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

Instructions

Rachata Anuwanuwong
6304641216

(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

n

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

R^2 model

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

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		
Model	145.390104	8	18.173763	Prob > F	0.0000
Residual	1864.95705	2,716	.686655763	R-squared	0.0723
Total	2010.34716	2,724	.738012906	Adj R-squared	0.0696
				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

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)

Model: $\widehat{narr86}_i = 0.71 - 0.15\,pcnv_i - 0.07\,avgsen_i + 0.01\,tottime_i - 0.04\,ptime86_i - 0.10\,qemp86_i$ ($\hat{\beta}_3$)

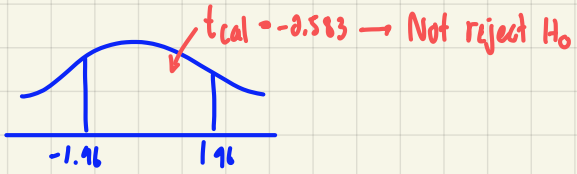
$avgsen_i$ = average sentence served from prior convictions (in months)

As $avgsen_i$ increase by one unit, narr86 decrease around 0.007 units

$SE_{\hat{\beta}_3} = 0.012$

Test ① $H_0: \beta_3 = 0$ $H_1: \beta_3 \neq 0$ ② $\alpha = 0.05$ ③ $t_{cal} = \frac{\hat{\beta}_3 - \beta_3}{SE_{\hat{\beta}_3}} \overset{0?}{=} \frac{-0.007 - 0}{0.012} = -0.583$

④ $t_{cri} = t_{n-k,0.05} = t_{2725-6,0.05} = t_{\infty,0.05} = \pm 1.96$

⑤ 

⑥ Conclusion: Not reject H_0
So, we can't sure that $\beta_3 \neq 0$ from 95 out of 100 times

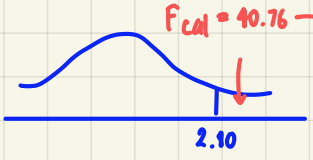
1.b) What is the overall significance of the regression from Model (1.1) and Model (1.2)? What test do you use? (Use α = 0.01)

model 1: $R^2 = 0.0421$, $k = 6$
model 2: $R^2 = 0.0723$, $k = 4$

Model 1

① $H_0: \beta_2 = \beta_3 = \ldots = \beta_6 = 0$ H_1: Otherwise ② $\alpha = 0.01$ ③ $F_{cal} = \frac{R^2/k-1}{1-R^2/n-k} = \frac{0.0421/5}{(1-0.0421)/(2725-6)} = \frac{0.00836}{0.00021} = 40.7619$

④ $F_{cri} = F_{n-k,0.01} = F_{2725-6,0.01} = F_{\infty,0.01} = 2.10$

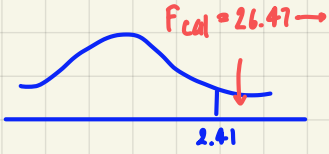
⑤ 

⑥ Conclusion: Reject H_0
So, we sure that $\beta_2 = \beta_2 = \ldots = \beta_6$ are not zero from 99 out of 100 times

Model

① $H_0: \beta_2 = \beta_3 = \ldots = \beta_9 = 0$ H_1: Otherwise ② $\alpha = 0.01$ ③ $F_{cal} = \frac{R^2/k-1}{1-R^2/n-k} = \frac{0.0723/9-1}{(1-0.0723)/(2725-9)} = \frac{0.0090}{0.00034} = 26.47$

④ $F_{cri} = F_{n-k,0.01} = F_{2725-9,0.01} = F_{\infty,0.01} = 2.41$

⑤ 

⑥ Conclusion: Reject H_0
So, we sure that $\beta_2 = \beta_2 = \ldots = \beta_9$ are not zero from 99 out of 100 times

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)

$black_i$ $inc86_i$

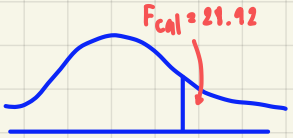
① H_0: $black_i$ & $inc86_i$ have no marginal contribution
H_1: Otherwise

