

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) OLS includes homoskedasticity, if not, it will be heteroskedasticity and the usual OLS t statistics being invalid

iii) if explanatory variable being omitted the data will be affected and affect result of regression, will cause OLS t statistic being invalid.

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):

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

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:

$$\widehat{\log(\text{salary})} = 4.32 + .280 \log(\text{sales}) + .0174 \text{roe} + .00024 \text{ros}$$

(.32) (.035) (.0041) (.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?

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.

$$\begin{aligned} i) \quad H_0: \beta_3 &= 0 \\ H_a: \beta_3 &> 0 \end{aligned}$$

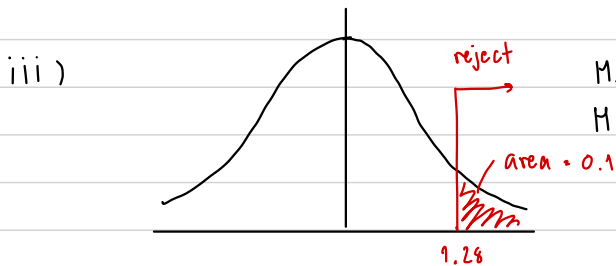
$$\begin{aligned} ii) \quad &\text{let salary be } y \\ &\text{let ros be } x_3 \\ &\text{let } 0.00024 = \beta_3 \end{aligned}$$

$$\frac{d \log(y)}{dx} = \beta_3 \Rightarrow \frac{\frac{1}{y} dy}{dx} = \beta_3$$

$$\frac{100 \frac{1}{y} \Delta y}{dx} = 100 \beta_3 \rightarrow \frac{\% \Delta y}{\Delta x} = 100 \beta_3$$

$$\rightarrow \frac{\% \Delta(\text{salary})}{\Delta(\text{ros})} = 100(0.00024) = 0.024$$

if ros increase by 50 point, salary will increase by $0.024 \times 50 = 1.2\%$.



$$\begin{aligned} H_0: \beta_3 &= 0 \\ H_a: \beta_3 &> 0 \end{aligned}$$

at 10% sig. level

d.f. $209 - 3 - 1 = 205 \rightarrow z\text{-table}$

$$z = \frac{\hat{\beta}_3 - \beta_3}{\text{S.E.}(\hat{\beta}_3)} = \frac{0.00024 - 0}{0.00054} = 0.444 < 0.129$$

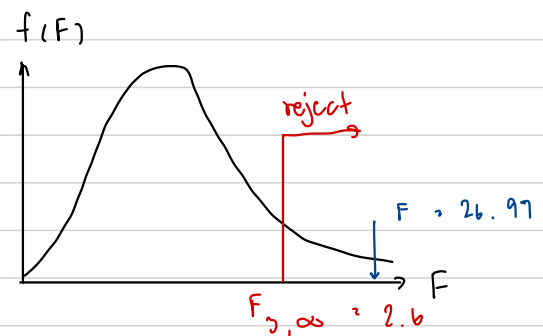
so, we accept H_0 at 10% sig. level

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

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

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

since, $F = 26.97 > 2.6$ we reject H_0 and conclude that firm performance have joint effect in explaining CEO compensation at 5% sig. level



C6. Use the data in WAGE2 for this exercise.

i. Consider the standard wage equation

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

State the null hypothesis that another year of general workforce experience has the same effect on $\log(\text{wage})$ 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?

$$\begin{aligned} \text{i)} \quad H_0: \beta_2 &= \beta_3 \\ H_A: \beta_2 &\neq \beta_3 \end{aligned}$$

$$\begin{aligned} \text{ii)} \quad H_0: \beta_2 &= \beta_3 \rightarrow \beta_2 - \beta_3 = 0 \\ H_A: \beta_2 &\neq \beta_3 \rightarrow \beta_2 - \beta_3 \neq 0 \end{aligned} \quad \left. \vphantom{\begin{aligned} H_0: \beta_2 &= \beta_3 \\ H_A: \beta_2 &\neq \beta_3 \end{aligned}} \right\} \text{let } \theta = \beta_2 - \beta_3 \quad \begin{aligned} H_0: \theta &= 0 \\ H_A: \theta &\neq 0 \end{aligned}$$

substitution $\beta_2 = \theta + \beta_3$ $\beta_2 = \theta + \beta_3$

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + (\theta + \beta_3) \text{exper} + \beta_3 \text{tenure} + u$$

