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

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.

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

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				
avgsen	-.0070487	.0124122				
tottime	.0120953	.0095768				
ptime86	-.0392585	.0089166				
qemp86	-.1030909	.0103972				
_cons	.7060607	.0331524				

Omitted for the purpose of this exam

Question 1

1.a) $H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$$t_{cal}(\hat{\beta}_3) = \frac{\hat{\beta}_3 - \beta_3}{se\,\hat{\beta}_3} = \frac{-0.00704 - 0}{0.0124122} = -0.5672$$

$t_{0.025, 2719} = 1.96$ ∴ Cannot reject H_0

1.b) Model 1.1

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

H_1 : Not all the slope cofficient are simultaneously 0

$$F\text{-test} : \frac{R^2}{k-1} \div \frac{1-R^2}{n-k}$$

$= 24.3153$

$F_{0.01, 8, 52919} = 3.25$ ∴ Reject H_0

Model 1.2

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

H_1 : Not all the slope cofficient are simultaneously 0

$$F\text{-test} : \frac{R^2}{k-1} \div \frac{1-R^2}{n-k}$$

$= 26.4588$

$F_{0.01, 8, 2716} = 2.74$ ∴ Reject H_0

1.C) $H_0 : \beta_7 = \beta_8 = \beta_9 = 0$

$H_1 :$ otherwise

$$F_{cal} = \frac{R^2_{\hat{u}_i^2} / (k)}{(1 - R^2_{\hat{u}_i^2}) / (n-k-1)} \sim F_{cri}$$

$= 28.7888$ $\therefore$ Reject H_0

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

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				
avgsen	-.0113177	.0122401				
tottime	.0120224	.0094352				
ptime86	-.0408417	.008812				
qemp86	-.0505398	.0144397				
inc86	-.0014887	.0003406				
black	.3265035	.0454156				
hispan	.1939144	.0397113				
_cons	.5686855	.0360461				

Omitted for the purpose of this exam

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.

$$\widehat{\ln wage_i} = \underset{\beta_1}{9.1748} + \underset{\beta_2}{0.587}\, civil_i - \underset{\beta_3}{0.0336}\, year_i + \underset{\beta_4}{0.0444}\, civil_i \cdot year_i + u_i$$

(0.0035) (0.0072) (0.005) (0.0102)

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 2

2.a)

#STEP 1

- For β_1: $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$
- For β_2: $H_0 : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$
- For β_3: $H_0 : \beta_3 = 0$, $H_1 : \beta_3 \neq 0$
- For β_4: $H_0 : \beta_4 = 0$, $H_1 : \beta_4 \neq 0$

#STEP 2

- $t_{cal}(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{se\,\hat{\beta}_1} = \frac{9.1748 - 0}{0.0035} = 2,621.3714$ #
- $t_{cal}(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = \frac{0.587 - 0}{0.0072} = 81.5278$ #
- $t_{cal}(\hat{\beta}_3) = \frac{\hat{\beta}_3 - \beta_3}{se\,\hat{\beta}_3} = \frac{0.0336 - 0}{0.005} = 6.72$ #
- $t_{cal}(\hat{\beta}_4) = \frac{\hat{\beta}_4 - \beta_4}{se\,\hat{\beta}_4} = \frac{0.0444 - 0}{0.0102} = 4.3529$ #

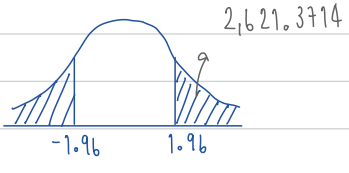
#STEP 3

- $n = 97878 - 4 = 97,874$
- $\alpha = 0.05$

$t_{cri} = t_{0.025,\,97,874} = 1.96$

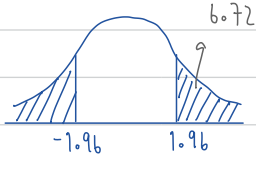
#STEP 4

- For β_1:



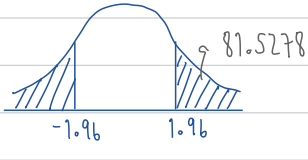
∴ reject H_0

- For β_3:



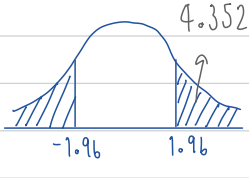
∴ reject H_0

- For β_2:



∴ reject H_0

- For β_4:



∴ reject H_0

2.b) $\log$ of β_2 = $\log$ of .587

$.587 = e$

$= 1.7986$ #

∴ The average monthy wage of civil servant and state employee is 1.7986 Baht higher than for non civil servants and state employee.

