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.

6104641268

OLS is method in statistic used for regression modeling. It is a type of linear least-squares method for estimating the unknow parameters in a linear regression model.

(i) Heteroskedasticity is correct

All random variables have same error term or finite variance or can be called "Homoskedasticity" as one of assumption, but if the data does not hold or "Heteroskedasticity" (the random variable does not have a common error term or varience) then the OLS t statistic will be invalid.

(iii) Omitting an important explanatory variable is also correct

Because the explanatory variable could affect the data and can affect the result of regression function analysis, so it will definitely cause the t-statistic to be invalid if the explanatory variable is omitted.

6104641268

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:

$$\widehat{\log(salary)} = 4.32 + .280 \log(sales) + .0174\, roe + .00024\, ros$$
$$\text{s.e.}\ (.32)\quad (.035)\ \text{s.e.}\quad (.0041)\ \text{s.e.}\quad (.00054)\ \text{s.e.}$$
$$n = 209,\ R^2 = .283.$$

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) $H_0: \beta_1 = 0,\ \beta_2 = 0,\ \beta_3 = 0$

$H_a: \beta_3 \neq 0$

ii) If ros increase by 50 point the percentage of salary will increase

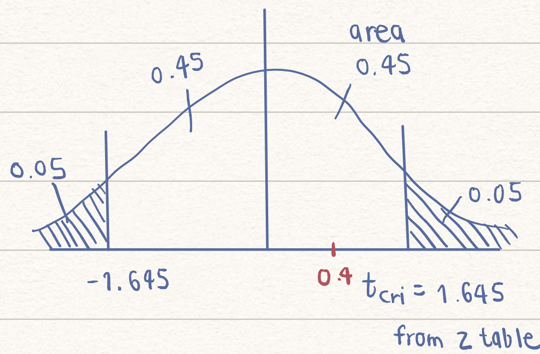
$\Delta \text{ros} = 0.00024 \times 50$

$= 0.012$

or equal to 1.2 % #

iii) $H_0: \beta_3 = 0$ $\quad$ d.f. $= n-k-1 = 209-3-1 = 205$

$H_a: \beta_3 \neq 0$ $\quad$ significant level $= 10\%$ or 0.1



$$t_{cal} = \frac{\hat{\beta}_3 - 0}{s.e.(\hat{\beta}_3)}$$

$$= \frac{0.00024}{0.00054} = 0.4$$

∴ We can't reject H_0 at the 10 % level of significant.

∴ β_1 has no statistically effect on salary at 10 % level of significant

iv) No, I would not include ros in final model since it has no impact on salaries of CEO as we prove from iii)

✓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 ?

(i) β_1 represents the positive relationship between the percentage of the vote received by Candidate A and the campaign expenditures 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.

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

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

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. $se(\hat{\beta}_x)$	t $\frac{\hat{\beta}-0}{s.e\hat{\beta}}$	P>|t|	[95% Conf.	Interval]
X_1 lexpendA	β_1 6.083316	.38215	15.92	0.000	5.328914	6.837719
X_2 lexpendB	β_2 -6.615417	.3788203	-17.46	0.000	-7.363246	-5.867588
X_3 prtystrA	β_3 .1519574	.0620181	2.45	0.015	.0295274	.2743873
intercept _cons	β_0 45.07893 (intercept)	3.926305	11.48	0.000	37.32801	52.82985

(iii) The A's expenditure positively affect the outcome by 6.083316.

The B's expenditure negatively affect the outcome by -6.615417

Therefore, the increase 1% in A's expenditure cannot be offset by the increase 1% in B's expenditure because $\beta_1 \neq \beta_2$

Regression Model

$$\text{Vote}A = \underset{(11.48)}{(45.1)} + \underset{(15.92)}{6.08} \log(\text{Expend}A) - \underset{(-17.46)}{6.62} \log(\text{Expend}B) + \underset{(2.45)}{0.152} (\text{Prtystr}A)$$

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.)