② $\alpha = 0.05$

model 1: $R^2 = 0.0421$, $k = 6$
model 2: $R^2 = 0.0723$, $k = 9$

③ $F_{cal} = \frac{R^2_{new} - R^2_{old} / additional\ k}{(1 - R^2_{new})/n - k_{new}} = \frac{0.0723 - 0.0421/9-6}{(1-0.0723)/2725-9} = \frac{0.0243/3}{0.9277/2719} = \frac{0.00913}{0.00034} = 28.92$

④ $F_{cri} = F_{\alpha(additional\ k, n-k_{new})} = F_{0.05(3,2725-9)} = F_{0.05(3,\infty)} = 2.6$

⑤ 

⑥ Reject H_0
So, we can sure that $black_i$ and $inc86_i$ have marginal contribution to the model at $\alpha = 0.05$

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.

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

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.

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

$$\widehat{\ln wage_i} = \underset{\hat{\beta}_1}{9.1748} + \underset{\hat{\beta}_2}{0.587}\, civil_i - \underset{\hat{\beta}_3}{0.0336}\, year_i + \underset{\hat{\beta}_4}{0.0444}\, civil_i \cdot year_i + u_i$$
$$se \rightarrow \quad (0.0035) \quad (0.0072) \quad (0.005) \quad (0.0102)$$

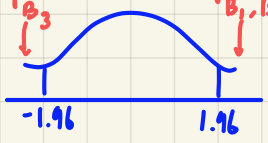
① $H_0: \beta_k = 0 \quad ; k = 1,2,3,4$
$H_1: \beta_k \neq 0 \quad ; k = 1,2,3,4$

② $\alpha = 0.05$

③
$$t_{cal}(\beta_1) = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{9.1748 - 0}{0.0035} = 2,621.37$$
$$t_{cal}(\beta_2) = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.587 - 0}{0.0072} = 81.53$$
$$t_{cal}(\beta_3) = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)} = \frac{-0.0336 - 0}{0.005} = -6.72$$
$$t_{cal}(\beta_4) = \frac{\hat{\beta}_4 - \beta_4}{se(\hat{\beta}_4)} = \frac{0.0444 - 0}{0.0102} = 4.35$$

④ $t_{cri} = t_{n-k} = t_{97176-4} = t_{\infty, 0.05} = \pm 1.96$

⑤ t_{β_3}, $t_{\beta_1, \beta_2, \beta_4}$ → all β lie on RR → Reject H_0



⑥ Conclusion : Reject H_0 for all cases
So, each β_k are not zero

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

log - lin model

$$\widehat{\ln wage_i} = 9.1748 + \underset{\hat{\beta}_2}{0.587}\, civil_i - 0.0336\, year_i + 0.0444\, civil_i \cdot year_i + u_i$$
$$(0.0035) \quad (0.0072) \quad (0.005) \quad (0.0102)$$

As the change of civil servant and state employee (β_2) is positive that means they got more wage than others
For the percentage change is $100 \times (e^{\hat{\beta}_2} - 1) = 100 \times (e^{0.587} - 1) = 100 \times (1.7986 - 1) = 79.86\%$.

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

To consider the effect on pandemic, we need to consider at year component by comparing 2nd quater of 2019 and 2020
Since β_3 is negative that means the pandemic makes wage fall by $100 \times (e^{\hat{\beta}_3} - 1) = 100 \times (1.0342 - 1) = 3.42\%$.

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.

Comparing both group during pandemic
Control group (civil): $\widehat{\ln wage_i} = 9.1748 + 0.587(1) - 0.0336(1) + 0.0444(1)(1) = 9.7726$
Treatment group : $\widehat{\ln wage_i} = 9.1748 + 0.587(0) - 0.0336(1) + 0.0444(0)(1) = 9.1412$
$\} \; 0.6314 = \hat{\beta}_2 + \hat{\beta}_4$