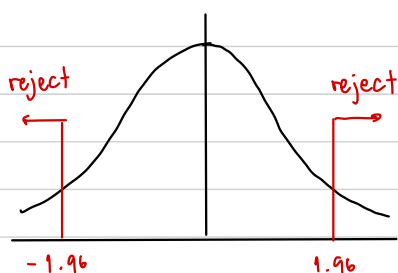
$$= \beta_0 + \beta_1 \text{educ} + \theta \text{exper} + \beta_3 \text{exper} + \beta_3 \text{tenure} + u$$

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

. regress lwage educ exper exper_tenure

Source	SS	df	MS	Number of obs	=	935
Model	25.6953242	3	8.56510806	F(3, 931)	=	56.97
Residual	139.960959	931	.150334005	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
exper_tenure	.0133748	.0025872	5.17	0.000	.0082974 .0184522
_cons	5.496696	.1105282	49.73	0.000	5.279782 5.713609



$0.41 < 1.96$, t -critical does not fall into rejection region, therefore we accept H_0 at 5% sig. level, both variable have the same impact on $\log(\text{wage})$

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) 2017

ii)

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

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

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

intercept = $\beta_0 = -43.03981$

$\beta_1 = 0.7993$ means if annual income increase by 1, net financial wealth will increase by 799.3 \$

$\beta_2 = 0.8427$, means if age of the survey respondent increase by 1 year, net financial wealth will increase by 842.7 \$

iii) Yes, because intercept shows what will net financial wealth be if income and age is 0, but it is not possible since minimum income = 10,008 and minimum wage is 25.

iv)

```
. 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

```
. di .087/2  
.0435
```

one-sided p -value = 0.435, which is greater than 0.01, so we accept H_0 at 1% sig. level.

v)

```
. reg nettfa inc
```

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

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

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

	inc	age
inc	1.0000	
age	0.0391	1.0000

coef. inc. is 0.820, from (ii) coef. inc. = 0.799, it is not so different because correlation between income and age is very small.

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

$$\text{voteA} = \beta_0 + \beta_1 \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u_i$$

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).

- What is the interpretation of β_1 ?
- 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.
- 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)?
- 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) $\beta_1 = 6.083$, if $\log(\text{expendA})$ increase by 1, vote A will increase by 6.083%.

ii) $H = \text{expendA}$, $i = \text{voteA}$, $j = \text{expendB}$

$$\beta_1 = \frac{d \log(H)}{d i} \quad \beta_2 = \frac{d \log(j)}{d i}$$

$$\frac{100 \beta_1 d i}{100 \beta_1 \Delta i} = 100 \frac{1}{H} \Delta H \quad \frac{100 \beta_2 d i}{100 \beta_2 \Delta i} = 100 \frac{1}{j} \Delta j$$

$$H_0: \% \Delta H = - \% \Delta j \quad H_a: \beta_1 \neq -\beta_2$$

$$100 \beta_1 \Delta i = 100 \beta_2 \Delta i$$

$$\beta_1 = -\beta_2$$

iii)

. regress voteA lexpendA lexpendB prtystrA						
Source	SS	df	MS	Number of obs		
Model	38405.1096	3	12801.7032	F(3, 169)	= 215.23	
Residual	10052.1389	169	59.480112	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]	
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 have positive effect on voteA
B's expenditure have negative effect on voteA
the result cannot be use because it does not directly give t statistic for testing, the result of $\beta_1 + \beta_2$ is not calculated

$$\begin{aligned} H_0: \beta_1 - \beta_2 &\rightarrow \beta_1 + \beta_2 = 0 \\ H_a: \beta_1 \neq -\beta_2 &\rightarrow \beta_1 + \beta_2 \neq 0 \end{aligned} \quad \left\{ \begin{array}{l} \text{let } \theta = \beta_1 + \beta_2 \\ \rightarrow H_0: \theta = 0 \\ \rightarrow H_a: \theta \neq 0 \end{array} \right.$$

substitute $\beta_1 = \theta - \beta_2$

$$\text{Vote A} = \beta_0 + (\theta - \beta_2) \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u$$

$$= \beta_0 + \theta \log(\text{expendA}) - \beta_2 \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u$$

$$\text{Vote A} = \beta_0 + \theta \log(\text{expendA}) + \beta_2 (\log(\text{expendB}) - \log(\text{expendA})) + \beta_3 \text{prtystrA} + u$$

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

voteA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lexpendA	-1.532101	.5330858	-1.00	0.320	-1.584466	.5202638
lab	-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

$-1 > -1.96$, t -critical does not fall into rejection region, we accept H_0 at 5% sig. level, 1% increase in A's expenditure is offset by 1% increase by B's expenditure