Athicha K.
6104640062

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.

Omitting an important explanatory variable can cause the usual OLS t statistics to be invalid. Because as we ignore the important explanatory variable, the coefficient of regression will be less than it should be which can result in the fact that the estimation become less precise leading to the unbiasedness.

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.

After the controlling for sales and roe,
we can test the null hypothesis such that

$H_o : \beta 3 = 0$

$H_a : \beta 3 \neq 0$

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

$$\underbrace{\widehat{\log(salary)}}^{y} = 4.32 + \overset{\beta 1}{.280} \overset{x1}{\log(sales)} + \overset{\beta 2}{.0174}\, \overset{x2}{roe} + \overset{\beta 3}{.00024}\, \overset{x3}{ros}$$
$$(.32)\quad (.035)\qquad (.0041)\qquad (.00054)$$
$$n = 209, R^2 = .283.$$

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

$$\beta 3 = \frac{d\overset{y}{\log(salary)}}{d(ros)\, x3} = \frac{d\log(y)}{d(x3)} = \frac{\frac{1}{y}dy}{d(x3)}$$

$$100\beta 3 = \frac{100\frac{1}{y}\Delta y}{\Delta x3}$$

$$100\beta 3 = \frac{\%\Delta y}{\Delta x3}$$

$$100(0.00024) = \frac{\%\Delta(salary)}{\Delta ros}$$

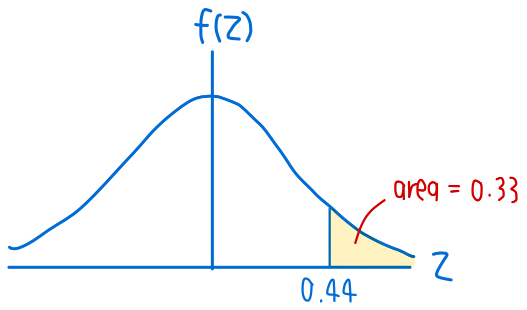
$$100(0.00024) = \frac{\%\Delta(salary)}{(50-0)}$$

$$\%\Delta(salary) = 1.2\%$$

Salary is predicted to increase by 1.2% if ros increases by 50 points.
However, ros does not have a practically large effect on salary.

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.

$$\left.\begin{array}{l} H0: \beta_3 = 0 \\ Ha: \beta_3 > 0 \end{array}\right\} \text{z-table} \qquad t = \frac{\hat{\beta}_3 - \beta_3}{s.e.(\hat{\beta}_3)} = \frac{0.00024 - 0}{0.00054} = 0.44$$



P-value is 0.33 > 0.1 (Significant level)

So, we do not reject Ho at 10% significant level.
Therefore, ros has no impact on salary.

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

No, as the return on firm's stock has no impacts on CEO compensation. So, the increase in ros will not have any impact on CEO salary which indicates firm performance. Therefore, we will not include ros in a final model explaining CEO compensation.

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 ?

The slope of β1 means of the campaign expenditures by Candidate A increase by $1, the percentage of the vote received by Candidate A is expected to change by the value of B1.

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.

$H_o: \beta 2 = -\beta 1$, $H_o: \beta 2+\beta 1 = 0$

$H_a: \beta 2 \neq -\beta 1$, $H_a: \beta 2+\beta 1 \neq 0$

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
Model	38405.1096	3	12801.7032
Residual	10052.1389	169	59.480112
Total	48457.2486	172	281.728189

Number of obs	=	173
F(3, 169)	=	215.23
Prob > F	=	0.0000
R-squared	=	0.7926
Adj R-squared	=	0.7889
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

The usual form:

$$VoteA = 45.0789 + 6.0833\log(expendA) - 6.6154\log(expendB) + 0.1520(prtystrA)$$

As A's expenditure and B's expenditure have a different value and sign, the change in both A and Bexpenditure will result in different outcome. Therefore, we can use these results to test the hypothesis in part (ii).

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.)

