

## Instructions

6304641050 Phupiruth

- (1) Please read the instruction carefully.
- (2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at least 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  $\alpha = 0.05$ )

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

**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  $\alpha = 0.05$ )

**Estimate the model (1.1) reports in the Table 1.1**

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

**Table 1.1**

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

$$narr86_i = \beta_1 + \beta_2pcnv_i + \beta_3avgsen_i + \beta_4tottime_i + \beta_5ptime86_i + \beta_6qemp86_i + \beta_4inc86_i + \beta_5black_i + \beta_6hispan_i + u_i \tag{1.2}$$

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
Model	145.390104	8	18.173763	F(8, 2716)	=	26.47
Residual	1864.95705	2,716	.686655763	Prob > F	=	0.0000
				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.

$$\ln \widehat{wage}_i = 9.1748 + 0.587 civil_i - 0.0336 year_i + 0.0444 civil_i \cdot year_i + u_i$$

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

- 2.a)** Test all the parameters individually if each of them is significantly different from zero or not.
- 2.b)** How much on average does a civil servant and state employee earn 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**

$$Taste = \beta_0 + \beta_1 acetic + \beta_2 h2s + \beta_3 lactic + u \tag{3.1}$$

- 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	Number of obs	=	30
Model	<b>5020.64468</b>	<b>3</b>	<b>1673.54823</b>	F(3, 26)	=	<b>16.47</b>
Residual	<b>2642.24237</b>	<b>26</b>	<b>101.624706</b>	Prob > F	=	<b>0.0000</b>
Total	<b>7662.88705</b>	<b>29</b>	<b>264.237485</b>	R-squared	=	<b>0.6552</b>
				Adj R-squared	=	<b>0.6154</b>
				Root MSE	=	<b>10.081</b>

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

	acetic	h2s	lactic	Variable	VIF	1/VIF
acetic	<b>1.0000</b>			lactic	<b>1.83</b>	<b>0.546648</b>
h2s	<b>0.2700</b>	<b>1.0000</b>		h2s	<b>1.72</b>	<b>0.582609</b>
lactic	<b>0.3607</b>	<b>0.6448</b>	<b>1.0000</b>	acetic	<b>1.15</b>	<b>0.867477</b>
				Mean VIF	<b>1.57</b>	

**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

$$Inf_t = \beta_1 + \beta_2 unem_t + u_t \tag{4.1}$$

where  $Inf_t$  = inflation rates (%)

$unem_t$  = unemployment rates (%)

**Table 4.1**

Source	SS	df	MS	Number of obs	=	59
Model	32.3284496	1	32.3284496	F(1, 57)	=	3.85
Residual	478.096987	57	8.38766644	Prob > F	=	0.0545
Total	510.425437	58	8.80043856	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			
_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?

\*\*\*\*\*

# "Question 1."

$$1. a) \hat{narr}_{86} = 0.71 - 0.15 pcnv - 0.007 avg_{sen} + 0.012 tottime - 0.04 ptime_{86} - 0.1qemp_{86}$$

• Interpret : if  $avg_{sen}$  increases by 1 unit,  $narr_{86}$  will decrease by 0.007 unit

$$\text{Step 1 : } H_0 = \beta_{avg_{sen}} = 0$$

$$H_a = \beta_{avg_{sen}} \neq 0$$

$$\text{Step 2 : For } \beta_{avg_{sen}} \quad t_{cal}(\beta_{avg_{sen}}) = \frac{\hat{\beta}_{avg_{sen}} - \beta_{avg_{sen}}}{\text{se } \hat{\beta}_{avg_{sen}}} = \frac{-0.0070487 - 0}{0.0124122} = -0.56788 //$$

$$\text{Step 3 : } \alpha = 0.05$$

$$t_{lower} = -1.96$$

$$t_{upper} = 1.96$$

Step 4 : concluding the test result



∴ Cannot reject  $H_0$

So, average sentence has no impact on  $narr_{86}$

1. b) Model (1.1) • F-test

$$\text{Step 1 : } H_0 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

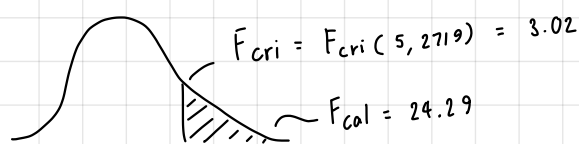
$$H_a = \text{otherwise}$$

$$\text{Step 2 : Calculate } F(5, 2719) = 24.29$$

$$\text{Step 3 : } \alpha = 0.01$$

$$F_{upper, \alpha}(5, 2719) = 3.02$$

Step 4 : Concluding the test : reject  $H_0$  | So, at least one  $\beta$  is not equal to 0.



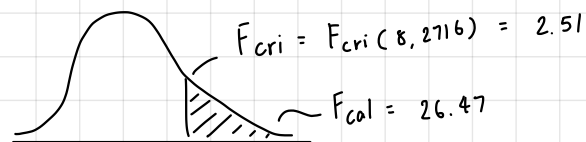
Model (1.2) • F-test

$$\text{Step 1 : } H_0 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 \\ H_a = \text{otherwise}$$

$$\text{Step 2 : } \text{Calculate } F(8, 2716) = 26.47$$

$$\text{Step 3 : } \alpha = 0.01 \\ F_{upper, \alpha}(8, 2716) = 2.51$$

Step 4 : Concluding the test : reject  $H_0$  | So, at least one  $\beta$  is not equal to 0.



