ptime86_i + \beta_6 qemp86 + u_i$

$narr86_i = 0.7061 - 0.1512(pncnv) - 0.0070(avgsen) + 0.0121(tottime)$

$- 0.3926(ptime86) - 0.1031(qemp86) + u_i$

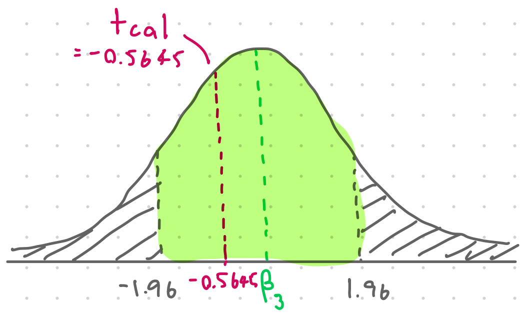
① $H_0: \beta_3 = 0$

$H_a: \beta_3 \neq 0$

② $t_{cal}: \dfrac{\hat{\beta}_3 - \beta_3}{se\beta_3} = \dfrac{-0.0070 - 0}{0.0124} = -0.5645$

③ $t_{lower} = t_{\frac{0.05}{2}, n-k} = t_{0.025(2725-6)} = -1.96$

$t_{upper} = t_{\frac{0.05}{2}, n-k} = t_{0.025(2725-6)} = 1.96$



Cannot Reject H_0: we cannot make sure that $\beta_3 \neq 0$ 95% out of times

∴ The average sentence served from prior conviction has no impact on the number of arrests in the current year (1986) 95% out of times #

1.b) What is the overall significance of the regression from Model (1.1) and Model (1.2)?
What test do you use? (Use α = 0.01)

I'll use the F-test to check the overall significance of the regression

Model 1.1

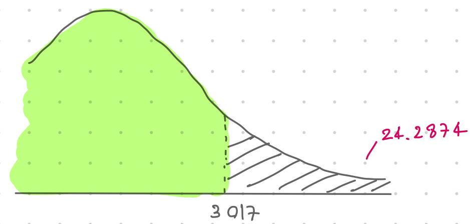
① $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

H_a: otherwise

② $F_{cal} = \dfrac{ESS/K-1}{RSS/n-k} = \dfrac{85.9532/6-1}{1924.3939/2725-6}$

$= \dfrac{17.19064}{0.7078}$

$= 24.2874 \approx 24.29$



③ F_{cri} : $F_{0.01}(6-1, 2719) = 3.017$

Reject H_0: we can make sure that $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, and β_6 are not simulteneously equal to zero 99% out of the times, thus this model is not valid

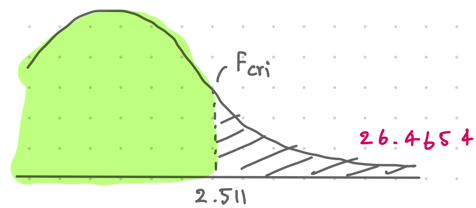
Model 1.2

① $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6$

H_a: otherwise

② $F_{cal} = \dfrac{145.3901 \; /9-1}{1864.9571/2725-9} = \dfrac{18.1738}{0.6867}$

$= 26.4654 \approx 26.47$



③ $F_{cri} = F_{0.01}(9-1, 2716) = 2.511$

Reject H_0: we can make sure that $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, and β_6 are not equal to zero simulteneously, therefore, the model is not valid #

1.c) If we are interested in testing whether "ethnic background and legal income" has an impact on the number of arrests in the current year (1986), what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use α = 0.05)

Marginal contribution test would be tested whether "ethnic background and legal income" has and impact on the number of arrests in 1986

① H_0 : ethnic background and legal income has no impact on the number of arrest

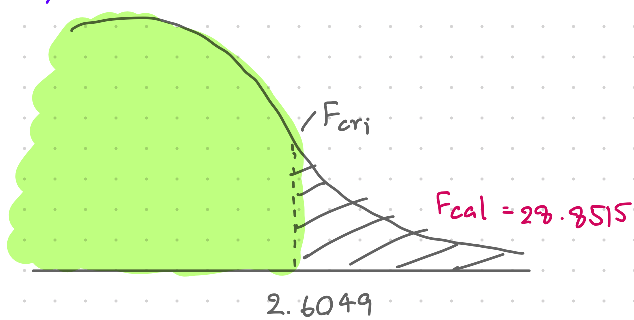
H_a : otherwise

② $F_{cal} = \dfrac{ESS_{new} - ESS_{old} / \#\text{ of new regressor}}{RSS_{new} / N - k_{new}}$