$H_0: \beta_2 = -\beta_1 \quad , \quad H_0: \beta_2 + \beta_1 = 0 \quad \longrightarrow \quad t = \dfrac{(\hat{\beta}_2 + \hat{\beta}_1) - 0}{s.e.(\hat{\beta}_2 + \hat{\beta}_1)}$

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

Let $\hat{\theta}_1 = \hat{\beta}_2 + \hat{\beta}_1 \longrightarrow H_0: \theta_1 = 0 \quad , \quad t = \dfrac{\hat{\theta}_1 - 0}{s.e.(\hat{\theta}_1)}$

$H_a: \theta_1 \neq 0$

if rearrange $\hat{\theta}_1 = \hat{\beta}_2 + \hat{\beta}_1$, we have $\hat{\beta}_1 = \hat{\theta}_1 - \hat{\beta}_2$ or $\beta_1 = \theta_1 - \beta_2$

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

$$voteA = \beta_0 + (\theta_1 - \beta_2)\log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u$$

$$voteA = \beta_0 + \theta_1 \log(expendA) - \beta_2 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u$$

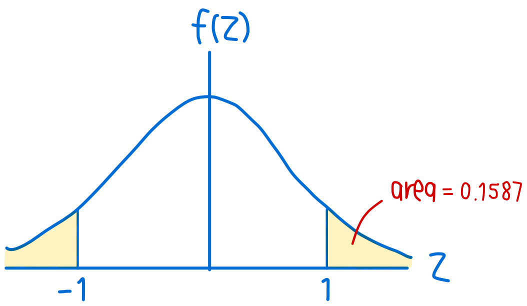
$$voteA = \beta_0 + \theta_1 \log(expendA) + \beta_2 [\underbrace{\log(expendB) - \log(expendA)}_{expend\_3}] + \beta_3 prtystrA + u$$

Then do the regression again.

```
. regress voteA lexpendA expend_3 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]
-------------+----------------------------------------------------------------
θ   lexpendA |   -.532101   .5330858    -1.00   0.320    -1.584466    .5202638
    expend_3 |  -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
------------------------------------------------------------------------------
```



$$t = \frac{\hat{\theta}_1 - 0}{s.e.(\hat{\theta}_1)} = \frac{-0.5321 - 0}{0.5331} = -0.998 \approx -1$$

Assuming that we are testing the null hypothesis at 1% significant level.

P-value = 0.1587 > 0.005 (Significant level)

So, we will not reject the Ho as the P-value is greater than the significant level. That means a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

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.

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

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

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?

$\left.\begin{array}{l} H_0: \beta_2 = \beta_3 \longrightarrow H_0: \beta_2 - \beta_3 = 0 \\ H_a: \beta_2 \neq \beta_3 \longrightarrow H_a: \beta_2 - \beta_3 \neq 0 \end{array}\right\}$ two-tailed test $\longrightarrow t = \dfrac{(\hat\beta_2 - \hat\beta_3) - 0}{s.e.(\hat\beta_2 - \hat\beta_3)}$

Let $\hat\theta_1 = \hat\beta_2 - \hat\beta_3 \longrightarrow H_0: \theta_1 = 0$, $H_a: \theta_1 \neq 0$, $t = \dfrac{\hat\theta_1 - 0}{s.e.(\hat\theta_1)}$

if rearrange $\hat\theta_1 = \hat\beta_2 - \hat\beta_3$, we have $\hat\beta_2 = \hat\theta_1 + \hat\beta_3$ or $\beta_2 = \theta_1 + \beta_3$

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

$$\log(wage) = \beta_0 + \beta_1(educ) + (\theta_1 + \beta_3)(exper) + \beta_3(tenure) + u$$

$$\log(wage) = \beta_0 + \beta_1(educ) + \theta_1(exper) + \beta_3(exper) + \beta_3(tenure) + u$$

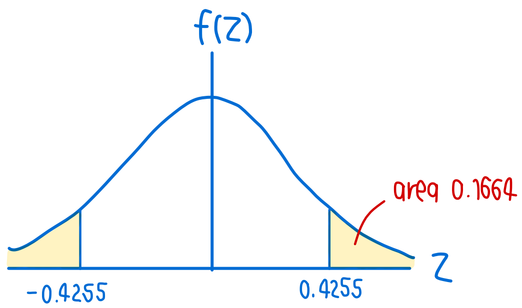
$$\log(wage) = \beta_0 + \beta_1(educ) + \theta_1(exper) + \beta_3\underbrace{(exper + tenure)}_{exper\_2} + u$$

