Paper & pencil question.

Vipharath Charoensukanun
6104640559

1. Which of the following can cause the usual OLS t statistics to be invalid (that is, not to have t distributions under H_0)?

   i. Heteroskedasticity.

   ii. A sample correlation coefficient of .95 between two independent variables that are in the model.

   iii. Omitting an important explanatory variable.

1. (i) and (ii)

(i) Heteroskedasticity is violated MLR5 - homoskedasticity, and it make OLS t statistic to be invalid.

(iii) An explanatory variable which could affect the data can thereby affect the results of regression analysis will definitely cause t statistic to be invalid if that explanatory variable is omitted. This violate MLR4 $E(u|x) = 0$.

2. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

   i. In terms of the model parameters, state the null hypothesis that, after controlling for *sales* and *roe*, *ros* has no effect on CEO salary. State the alternative that better stock market performance increases a CEO's salary.

$H_0: \beta_3 = 0$

$H_a: \beta_3 > 0$

   ii. Using the data in CEOSAL1, the following equation was obtained by OLS:

$$\widehat{\log(salary)} = \underset{(.32)}{4.32} + \underset{(.035)}{.280} \log(sales) + \underset{(.0041)}{.0174}\, roe + \underset{(.00054)}{.00024}\, ros$$

$$n = 209,\ R^2 = .283.$$

By what percentage is *salary* predicted to increase if *ros* increases by 50 points? Does *ros* have a practically large effect on *salary*?

$$\beta_3 = \frac{d \log(salary)}{d x_3}$$

$$\beta_3 = \frac{\frac{1}{salary}\Delta salary}{\Delta x_3}$$

$$100 \times .00024 = \frac{\frac{1}{salary}\Delta salary}{50} \times 100$$

$$100 \times .00024(50) = \frac{1}{salary}\Delta salary \times 100$$

$$1.2 = \%\Delta\, salary$$

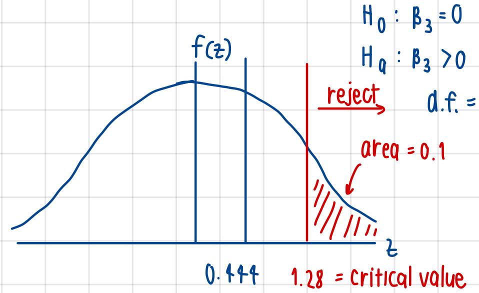
holding other factors constant, if ros increases by 50 points log(salary) will increase by 1.2%.

   iii. Test the null hypothesis that *ros* has no effect on *salary* against the alternative that *ros* has a positive effect. Carry out the test at the 10% significance level.

$H_0: \beta_3 = 0$

$H_a: \beta_3 > 0$

d.f. = 209 - 3 - 1 = 205 > 30 use z-table



$$z = \frac{\hat{\beta}_3 - \beta_3}{s.e.\,\hat{\beta}_3} = \frac{.00024}{.00054} = 0.444 < 1.28$$

So, we fail to reject H_0 at 10% significant level.

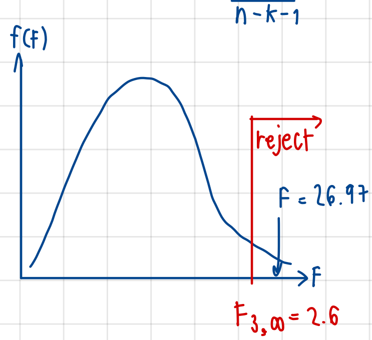
   iv. Would you include *ros* in a final model explaining CEO compensation in terms of firm performance? Explain.

r : $\widehat{\log(salary)} = 4.32$ → $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

ur : $\widehat{\log(salary)} = 4.32 + 0.28 \log(sales) + 0.0174(roe) + 0.00024(ros)$

H_a : otherwise

$$F = \frac{\frac{R^2}{k}}{\frac{(1-R^2)}{n-k-1}} = \frac{\frac{0.283}{3}}{\frac{0.717}{205}} = 26.97$$



Since, F = 26.97 > 2.6 , we rejected H_0 and conclude that firm performance have joint effect in explaning CEO compensation at 5% significant level

Computer questions.

C6. Use the data in WAGE2 for this exercise.

i. Consider the standard wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

State the null hypothesis that another year of general workforce experience has the same effect on $\log(wage)$ as another year of tenure with the current employer.

