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

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

Table 1.1

Source	SS	df	MS		
Model	85.9532425 ESS	5	17.1906485	Number of obs	= 2,725
Residual	1924.39391 RSS	2,719	.707757967	F(5, 2719)	= 24.29
				Prob > F	= 0.0000
				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)
๒	$pcnv_i$	= the proportion of prior arrests that led to a conviction
๓	$avgsen_i$	= the average sentence served from prior convictions (in months)
๔	$tottime_i$	= months spent in prison since age 18 prior to 1986
๕	$ptime86_i$	= months spent in prison in 1986
๖	$qemp86_i$	= the number of quarters that the man was legally employed in 1986
๗	$inc86_i$	= legal income, 1986, (hundred dollars)
๘	$black_i$	= 1 if black ethnic background
๙	$hispan_i$	= 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

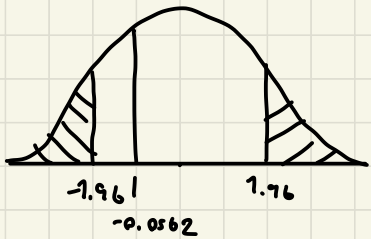
narr86	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
pcnv	-.1332344	.0403502			
avgsen	-.0113177	.0122401			
tottime	.0120224	.0094352			
ptime86	-.0408417	.008812			
qemp86	-.0505398	.0144397			
inc86	-.0014887	.0003406			
black	.3265035	.0454156			
hispan	.1939144	.0397113			
_cons	.5686855	.0360461			

Omitted for the purpose of this exam

1a.) -0.00704 : If the average sentence served from prior conviction raise by 1 month, the number of arrest in 1986 will decrease by 0.00704.

$H_0: \beta_3 = 0$
$H_1: \beta_3 \neq 0$

$$t = \frac{\hat{\beta}_3 - \beta_3}{SE(\beta_3)} = \frac{-0.00704 - 0}{0.0124122} = -0.5672$$



$t_{0.025, 2719} = 1.96$

∴ Cannot reject H_0, the average sentence served from prior conviction has no impact on the number of arrest in current year at 95% confidence level

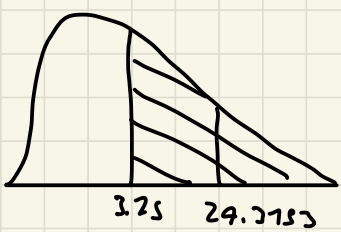
1b.) Model 1.1

$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$
H_1: Not all the slope coefficient are simultaneously zero

F test

$$F_{test} = \frac{R^2}{k-1} \div \frac{1-R^2}{n-k} = \frac{0.0423}{5} \div \frac{0.9577}{2719} = 24.3153$$

$F_{0.01, 5, 2719} = 3.25$



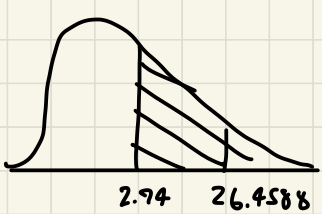
∴ Reject H_0,
There is enough evidence to say that not all the slope coefficient are simultaneously zero at 99% confidence level

Model 1.2

$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9$
H_1: Not all the slope coefficient are simultaneously zero

$$F_{test} = \frac{R^2}{k-1} \div \frac{1-R^2}{n-k} = \frac{0.0723}{8} \div \frac{0.9277}{2716} = 26.4588$$

$F_{0.01, 8, 2716} = 2.74$



∴ Reject H_0,
There is enough evidence to say that not all the slope coefficient are simultaneously zero

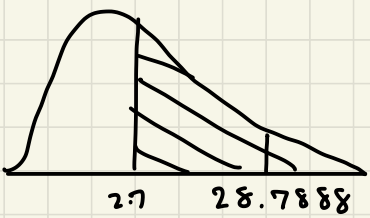
1C.) $H_0: \beta_7 = \beta_8 = \beta_9 = 0$

H_1: otherwise

$$F_{cal} = \frac{R^2_R - R^2_{UR}}{m} \div \frac{1 - R^2_{UR}}{n - k_{UR}} \sim F_{cri} = F_{\alpha, m, n-k_{UR}}$$

$$= \frac{0.0428 - 0.0723}{3} \div \frac{0.9277}{2916} \sim F_{0.05, 3, 2916} = 2.7$$

$$= 28.7888$$



∴ Reject H_0, legal income and ethnic background has an impact on the number of arrest in current year (1986) at 95% confidence level

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.