Then do the regression again.

```
. regress lwage educ exper exper_2
```

Source	SS	df	MS			
Model	25.6953242	3	8.56510806	Number of obs	=	935
Residual	139.960959	931	.150334005	F(3, 931)	=	56.97
				Prob > F	=	0.0000
				R-squared	=	0.1551
				Adj R-squared	=	0.1524
Total	165.656283	934	.177362188	Root MSE	=	.38773

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0748638	.0065124	11.50	0.000	.062083	.0876446
θ1 exper	.0019537	.0047434	0.41	0.681	-.0073554	.0112627
exper+tenure exper_2	.0133748	.0025872	5.17	0.000	.0082974	.0184522
_cons	5.496696	.1105282	49.73	0.000	5.279782	5.713609



$t = \dfrac{\hat\theta_1 - 0}{s.e.(\hat\theta_1)} = \dfrac{0.002 - 0}{0.0047} = 0.4255$ | P-value is 0.1664 > 0.025 (significant level)

So, we do not reject Ho at 5% significant level as P-value is greater than the significant level. That means another year of general workforce experience has the same effect on log(wage) as another year of tenure with the current employer.

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?

	fsize
2014	1
2015	1
2016	1
2017	1
2018	2
2019	2
2020	2

From the data given, the single-person households which has the family size of 1 have all together 2017 households.

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?

```
. regress nettfa inc age if fsize == 1

      Source |       SS           df       MS      Number of obs   =     2,017
-------------+----------------------------------   F(2, 2014)      =    136.46
       Model |  544916.989         2  272458.495   Prob > F        =    0.0000
    Residual |  4021048.06     2,014  1996.54819   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
------------------------------------------------------------------------------
```

The usual form:

Nettfa = -43.0398 + 0.7993(inc) + 0.8427(age)

The slope for annual family income of 0.7993 means if annual family income increases by $1000, the net financial wealth is expected to increase by $799.32.

The slope for age survey respondents of 0.8427 means if the age of survey respondents increases by 1 year, the net financial wealth is expected to increase by $842.7.

Therefore, we can conclude that age of survey respondents has a greater impact on net financial wealth that annual family income.

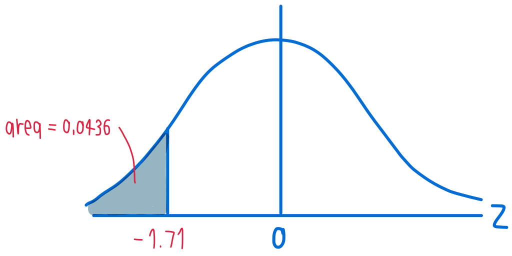
iii. Does the intercept from the regression in part (ii) have an interesting meaning? Explain.

The slope of -43.0398 means if the age of the survey respondents and annual family income is zero, the net financial wealth is expected to have the loss of $43039.8

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?

$$\left.\begin{array}{l} Ho : \beta 2 = 1 \\ Ha : \beta 2 < 1 \end{array}\right\} \text{one-tailed test , Z-table}$$

$$t = \frac{\hat{\beta}2 - \beta 2}{s.e.(\hat{\beta}2)} = \frac{0.843 - 1}{0.092} = -1.71$$



P-value for the hypothesis test is 0.0436. > 0.01 (significant level)

So, as the P-value for test hypothesis is greater than the significant level (0.0436 > 0.01), we not reject Ho 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 inc nettfa if fsize == 1
```

Source	SS	df	MS			
Model	46335.1731	1	46335.1731	Number of obs	=	2,017
Residual	514127.962	2,015	255.150354	F(1, 2015)	=	181.60
				Prob > F	=	0.0000
				R-squared	=	0.0827
				Adj R-squared	=	0.0822
Total	560463.135	2,016	278.007508	Root MSE	=	15.973

inc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nettfa	.100737	.0074754	13.48	0.000	.0860768	.1153973
_cons	28.07666	.3699027	75.90	0.000	27.35123	28.80209

The intercept of the simple regression of net financial wealth and annual income is quite different in both value and sign. In this case, even the net financial wealth is zero, the annual family income is still expected to be positive with $28,076.66

In terms of slope, the slope for net financial wealth of 0.1007 means if the net financial wealth increases by $1000, the annual family income is expected to increase by $100.7

To conclude, the slope is changed a lot as it reduces from $799.32 (inc on nettfa) to 100.7 (nettfa on inc). So, there is a greater impact of annual family income on net financial wealth (inc on nettfa) than that of net financial wealth on annual family income (nettfa on inc).