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.

i) For the OLS disirbution to be vaild or BLUE, Assumption MLR5 : Homoskedasticity which make $\hat{\beta}_{OLS}$ the most efficient among all linear estimator.

Therefore heteroskedasticity make the usual OLS t statistics invalid.

ii) does not cause OLS t statistic to be invalid

iii) Omitting an important explanatory variable does not make the ols invalid but makes it bias

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*?

i) $H_0: \beta_3 = 0$ ; ros have no effect

$H_a: \beta_3 > 0$ ; ros have positive effect

ii) if ros increase by 50points, salary will increase by 0.012 or percentage increase is 1.2%.

ros does not have a particular large effect on salary. This can be seen when compared to log(sale) and roe. For example 50point increase in log(sale) and roe will result in salary to increase by 14 and 0.87 respectively.

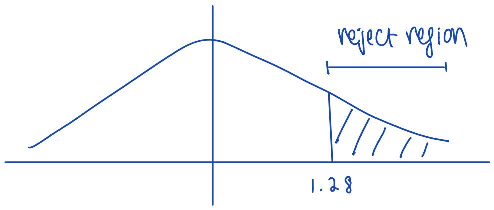
iii. Test the null hypothesis that *ros* has no effect on *salary* against the alternative that *ros* has a positive effect. Carry out the test at the 10% significance level.

$n-k-1 > 30 \rightarrow z$-table

$$\frac{\hat{\beta}_3 - 0}{\sigma} \sim 209-3-1$$

$$= \frac{0.00024}{0.00054}$$

$$= 0.\dot{4}$$




since $\frac{\hat{\beta}_3 - 0}{\sigma} < 1.28$, $H_0$ is accepted at 10% significant level; thus no effect

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

No. since as tested above, ros have no impact on the firm performance thus we would just be including irrelevant variable to 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 $\beta_1$ ?

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)?

i) Vote A will increase by $\beta_1$ if there is an increase in 1 unit of (log(expendA)

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

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

| 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 |

The usual form:

$\widehat{voteA} = 45.07893 + 6.083316 \log(expendA) - 6.615417 \log(expendB) + 0.1519574\, prtystrA$

For $\beta_1$; $H_0: \beta_1 = 0$ $H_a: \beta_1 \neq 0$ reject $H_0: \beta_1 = 0$ as pvalue < 5% thus affect

For $\beta_2$; $H_0: \beta_2 = 0$ $H_a: \beta_2 \neq 0$ reject $H_0: \beta_2 = 0$ as pvalue < 5% thus affect

we cannot use this result to test the hypotheses in part (ii) as it require $se(\beta_1 + \beta_2)$ and the t-statistic for $(\beta_1 + \beta_2)$ which this result lack.

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 $\theta = \beta_1 + \beta_2 \qquad \beta_1 = \theta - \beta_2$

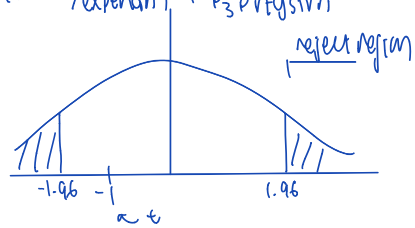
$$voteA = \beta_0 + \theta - \beta_2(\log(expendA)) + \beta_2 \log(expendB) + \beta_3 prtystrA$$

$$= \beta_0 + \theta \log(expendA) + \beta_2(\log(expendB) - \log(expendA)) + \beta_3 prtystrA$$

$$= \beta_0 + \theta \log(expendA) + \beta_2 \log(expendB/expendA) + \beta_3 prtystrA$$

| Source | SS | df | MS | | | |
|---|---|---|---|---|---|---|
| Model | 38405.1097 | 3 | 12801.7032 | Number of obs | = | 173 |
| Residual | 10052.1388 | 169 | 59.4801115 | F(3, 169) | = | 215.23 |
| | | | | Prob > F | = | 0.0000 |
| | | | | 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] |
|---|---|---|---|---|---|---|
| $\theta$ lexpendA | -.532101 | .5330858 | -1.00 | 0.320 | -1.584466 | .5202638 |
| expend3 | 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 |




thus does not reject $H_0: \beta_1 + \beta_2 = 0$

1% increase in A's expenditure is offset by a 1% increase of B's

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.

i) $H_0: \beta_2 = \beta_3 \;;\; \beta_2 - \beta_3 = 0$

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

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?

ii) let $\theta = \beta_2 - \beta_3 \longrightarrow \beta_2 = \theta + \beta_3$

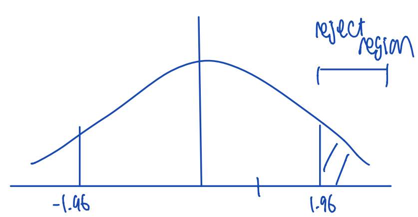
$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

$$= \beta_0 + \beta_1 educ + (\theta + \beta_3) exper + \beta_3 tenure + u$$

$$= \beta_0 + \beta_1 educ + \theta exper + \beta_3(exper + tenure) + u$$

| 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 |
| exper | .0019537 | .0047434 | 0.41 | 0.681 | -.0073554 | .0112627 |
| new | .0133748 | .0025872 | 5.17 | 0.000 | .0082974 | .0184522 |
| _cons | 5.496696 | .1105282 | 49.73 | 0.000 | 5.279782 | 5.713609 |




therefore do not reject $H_0: \beta_2 = \beta_3$, general workforce experience give the same effect 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)

```
count if fsize == 1
2,017
```

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?

. 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 |

usual form:

$nettfa = -43.03981 + 0.7993167(inc) + 0.8426563(age)$

For $\beta_1$: if inc increase by 1 unit, nettfa will increase by 0.799316

For $\beta_2$: if age increase by 1 unit, nettfa will increase by 0.8426563

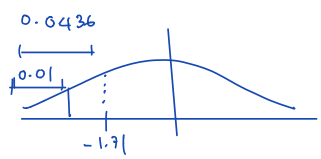
age have greater impact on nettfa.

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

if the age and annual income is equal to zero, net total financial asset will be -43,0398, meaning negative asset.

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? 0.01

$$z = \frac{0.8426563 - 1}{0.0920169} = -1.7099435 \approx 1.71$$




pvalue = 0.5 - 0.4564 = 0.0436

we do not reject $H_0: \beta_2 = 1$ as pvalue > 0.01.

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 | | | |
|---|---|---|---|---|---|---|
| | | | | Number of obs | = | 2,017 |
| | | | | F(1, 2015) | = | 181.60 |
| Model | 46335.1731 | 1 | 46335.1731 | Prob > F | = | 0.0000 |
| Residual | 514127.962 | 2,015 | 255.150354 | 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 |

if nettfa = 0, inc = \$28,0767, if nettfa increase by 1 unit, inc would increase by 0.100737 unit.

This is different because this is the effect of nettfa on inc while the previous regression is the effect of inc and age on nettfa.