2.C) $\ln(\widehat{wage}_i | year_i = 0) = \hat{\beta}_1 + \hat{\beta}_2 civil_i$

$= 9.1748 + 0.587\, civil_i$ #

$\ln(\widehat{wage}_i | year_i = 1) = (\hat{\beta}_1 + \hat{\beta}_3) + (\hat{\beta}_2 + \hat{\beta}_4) civil_i$

$= 9.1412 + 0.6314\, civil_i$ #

year 2019 : civil servant state employee : $e^{9.7618} = 17,357.8477$

otherwise : $e^{9.1748} = 9650.8377$

year 2020 : civil servant state employee : $e^{9.7726} = 17,546.3284$

otherwise : $e^{9.1412} = 9,331.9568$

∴ During pandemic, the civil servant and state employee monthly wage, on average, has increase by 188.4807 Baht #

∴ During pandemic, the private firm or freelance, etc monthly wage, on average, has decrease by 318.8809 Baht #

2.d) Control group is better off because monthly wage raise by 188.4807 baht. But, the treatment group worse off, as their monthly drop by 318.8809 baht. The treatment group is worse off because the closed business are mostly private firm and mostly are servic sector.

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	Number of obs	=	30
Residual	2642.24237	26	101.624706	F(3, 26)	=	16.47
				Prob > F	=	0.0000
Total	7662.88705	29	264.237485	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				
h2s	3.915242	1.153106				
lactic	18.80235	8.342614				
_cons	-34.13491	15.67628				

Omitted for the purpose of this exam

	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

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

where Inf_t = inflation rates (%)

$unem_t$ = unemployment rates (%)

Table 4.1

Source	SS	df	MS	Number of obs	=	59
				F(1, 57)	=	3.85
Model	32.3284496	1	32.3284496	Prob > F	=	0.0545
Residual	478.096987	57	8.38766644	R-squared	=	0.0633
				Adj R-squared	=	0.0469
Total	510.425437	58	8.80043856	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

Question 3

3.a) #STEP 1

- For β_1: $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$
- For β_2: $H_0 : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$
- For β_3: $H_0 : \beta_3 = 0$, $H_1 : \beta_3 \neq 0$

#STEP 2

- $t_{cal}(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{se\,\hat{\beta}_1} = \frac{1.5386 - 0}{3.0005} = 0.5128$ #
- $t_{cal}(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = \frac{3.9152 - 0}{1.1537} = 3.3954$ #
- $t_{cal}(\hat{\beta}_3) = \frac{\hat{\beta}_3 - \beta_3}{se\,\hat{\beta}_3} = \frac{18.8024 - 0}{8.326} = 2.2538$ #

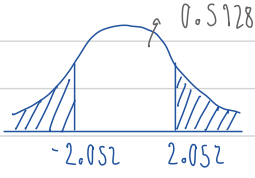
#STEP 3

- $n = 30 - 3 = 27$
- $\alpha = 0.05$

$t_{cri} = t_{0.025,\,27} = 2.052$

#STEP 4

- For β_1:

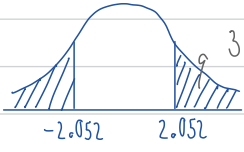


∴ cannot reject H_0

- For β_3:

∴ reject H_0

- For β_2:



∴ reject H_0

- P-test : $\frac{R^2}{k-1} \div \frac{1-R^2}{n-K} \sim F_{0.05,3,26} = 2.98$

$\Rightarrow 16.4687$

∴ Reject H_0

3.b) BLUE or best linear unbiased estimator

- It is linear function of data and have minimum of variance among linear unbiased estimator

- The OLS estimator still remain the property of BLUE although there is multicollinearity.

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?

****

Question 4

4.a) This mean that, in period of 1948 to 2006, if the unemployment rate is 0, inflation rate on average, is about 1.01 percent. The slope coefficient is 0.5054734, it can indicate that in the period of 1948 to 2006, inflation rate on average, increase by around 0.5 percent when unemployment rates go up 0.5 percent.

4.b) H_0: Homoscedasticity in the model

H_1: Heteroscedasticity in the model

According to table 4.1, white's general heteroscedasticity with a test statistic of 1.0266 and the critical chi-squared distribution with df = 1 and $\alpha = 0.05$

$$\begin{aligned} \chi^2_{cri} &= \chi^2_{4,k-1} \\ &= \chi^2_{(0.05,1)} \\ &= 0.0039 \end{aligned}$$

So, this have sufficient evidence to reject the null hypothesis, there is heteroscedasticity in residual of this regression model

4.c) The OLS estimator does not still retain the property of BLUE due to the fact that they are still unbiasedness and consistency