$H_0: \beta_2 = \beta_3$ or $H_0: \beta_2 - \beta_3 = 0$

$H_a: \beta_2 \neq \beta_3$ $\quad H_a: \beta_2 - \beta_3 \neq 0$

ii. Test the null hypothesis in part (i) against a two-sided alternative, at the 5% significance level, by constructing a 95% confidence interval. What do you conclude?

let $\hat{\theta}_1 = \hat{\beta}_2 - \hat{\beta}_3$

So, $H_0: \hat{\theta}_1 = 0$ $\qquad t = \frac{\hat{\theta}_1}{s.e.\hat{\theta}_1}$

$H_a: \hat{\theta}_1 \neq 0$

rearrange $\hat{\theta}_1 = \hat{\beta}_2 - \hat{\beta}_3$, get $\hat{\beta}_2 = \hat{\theta}_1 + \hat{\beta}_3$ and substitute it into regression model

$$\log(wage) = \beta_0 + \beta_1 educ + (\hat{\theta}_1 + \hat{\beta}_3) exper + \beta_3 tenure + u$$
$$= \beta_0 + \beta_1 educ + \hat{\theta}_1 exper + \hat{\beta}_3 exper + \beta_3 tenure + u$$
$$\log(wage) = \beta_0 + \beta_1 educ + \hat{\theta}_1 exper + \hat{\beta}_3 (exper + tenure) + u$$

```
. generate expertenure = exper+tenure

.
end of do-file

. do "C:\Users\610464~1\AppData\Local\Temp\STD1224_000000.tmp"

. regress lwage educ exper expertenure
```

Source	SS	df	MS
Model	25.6953242	3	8.56510806
Residual	139.960959	931	.150334005
Total	165.656283	934	.177362188

Number of obs	=	935
F(3, 931)	=	56.97
Prob > F	=	0.0000
R-squared	=	0.1551
Adj R-squared	=	0.1524
Root MSE	=	.38773

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0748638	.0065124	11.50	0.000	.062083	.0876446
exper	.0019537	.0047434	0.41	0.681	-.0073554	.0112627
expertenure	.0133748	.0025872	5.17	0.000	.0082974	.0184522
_cons	5.496696	.1105282	49.73	0.000	5.279782	5.713609

$t_{critical} = 0.41 < 1.96$

Which means we fail to reject H_0 at 95% confidence interval.
So, both of variables has not have the same impact.

C8. The data set 401KSUBS contains information on net financial wealth (*nettfa*), age of the survey respondent (*age*), annual family income (*inc*), family size (*fsize*), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question, use only the data for single-person households (so $fsize = 1$).

i. How many single-person households are there in the data set?

```
. tab fsize
```

family size	Freq.	Percent	Cum.
1	2,017	21.75	21.75
2	2,199	23.71	45.46
3	1,829	19.72	65.18
4	1,990	21.46	86.63
5	816	8.80	95.43
6	268	2.89	98.32
7	95	1.02	99.34
8	38	0.41	99.75
9	7	0.08	99.83
10	7	0.08	99.90
11	3	0.03	99.94
12	4	0.04	99.98
13	2	0.02	100.00
Total	9,275	100.00	

```
.
end of do-file
```

There are 2,017 single-person household in the data.

ii. Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u,$$

and report the results using the usual format. Be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?

```
. keep if fsize == 1
(7,258 observations deleted)

.
end of do-file

. do "C:\Users\610464~1\AppData\Local\Temp\STD17e4_000000.tmp"

. regress nettfa inc age
```

Source	SS	df	MS		
Model	544916.989	2	272458.495	Number of obs =	2,017
Residual	4021048.06	2,014	1996.54819	F(2, 2014) =	136.46
				Prob > F =	0.0000
				R-squared =	0.1193
				Adj R-squared =	0.1185
Total	4565965.05	2,016	2264.86361	Root MSE =	44.683

nettfa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
inc	.7993167	.0597307	13.38	0.000	.6821762	.9164572
age	.8426563	.0920169	9.16	0.000	.6621982	1.023115
_cons	-43.03981	4.080393	-10.55	0.000	-51.04204	-35.03758

```
.
end of do-file
```

Intercept = -43.03981 ;
slope on inc = .7993167; slope on age = .8426563
If income increases by 1 unit or $1,000 then nettfa increased by .7993167 or $799.3167.
If age increases by 1 unit or 1 year the nettfa increased by .8426563 or $843

iii. Does the intercept from the regression in part (ii) have an interesting meaning? Explain.

```
. summarize nettfa
```

Variable	Obs	Mean	Std. Dev.	Min	Max
nettfa	2,017	13.59498	47.59058	-143.5	1134.098

```
.
end of do-file

. do "C:\Users\610464~1\AppData\Local\Temp\STD17e4_000000.tmp"

. count if nettfa < 0
  524

.
end of do-file

. do "C:\Users\610464~1\AppData\Local\Temp\STD17e4_000000.tmp"

. summarize inc age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
inc	2,017	29.44618	16.67356	10.008	143.067
age	2,017	39.27814	10.82328	25	64

```
.
end of do-file
```

Intercept gives us the predicted nettfa when inc=0 and age=0 but the lowest inc is $10,000 and the lowest age is 25.

iv. Find the *p*-value for the test $H_0: \beta_2 = 1$ against $H_1: \beta_2 < 1$. Do you reject H_0 at the 1% significance level?

```
. lincom age - 1

 ( 1)  age = 1
```

nettfa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-.1573437	.0920169	-1.71	0.087	-.3378018	.0231145

One-sided p-value = .087/2 = .0435 which is greater than .01. So, we accepted H₀ at 1% significant level.

v. If you do a simple regression of *nettfa* on *inc*, is the estimated coefficient on *inc* much different from the estimate in part (ii)? Why or why not?