$= \dfrac{145.3901 - 85.9532 / 3}{1964.9571 / 2725 - 9} = \dfrac{19.8123}{0.6867}$

$= 28.8515$

③ $F_{cri} = F_{0.05}(3, 2716) = 2.6049$



Reject H_0 : We cannot make sure that ethnic background and legal income has no impact on number of arrest of the current year (1986) 95 times out of 100 #

2.a) Test all the parameters individually if each of them is significantly different from zero of not.

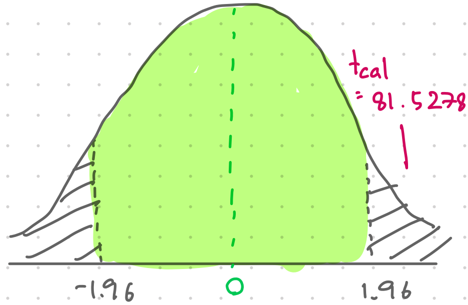
β_2

① $H_0 : \beta_2 = 0$

H_a : Otherwise

② $t_{cal} : \dfrac{\hat{\beta}_2 - \beta_2}{se\hat{\beta}_2} = \dfrac{0.587 - 0}{0.0072} = 81.5278$

③ $t_{cri} : t_{\frac{0.05}{2}}(n-k) = t_{\frac{0.25}{2}}(97{,}874) = \pm 1.96$



Reject H_0 : we can make sure that β_2 is significantly different from zero 95% of the times #

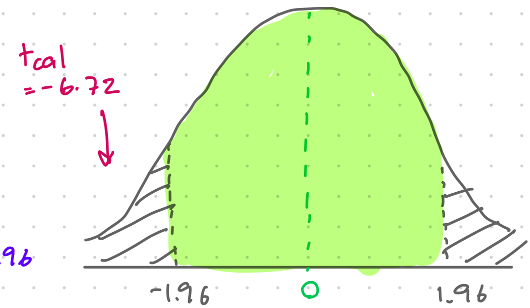
β_2

① $H_0 : \beta_3 = 0$

H_0 : Otherwise

② $t_{cal} : \dfrac{\hat{\beta}_3 - \beta_3}{se\hat{\beta}_3} = \dfrac{-0.0336 - 0}{0.005} = -6.72$

③ $t_{cri} : t_{\frac{0.05}{2}}(n-k) = t_{\frac{0.25}{2}}(97{,}874) = \pm 1.96$



Reject H_0: we can make sure that β_3 is significantly different from zero 95% out of the times #

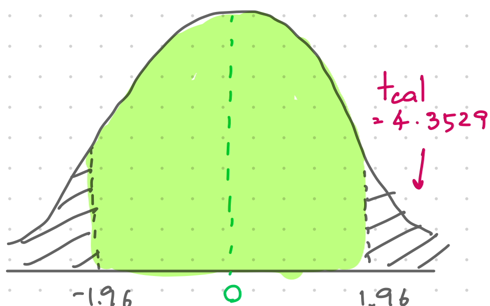
β_2

① $H_0 : \beta_4 = 0$

H_0 : Otherwise

② $t_{cal} : \dfrac{\hat{\beta}_4 - \beta_4}{se\hat{\beta}_4} = \dfrac{0.0444 - 0}{0.0102} = 4.3529$

③ $t_{cri} : t_{\frac{0.05}{2}}(n-k) = t_{\frac{0.25}{2}}(97{,}874) = \pm 1.96$



Reject H_0 : we can make sure that β_4 is significantly different from zero 95% out of the times #

2.b) How much on average does a civil servant and state employee earns more or less than the others disregarding the year?

