Thanyalak Samreeloy
6104641284

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.

Heteroskedasticity is the assumption for MLR that used to find the OLS estimation. Meaning that $\beta_0, \beta_1, \ldots \beta_k$ will not be efficient without $Var(U \mid x_1, x_2, \ldots, x_k) = \sigma^2$. Moreover, the explanatory variables should be includet to make the OLS unbiased in order to seperate them with error term. If our OLS has sample correlation ceefficient of 0.95 between 2 independent variables, It is OK because it doesn't violate any MLR assumption.

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.

   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*?

   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.

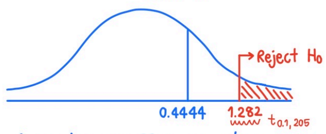
   iv. Would you include *ros* in a final model explaining CEO compensation in terms of firm performance? Explain.

(i) $H_0 : \beta_3 = 0$
$H_a : \beta_3 > 0$

(ii) Effect on salary = 50(0.00024) = 0.012
∴ Ros doesn't has a large effect on salary which only 1.2%

(iii) $\alpha = 0.1$, $n = 209$, $d.f. = 209-3-1 = 205$

$$z = \frac{\hat{\beta}_3 - \beta_3}{se.\hat{\beta}_3} = \frac{0.00024 - 0}{0.00054}$$
$$= 0.4444$$



∴ ros has no effect on salary

(iv) No, ros has no effect on salary at 0.1 level of significant. If we add ros, it will rise R^2, but it worsen the variance. We only want thing that explain the model.

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?

   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.

   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)?

   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.)

(i) β_1 can be interpreted that 1% increase in a campain expenditure by candidates A will increase β_1 percentage of the vote recieved by candidate A

(ii) $H_0: \beta_2 = -\beta_1$
$H_a: \beta_2 \neq -\beta_1$

```
. 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

(iii) $\widehat{voteA} = 45.08 + 6.083 \log(expendA) - 6.615 \log(expendB) + 0.152\, prtystrA$

A's expenditure and B's expenditure affect the outcome which is the percentage of the vote recieved by candidate A. While A's expenditure has a positive impact, B's expenditure a negative impact on vote A.
We cannot use this result to test the hypothesis in part (ii) because we need the standard error of $\beta_1 + \beta_2$

```
. gen lexpendA_lexpendB = lexpendA - lexpendB

. regress voteA lexpendA lexpendA_lexpendB 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
lexpendA_lexpendB	6.615417	.3788203	17.46	0.000	5.867588	7.363246
prtystrA	.1519574	.0620181	2.45	0.015	.0295274	.2743873
_cons	45.07893	3.926305	11.48	0.000	37.32801	52.82985

(iv) $\theta_1 = \beta_1 + \beta_2$
$\beta_1 = \theta_1 - \beta_2$
$Vote\,A = \beta_0 + (\theta_1 - \beta_2)\log(expend\,A) + \beta(expend\,B) + \beta_3\, prtystr\,A + u$
$Vote\,A = \beta_0 + \theta_1 \log(expend\,A) + \beta_2(\log(expend\,B) - \log(expend\,A)) + \beta_3\, prtystr\,A + u$

$\hat{\theta}_1 = -0.532$
since p-value of $\hat{\theta}_1$, is equal to 0.320 which is more than 0.05 which is 5% level of significant, we accept Ho at 5% signifinicant level. So, 1% increase in A's expenditure is not offset by one percent increas in B's expenditure

Thanyalak Samreeloy
6104641284

C6. Use the data in WAGE2 for this exercise.

i. Consider the standard wage equation

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

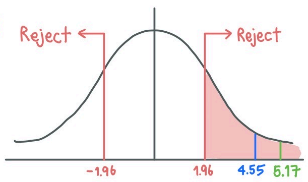
State the null hypothesis that another year of general workforce experience has the same effect on $\log(wage)$ as another year of tenure with the current employer.

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?

(i) $H_0: \beta_2 = \beta_3$
$H_a: \beta_2 \neq \beta_3$

```
. regress lwage educ exper tenure

      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(3, 931)       =     56.97
       Model |  25.6953242         3  8.56510806   Prob > F        =    0.0000
    Residual |  139.960959       931  .150334005   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
       exper |   .0153285   .0033696     4.55   0.000     .0087156    .0219413
      tenure |   .0133748   .0025872     5.17   0.000     .0082974    .0184522
       _cons |   5.496696   .1105282    49.73   0.000     5.279782    5.713609
------------------------------------------------------------------------------
```

(ii)



Since, 4.55 > 1.96 and 5.17 > 1.96 , we reject Ho at 5% level of significant level and conclude the exper and tenure have joint effect on wage

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?

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?

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

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?

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?

(i) There are 2,017 single-person houshold in the data set

```
. sum if fsize == 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       e401k |      2,017    .3604363    .4802461          0          1
         inc |      2,017    29.44618    16.67356     10.008    143.067
        marr |      2,017    .0183441    .1342256          0          1
        male |      2,017    .5418939    .4983654          0          1
         age |      2,017    39.27814    10.82328         25         64
-------------+---------------------------------------------------------
       fsize |      2,017           1           0          1          1
      nettfa |      2,017    13.59498    47.59058     -143.5   1134.098
       p401k |      2,017    .2429351    .4289625          0          1
        pira |      2,017    .2141795    .4103536          0          1
       incsq |      2,017    1144.947    1581.761   100.1601   20468.17
-------------+---------------------------------------------------------
       agesq |      2,017    1659.857    922.5799        625       4096
```

(ii) $\widehat{nettfa} = -43.04 + 0.799\,inc + 0.843\,age$
when income increase 1 unit, the net financial wealth will increase 0.799 unit. Also, the increasing in age 1 unit, the net financial wealth will rise 0.843 unit.

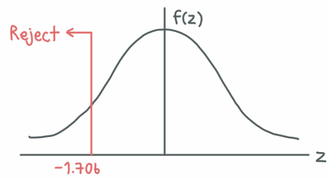
```
. regress nettfa inc age if fsize == 1

      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
------------------------------------------------------------------------------
```

(iii) The intercept can tell that when other factors are zero, the net financial wealth will be equal to −43.04

(iv) $H_0: \beta_2 = 1$
$H_a: \beta_2 < 1$

$$t = \frac{\hat{\beta}_2 - \beta_2}{s.e.\,\hat{\beta}_2} = \frac{0.843 - 1}{0.092} = -1.706$$

area = 0.5 − 0.4608 = 0.0392
Since p-value > 0.01 , we do not reject Ho



(v) The coef. in (v) is greater than (ii) because (ii) has more x variables (inc, age) than (v).
So, (ii) will share the impact on the explained variables (less value of coef.)

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

      Source |       SS           df       MS      Number of obs   =     2,017
-------------+----------------------------------   F(1, 2015)      =    181.60
       Model |  377482.064         1  377482.064   Prob > F        =    0.0000
    Residual |  4188482.98     2,015   2078.6516   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
------------------------------------------------------------------------------
```