The civil group got more wage than others because civil gain more by $\hat{\beta}_2$ when comparing with others
and also gain more by $\hat{\beta}_4$ which is civil during covid-19, eventhough their wage drop by $\hat{\beta}_3$, they gain more by $\hat{\beta}_2 + \hat{\beta}_4$
On the other hand, treatment group got only negative effect by $\hat{\beta}_3$ which reduce their wage
Change of control group : $100 \times (e^{\hat{\beta}_2 - \hat{\beta}_3 + \hat{\beta}_4} - 1) = 100 \times (e^{0.5978} - 1) = 81.82\%$ → civil group better-off by 81.82%
Change of treatment group : $100 \times (e^{\hat{\beta}_3} - 1) = 100 \times (e^{0.0336} - 1) = 3.42\%$ → treatment group worst-off by 3.42%
This makes economic sense that the civil servant's wage because they still working but otherwise may work-less or get lay-off

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
Total	7662.88705	29	264.237485	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]
$\hat{\beta}_2$ acetic	1.538645	3.000501			
$\hat{\beta}_3$ h2s	3.915242	1.153106			
$\hat{\beta}_4$ lactic	18.80235	8.342614			
$\hat{\beta}_1$ _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.

Multicollinearity → same value or scale of beta

Detecting multicollinearity
- Rule of thumb → $|R^2| < 0.8$ ✓
- VIF & TOL → VIF < 10 ✓
  - TOL close to 1 not 0 ✓
- Check all $B_k = 0$?

Test using t-test; against zero

① $H_0: B_k = 0$; $k = 1,2,3,4$
$H_1: B_k \neq 0$; $k = 1,2,3,4$

② $\alpha = 0.05$

③ $t_{cal} = \frac{\hat{B}_k - B_k}{SE_{\hat{B}_k}}$ (B_k = 0)

$t_{cal}(\hat{B}_1) = \frac{-34.13441 - 0}{15.67629} = -2.18$

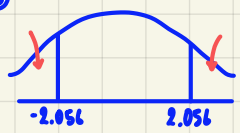
$t_{cal}(\hat{B}_2) = \frac{1.538645 - 0}{3.000501} = 0.51$

$t_{cal}(\hat{B}_3) = \frac{3.916242 - 0}{1.153181} = 3.40$

$t_{cal}(\hat{B}_4) = \frac{18.80235 - 0}{8.342614} = 22.54$

④ $t_{cri} = t_{n-k, 0.05} = t_{26, 0.05} = \pm 2.056$

⑤



each $\hat{B}_k$ fall on rejection region → reject H_0

⑥ Conclusion: Reject H_0

So, we can sure at 95% confidence that each B_k ; $k = 1,2,3,4$ are not zero

From all testing above, we doesn't see the multicollinearity problem

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?

high corr → high r^2 → high var → high SE. → low t → less chance to conclude or to reject H_0 ($P > |t|$)

BLUE is best linear unbias estimater
Multicollinearity problem is high correlation in the model that causeing the r^2 will be high following with the variance of the model also high affacting to the high SE., finally the conclusion may be wrong by the high error because of low t-value
As the variance is high the B-Best (lowest variance property) may violate

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]
$\hat{\beta}_2$ unem	.5054734	.2574699	Omitted for the purpose of this exam		
$\hat{\beta}_1$ _cons	1.010847	1.491583			

White's general test statistic: 1.0266 Chi-sq (2) $= LM_{cal}$ for χ^2_{k-1}

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

chi2(1) = 1.12

4.a) Interpret the intercept and slope coefficients.

Intercept : $\hat{B}_1 = 1.01$ means at unemployment rate = 0%, the inflation rate is around 1.01%

Slope : $\hat{B}_2 = 0.51$ means when unemployment rate increase by 1%, the inflation rate will increase around 0.51%

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)

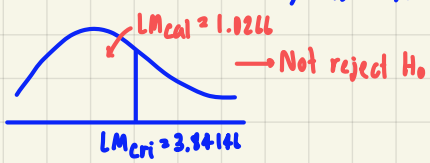
Heteroscedasticity → high value → high variance
high variance, data not fits line

White's test given: $LM_{cal} = 1.0266$

① H_0: No Heteroscedasticity
H_1: Heteroscedasticity present

② Find $LM_{cri} = \chi^2_{k-1} = \chi^2_{2-1,(0.05)} = \chi^2_{1,(0.05)} = 3.84146$

③



④ Conclusion: Can't sure that no Heteroscedasticity present

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 not sure that Heteroscedasticity present (from b.)
So we can sure that BLUE is not violate

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?

****