(β_2) $Civil_i = 1$ where it is a civil servant and state employee

$= 0$ where it is otherwise

From $\widehat{\ln wage_i} = 9.1748 + 0.587\, civil_i$

a civil servant and state employee's $\widehat{\ln wage} = 9.1748 + 0.587(1)$

$= 9.7618$

a civil servant and state employee's wage $= e^{9.7618} = 17357.8477$

otherwise $\widehat{\ln wage} = 9.1748 + 0.587(0)$

$= 9.1748$

otherwise's wage $= e^{9.1748} = 9650.8377$

different in wage $= 17,357.8477 - 9,650.8377$

$= 7707.0010$

$\therefore$ On average, a civil servant and state employee earns more than other 7,707.0010 Baht disregarding the year. #

2.c) How much on average does the pandemic affect wage overall?

(β_3) $year_i = 1$ where the year is 2020

$= 0$ where the year is otherwise (2019)

From $\widehat{\ln wage_i} = 9.1748 - 0.0336 year_i$

$\ln \widehat{wage}$ in 2020 $= 9.1748 - 0.0336(1)$

$= 9.1412$

wage in 2020 $= e^{9.1412} = 9331.9568$

$\ln \widehat{wage}$ in otherwise (2019) $= 9.1748 - 0.0336(0)$

$= 9.1748$

wage in otherwise (2019) $= e^{9.1748} = 9650.8877$

different in wage $= 9331.9568 - 9650.8877$

$= -318.8809$

$\therefore$ On averge, the pandamic decreases overall wage by 318.8809 Baht #

2.d) Are the control group and the treatment group better-off or worse-off during the pandemic. Discuss each group separately, show your work and explain with economic reasons according to the intention of this model.

From $\widehat{\ln wage_i} = 9.1748 + 0.587\, civil_i - 0.0336\, year_i + 0.0444(civil_i \cdot year_i)$

① control group — civil servant and state employee ; $civil_i = 1$

$$\rightarrow \widehat{\ln wage_i} = (9.1748 + 0.587) + (-0.0336 + 0.0444)\, year_i$$

$$= 9.7618 + 0.0108\, year_i$$

During pandamic (2020): $year_i = 1$

$$\widehat{\ln wage} = 9.7618 + 0.0108(1) = 9.7726$$

$$wage = e^{9.7726} = 17{,}546.3284$$

Before pandemic (2019) : $year_i = 0$

$$\widehat{\ln wage} = 9.7618 + 0.0108(0) = 9.7618$$

$$wage = e^{9.7618} = 17{,}357.8477$$

different $= 17{,}546.3284 - 17{,}357.8477 = 188.4807$

$\therefore$ The control group is better-off during the pandemic by 188.4807 Baht #

The reason behind this is, for control group, when considered only on the year, their wage might reduce due to the pandemic. But when take their job into consideration, we'll see that their benefit as a civil servant and state employee can overcome the reduction in wage during the pandemic. Therefore, a civil servant and state employee is better-off during the pandemic #

(2) treatment group — otherwise ; $civil_i = 0$

$$\rightarrow \widehat{\ln wage} = 9.1748 - 0.0336\, year_i$$

During pandamic (2020): $year_i = 1$

$$\widehat{\ln wage} = 9.1748 - 0.0336(1) = 9.1412$$

$$wage = e^{9.1412} = 9331.9568$$

Before pandemic (2019) : $year_i = 0$

$$\widehat{\ln wage} = 9.1748 - 0.0336(0) = 9.1748$$

$$wage = e^{9.1748} = 9650.8377$$

$$\text{different} = 9331.9568 - 9650.8377 = -318.8809$$

$\therefore$ The treatment group is worse-off by 318.8809 Baht

The reason behind this is other jobs such as employee in private firms, freelance, etc., don't have benefit in wage to begin with like civil servant and state employee, according to the model. When the pandemic hits, others face a reduction in wage. Thus, they are worse-off during the pandemic. #

3.a) Is there evidence of multicollinearity in the data? How do you know? Explain your answers in detail and state the critical value for hypothesis testing to receive full points.

Checking the multicollinearity in the data

- VIF for each independent variable exceed 10 because the higher the value of VIF, the higher the correlation between each variables. If the correlation is too high, the model will be too fluctuate.