(iv) $H_0: \beta_1 - \beta_2 = 0$

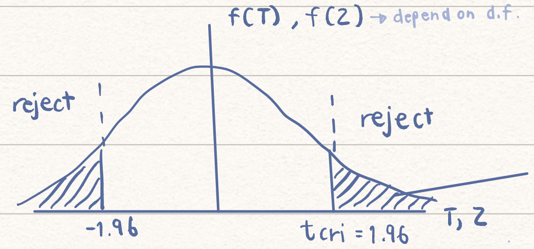
$H_a: \beta_1 - \beta_2 \neq 0$

$d.f = n - K - 1 = 173 - 3 - 1 = 169$

hypothesis value

$$t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{s.e.(\hat{\beta}_1 - \hat{\beta}_2)}$$

⇒ we compute this t-statistic and compare with critical value



$$\frac{\text{area of significant}}{2} = \frac{0.05}{2} = 0.025$$

∴ If t_{cal} is greater than t_{cri} (1.96) or lower than -1.96, then we will reject H_0 at the 5% level of significant

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?

```
. regress lwage educ exper tenure
```

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. ($s.e.\ \hat\beta$)	t ($\frac{\hat\beta - 0}{s.e.\ \hat\beta}$)	P>|t|	[95% Conf.	Interval]
x_1	educ	β_1 .0748638	.0065124	11.50	0.000	.062083	.0876446
x_2	exper	β_2 .0153285	.0033696	4.55	0.000	.0087156	.0219413
x_3	tenure	β_3 .0133748	.0025872	5.17	0.000	.0082974	.0184522
intercept	_cons	β_0 5.496696	.1105282	49.73	0.000	5.279782	5.713609

intercept

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

$H_a: \beta_2 \neq \beta_3 \ , \ \beta_1 - \beta_3 \neq 0$

ii) $H_0: \beta_2 = \beta_3 \ , \ \beta_1 - \beta_3 = 0$

$H_a: \beta_2 \neq \beta_3 \ , \ \beta_1 - \beta_3 \neq 0$

two-tailed test

$$t = \frac{(\hat\beta_2 - \hat\beta_3) - 0}{s.e.(\hat\beta_2 - \hat\beta_3)}$$

let $\hat\theta_1 = \hat\beta_2 - \hat\beta_3$; $H_0: \theta_1 = 0$, we have

$H_a: \theta_1 \neq 0$ $\qquad t = \dfrac{\hat\theta_1 - 0}{s.e.(\hat\theta_1)}$

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

$$\begin{aligned}
\log(wage) &= \beta_0 + \beta_1(educ) + \beta_2(exper) + \beta_3(tenture) + u \\
&= \beta_0 + \beta_1(educ) + (\theta_1 + \beta_3)(exper) + \beta_3(tenture) + u \\
&= \beta_0 + \beta_1(educ) + \theta_1(exper) + \beta_3(exper) + \beta_3(tenture) + u \\
&= \beta_0 + \beta_1(educ) + \theta_1(exper) + \beta_3\underbrace{(exper + tenure)}_{\text{new variable}} + u
\end{aligned}$$

Run the regression again

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?

. regress nettfa inc age if fsize ==1

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	$\hat{\beta}_1$ .7993167	.0597307	13.38	0.000	.6821762	.9164572
age	$\hat{\beta}_2$ .8426563	.0920169	9.16	0.000	.6621982	1.023115
_cons	$\hat{\beta}_0$ -43.03981	4.080393	-10.55	0.000	-51.04204	-35.03758

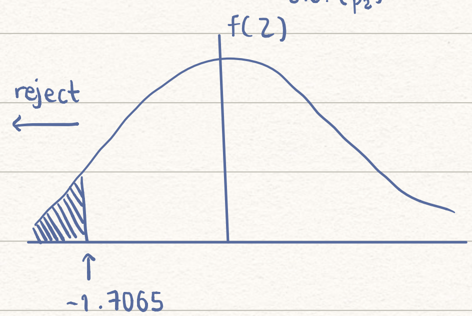
i)

ii) $\hat{\beta}_2 = 0.843$

$H_0 : \beta_2 = 1$ → one-tailed test

$H_a : \beta_2 < 1$

$$t = \frac{\hat{\beta}_2 - \beta_2}{s.e.(\hat{\beta}_2)} = \frac{0.843 - 1}{0.092} = -1.7065$$



Then find the P-value of -1.7065 using z-table (d.f. >30)

* If P-value < 0.05, we reject H_0 at 5% level of significant

* If P-value < 0.01, we reject H_0 at 1% level (or 99%)

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?