1.C)

Step 1 : state hypothesis

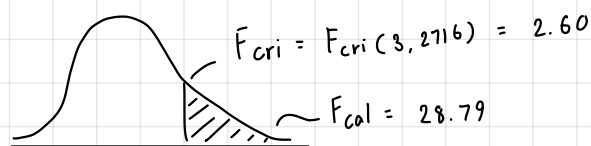
$$H_0 : \beta_7 = \beta_8 = \beta_9 = 0 \\ H_a : \text{otherwise}$$

Step 2 :

$$F_{cal} = \frac{R_{UR}^2 - R_R^2 / m}{1 - R_{UR}^2 / (n - k_{UR})} = \frac{0.0723 - 0.0428 / 3}{0.0428 / 2716} = 28.79$$

$$\text{Step 3 : } \alpha = 0.05 \\ F_{upper, \alpha}(3, 2716) = 2.60$$

Step 4 : Concluding the test



$\therefore$  reject  $H_0$   
since  $F_{cal} > F_{table}$   
 $28.79 > 2.60$

So, unconstrained model is better  
since we can reject  $H_0$

## "Question 2."

2.a) use t-test

$$\text{cons, } H_0: \begin{matrix} \beta_1 = 0 \\ \beta_1 \neq 0 \end{matrix} \quad \left| \quad t\text{-cal}(\beta_1) = \frac{\hat{\beta}_1 - \beta_1}{\text{se } \hat{\beta}_1} = \frac{9.1748 - 0}{0.0035} = 2,621.37 //$$

$$\text{civil, } H_0: \begin{matrix} \beta_2 = 0 \\ \beta_2 \neq 0 \end{matrix} \quad \left| \quad t\text{-cal}(\beta_2) = \frac{\hat{\beta}_2 - \beta_2}{\text{se } \hat{\beta}_2} = \frac{0.587 - 0}{0.0072} = 81.5278 //$$

$$\text{year, } H_0: \begin{matrix} \beta_3 = 0 \\ \beta_3 \neq 0 \end{matrix} \quad \left| \quad t\text{-cal}(\beta_3) = \frac{\hat{\beta}_3 - \beta_3}{\text{se } \hat{\beta}_3} = \frac{-0.0336 - 0}{0.005} = -6.72 //$$

$$\text{civ year, } H_0: \begin{matrix} \beta_4 = 0 \\ \beta_4 \neq 0 \end{matrix} \quad \left| \quad t\text{-cal}(\beta_4) = \frac{\hat{\beta}_4 - \beta_4}{\text{se } \hat{\beta}_4} = \frac{0.0444 - 0}{0.0102} = 4.3529 //$$

- The critical value for the degrees of freedom  $(n-k) = 97,874$  ( $\alpha = 0.05$ ) is  $\pm 1.96$ .
- All the  $t_{\text{cal}}$  exceeds the critical value. Therefore, we can reject all null hypothesis.

2.b) On average civil servants earns more than the other group by  $100 \times (e^{\hat{\beta}_2} - 1) = 79.86 \%$ .

2.c) Overall wage drops by  $100 \times (e^{\hat{\beta}_3} - 1) = 3.30$  percent and note that it is a negative coefficient.

2. D) When we consider the interaction term, we can see that once the civil servant variable is interacted with year (2020), there is a positive effect and significantly different from 0.

and it will display like this

$$\ln \widehat{\text{wage}} = 9.1748 + 0.587(1) - 0.0336(1) + 0.0444(1) \cdot (1) \text{ and the others are}$$

$$\ln \widehat{\text{wage}} = 9.1748 + 0.587(0) - 0.0336(1) + 0.0444(0) \cdot (1)$$

- In conclusion, the civil servant group or control group is better-off by  $100 \times (e^{\hat{\beta}_3 + \hat{\beta}_4} - 1) = 1.09\% \uparrow$ . Meanwhile, others are worse-off by  $100 \times (e^{\hat{\beta}_3} - 1) = 3.30\% \downarrow$  of overall wage.
- With economic reason, civil servants' wage didn't drop during the pandemic, while others work less due to lower wage.

### "Question 3"

3. a) There is no evidence of multicollinearity. First, correlation matrix must higher than 0.8, so, it will cause multicollinearity. However, in this case all of the coefficients are lower than 0.8. Secondly, VIF must be more than 10 to cause multicollinearity but according to the table all of the coefficients are less than 10. Then, there is no multicollinearity in this case.

3. b) BLUE is Best Linear Unbiased Estimator, in the other word is lowest variance possible. They are also unbiased when there are more samples than the parameters. By the way, multicollinearity doesn't cause BLUE but, multicollinearity causes misleading conclusion of hypothesis testing.

### "Question 4"

4. a) • Intercept ( $\hat{\beta}_1$ )

For  $\hat{\beta}_1$ , when unemployment rate = 0, inflation rate is 1.01 on average.

• Slope ( $\hat{\beta}_2$ )

For  $\hat{\beta}_2$ , when unemployment rate increases by 1 percent, inflation rate increases by 0.5% on average

4. b) If we use White's test, according to the given information,  $LM_{cal}$  for  $\chi^2_{k-1}$  is 1.0266. then, we have to find  $\chi^2_{\alpha}$  when  $\alpha = 0.05$  and the answer is 3.84146.

In conclusion  $LM_{cal} < \chi^2_{\alpha}$ , then, we can't reject the null hypothesis.

4. c) BLUE property is not violated because we can't reject the null hypothesis.