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

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.

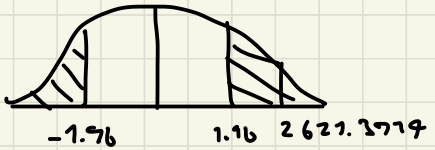
2a.) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

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

$$= \frac{9.1748 - 0}{0.0035}$$

$$= 2621.3719$$



$\therefore$ Reject H_0, β_1 is statistically significant

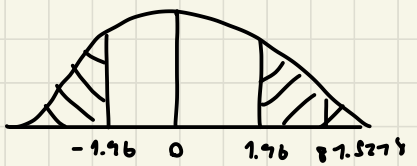
$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

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

$$= \frac{0.589 - 0}{0.0072}$$

$$= 81.5278$$



$\therefore$ Reject H_0, type of employment is statistically significant.

$t_{crit} = t_{0.025, 99.874} = 1.96$

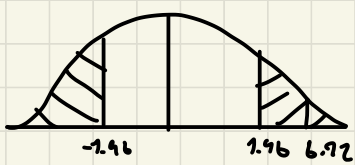
$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$$t\text{-test} = \frac{\hat{\beta}_3 - \beta_3}{SE(\beta_3)}$$

$$= \frac{0.0336 - 0}{0.005}$$

$$= 6.72$$



$\therefore$ Reject H_0, The pandemic effect is statistically significant.

$t_{crit} = 1.96$

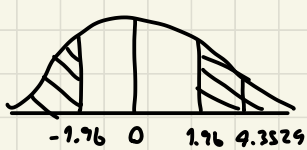
$H_0: \beta_4 = 0$

$H_1: \beta \neq 0$

$$t\text{-test} = \frac{\hat{\beta}_4 - \beta_4}{SE(\beta_4)}$$

$$= \frac{0.0444 - 0}{0.0102}$$

$$= 4.3529$$



$\therefore$ Reject H_0, interaction between type of employment and pandemic effect is statistically significant

2b.) Anti log of β_2 = anti log of 0.587

$= e^{0.587}$

$= 1.7986$

On average monthly wage of civil servant and state employee is 1.7986 baht more than who is not civil servant and state employee disgurding the year

2c.) year 2019 : year; =0

$= 9.1748 + 0.589\,civil_i - 0.0326(0) + 0.0444\,civil_i(0)$

$= 9.1748 + 0.589\,civil_i$

year 2020 : year; = 1

$= 9.1748 + 0.589\,civil_i - 0.0326(1) + 0.0444\,civil_i(1)$

$= 9.1412 + 0.6334\,civil_i$

	2019	2020
civil servant state employee	(1,0) 9.7618	(1,1) 9.7726
otherwise	(0,0) 9.1748	(0,1) 9.1412

$e^{9.7618} = 17,357.8477$

$e^{9.7726} = 17,546.3284$

$e^{9.1748} = 9650.8377$

$e^{9.1412} = 9331.9568$

During pandemic

∴ On average, the civil servant and state employee monthly wage has increase by 188.4807 baht.

∴ On average, the private firm, freelance, etc. monthly wage has decrease by 318.8809 baht.

2d.) Control group is better off, as their monthly wage raise by 188.4807 baht (17546.3284 - 17357.8477). However, the treatment group worse off, as their monthly wage decrease by 318.8809 baht (9331.9568 - 9650.8377). The treatment group is worse off as the closed business are mostly private firm and mostly are service sector. Moreover, reduce time for open for example, convicience store, and department store, as a result more employee are being lay off and their income decrease.