```
. regress nettfa inc
```

Source	SS	df	MS		
Model	377482.064	1	377482.064	Number of obs =	2,017
Residual	4188482.98	2,015	2078.6516	F(1, 2015) =	181.60
				Prob > F =	0.0000
				R-squared =	0.0827
				Adj R-squared =	0.0822
Total	4565965.05	2,016	2264.86361	Root MSE =	45.592

nettfa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
inc	.8206815	.0609	13.48	0.000	.7012479	.940115
_cons	-10.57095	2.060678	-5.13	0.000	-14.61223	-6.529671

```
.
end of do-file

. do "C:\Users\610464~1\AppData\Local\Temp\STD17e4_000000.tmp"

. correlate inc age
(obs=2,017)
```

	inc	age
inc	1.0000	
age	0.0391	1.0000

From (ii), estimated coefficient on inc is .799 versus the new one which is .821 . They are not very different because income and age are not quite correlated.

C1. The following model can be used to study whether campaign expenditures affect election outcomes:

$$voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u,$$

where *voteA* is the percentage of the vote received by Candidate A, *expendA* and *expendB* are campaign expenditures by Candidates A and B, and *prtystrA* is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

i. What is the interpretation of β_1 ?

1% increases in campaign expenditure by candidate A, increase 1% of the vote received by candidate A.

ii. In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

$H_0: \beta_1 = -\beta_2$ or $H_0: \beta_1 + \beta_2 = 0$

$H_a: \beta_1 \neq -\beta_2$ $\quad$ $H_a: \beta_1 + \beta_2 \neq 0$

iii. Estimate the given model using the data in VOTE1 and report the results in usual form. Do A's expenditures affect the outcome? What about B's expenditures? Can you use these results to test the hypothesis in part (ii)?

```
. regress voteA lexpendA lexpendB prtystrA

      Source |       SS           df       MS      Number of obs   =       173
-------------+----------------------------------   F(3, 169)       =    215.23
       Model |  38405.1096         3  12801.7032   Prob > F        =    0.0000
    Residual |  10052.1389       169   59.480112   R-squared       =    0.7926
-------------+----------------------------------   Adj R-squared   =    0.7889
       Total |  48457.2486       172  281.728189   Root MSE        =    7.7123

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   6.083316     .38215    15.92   0.000     5.328914    6.837719
    lexpendB |  -6.615417   .3788203   -17.46   0.000    -7.363246   -5.867588
    prtystrA |   .1519574   .0620181     2.45   0.015     .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------

.
```

A's expenditure positively affect the outcome.
B's expenditure negatively affect the outcome.
No, the data can not be use to test the hypothesis in part (ii) Because $\beta_1 + \beta_2$ is not provided.

iv. Estimate a model that directly gives the t statistic for testing the hypothesis in part (ii). What do you conclude? (Use a two-sided alternative.)

Let's $\hat{\theta}_1 = \hat{\beta}_1 + \hat{\beta}_2$ $\quad$ $H_0: \hat{\theta}_1 = 0$ $\quad$ $t = \dfrac{\hat{\theta}_1}{s.e.\,\hat{\theta}_1}$

$H_a: \hat{\theta}_1 \neq 0$

Now, $\hat{\beta}_1 = \hat{\theta}_1 - \hat{\beta}_2$ $\quad$ $\beta_1 = \theta_1 - \beta_2$ substitute into regression.

$voteA = \beta_0 + (\theta_1 - \beta_2)\log(expend\,A) + \beta_2 \log(expend\,B) + \beta_3\, prtystrA + u.$

$= \beta_0 + \theta_1 \log(expend\,A) - \beta_2 \log(expend\,A) + \beta_2 \log(expend\,B) + \beta_3\, prtystrA + u.$

$= \beta_0 + \theta_1 \log(expend\,A) - \beta_2 (\log(expend\,A) + \log(expend\,B)) + \beta_3\, prtystrA + u.$

$voteA = \beta_0 + \theta_1 \log(expend\,A) + \beta_2 (\log(expend\,B) - \log(expend\,A)) + \beta_3\, prtystrA + u.$

```
. generate lexpendAlexpendB = lexpendB - lexpendA

. regress voteA lexpendA lexpendAlexpendB prtystrA

      Source |       SS           df       MS      Number of obs   =       173
-------------+----------------------------------   F(3, 169)       =    215.23
       Model |  38405.1097         3  12801.7032   Prob > F        =    0.0000
    Residual |  10052.1388       169  59.4801115   R-squared       =    0.7926
-------------+----------------------------------   Adj R-squared   =    0.7889
       Total |  48457.2486       172  281.728189   Root MSE        =    7.7123

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   -.532101   .5330858    -1.00   0.320    -1.584466    .5202638
lexpendAlexpe~B | -6.615417 .3788203  -17.46   0.000    -7.363246   -5.867588
    prtystrA |   .1519574   .0620181     2.45   0.015     .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------
```

$t_{critical} = -1.00 > -1.96$

Therefore, we fail to reject $H_0: \beta_1 = -\beta_2$ at 5% significant level. 1% increased in A's expenditure is offset by 1% increased in B's expenditure.