VIF of lactic = 1.83 < 10

VIF of h2s = 1.72 < 10

VIF of acetic = 1.15 < 10

- testing t-test and R^2

① $H_0 : \beta_k = 0$

$H_a : \beta_k \neq 0$

where k = 1, 2, 3, 4

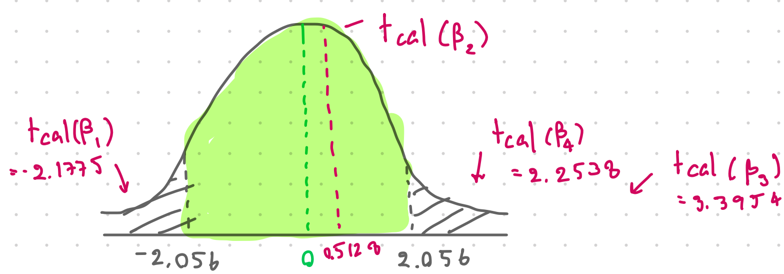
② $t_{cal}(\beta_1) = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{-34.1349 - 0}{15.6762} = -2.1775$

$t_{cal}(\beta_2) = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{1.5386 - 0}{3.0005} = 0.5128$

$t_{cal}(\beta_3) = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = \frac{3.9152 - 0}{1.1531} = 3.3954$

$t_{cal}(\beta_4) = \frac{\hat{\beta}_4 - \beta_4}{se(\hat{\beta}_4)} = \frac{18.8023 - 0}{8.3426} = 2.2538$

③ $t_{cri} = t_{\frac{0.05}{2}(n-k)} = t_{0.025(30-4)} = \pm 2.056$



Reject H_0: we can make sure that β_1, β_2, and β_4 are significantly different from zero 95% of the times

Cannot Reject H_0: we can make sure that β_3 is not significantly different from zero 95% of the time

Therefore, we can make sure that $R^2 = 0.6552$ and most of the parameters are significantly different from zero, thus, there is no conflict in the test. In conclusion, there is no evidence of multicollinearity in this data. #

3.b) What is the property of BLUE? If there is the multicollinearity problem, is the OLS estimators still retain the property of BLUE? If not, which properties are violated?

BLUE or Best Linear Unbiased Estimator is the minimum variance or the narrowest sampling distribution, in other word, fit linear regression the best. OLS coefficient estimate the tightest possible sampling distribution of unbiased estimates. BLUE is not violated by multicollinearity, but multicollinearity can lead to false conclusion of hypothesis testing. #

4.c) Given your test results in a), do the OLS estimators still retain the property of BLUE? If not, which properties are violated?

Since we cannot reject H_0, BLUE property is not violated. #

4.a) Interpret the intercept and slope coefficients.

From $Inf_t = \beta_1 + \beta_2 \, unem_t + u_t$

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

the regression function is

$$Inf_t = 1.0108 + 0.50547 unem_t + u_t$$

The intercept of this model (β_1) is equal to 1.0108, which means that without unemployment, the inflation rate will be at 1.0108%. And the slope of the model (β_2) is equal to 0.5055, can tell that if the unemployment rate increase by 1%, the inflation rate will increase by 0.5055% #

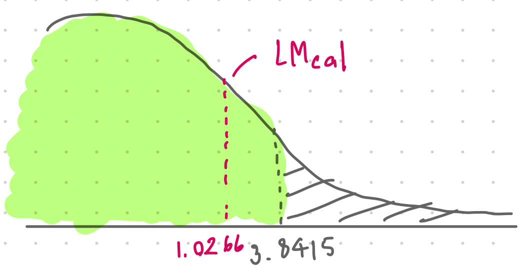
4.b) According to the test statistics given after Table 4.1 below, is there any sufficient evidence to conclude that there is heteroscedasticity problem? Show your work on the hypothesis testing. (Use α = 0.05)

Using White's general test

1. H_0 = heteroscedasticity

   H_a = otherwise

2. $LM_{cal} = \chi^2_{K-1} = \chi^2_1 = 1.0266$
3. Critical value of $\chi^2_{1\ 0.05} = 3.8415$



Cannot Reject H_0 : we can say for sure that there is heteroscedasticity problem in this data 95% out of times #