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
				R-squared	= 0.6552
				Adj R-squared	= 0.6154
Total	7662.88705	29	264.237485	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.) $H_0: \beta_1 = 0$

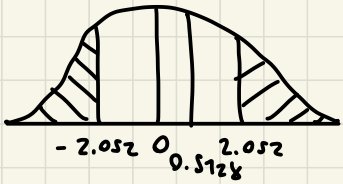
$H_1: \beta_1 \neq 0$

$t\text{-test} = \dfrac{\hat{\beta}_1 - \beta_1}{SE(\beta_1)}$

$= \dfrac{1.5356 - 0}{3.0005}$

$= 0.5128$

$t_{0.025,29} = 2.052$



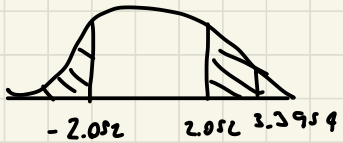
∴ Cannot reject H_0, the concentration of acetic is statistically insignificant

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$t\text{-test} = \dfrac{\hat{\beta}_2 - \beta_2}{SE(\beta_2)}$

$= \dfrac{3.9152}{1.1531} = 3.3954$



∴ Reject H_0, the concentration of hydrogen sulfide is statistically significant

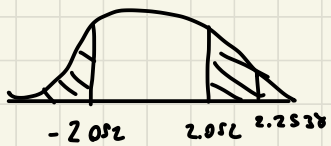
$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$t\text{-test} = \dfrac{\hat{\beta}_3 - \beta_3}{SE(\beta_3)}$

$= \dfrac{18.8024}{8.3426}$

$= 2.2538$



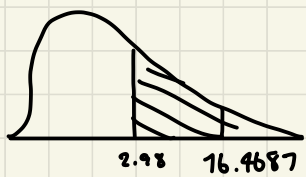
∴ Reject H_0, Lactic is statistically significant.

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1: Not all the slope coefficient are simultaneously zero

$F\text{-test}: \dfrac{R^2}{k-1} \div \dfrac{1-R^2}{n-k} \sim F_{0.05,3,26} = 2.98$

$= \dfrac{0.6552}{3} \div \dfrac{0.3448}{26} = 16.4687$



∴ Reject H_0, There is enough evidence to say that not all the slope coefficient are simultaneously zero

∴ $\hat{\beta}_2$ is insignificant at $\alpha = 0.05$, therefore the partial slope coefficient are individually statistically insignificant but the overall F test is significant at the same level of significant. According to above reason, there is evidence to conclude that there is multicollinearity of the regression

3b.) "BLUE" or best linear unbiased estimator

It is linear function of data and has minimum variance among linear unbiased estimator.

The OLS estimator still retain the property of BLUE although there is multicollinearity in the data.

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

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?

****

4a.) Intercept is 1.010847, it mean that, in the period 1948-2006, if unemployment rate is about zero, inflation rate on average, is about 1.01 percent
The slope coefficient is 0.5054734, We can indipret that in the period 1948-2006 inflation rate on average increase by about 0.5 percent when unemployment rates raise 0.5 percent

4b.) H_0: there is Homoscedasticity in the model.
H_1: there is Heteroscedasticity in the model.
From the table 4.1, White's General Heteroscedasticity with a test statistic of 1.0266 and the critical chi-squared distribution with df = 1 at $\alpha = 0.05$ is 0.0039

$$\chi^2_{cri} = \chi^2_{\alpha, k-1}$$
$$= \chi^2_{(0.05, 1)}$$
$$= 0.0039$$

We have sufficient evidence to reject the null hypothesis, there is Heteroscedasticity in the residual of this regression model.

4c.) Under Heteroscedasticity, the OLS estimator does not still retain the property of BLUE due to the fact that they are still unbiasedness and consistency but these estimator are no longer efficient or minimum variance.