

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)?

- Heteroskedasticity.
- A sample correlation coefficient of .95 between two independent variables that are in the model.
- Omitting an important explanatory variable.

i) Heteroskedasticity

- If the data appears to be heteroskedastic, it means that random variables do not have a constant error term or variance, thus the OLS t -statistics will be invalid.

iii) Omitting an important explanatory variable

If the omitted x is correlated to other independent variables, $E(u|x) = 0$ is violated

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):

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

- 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.
- 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?

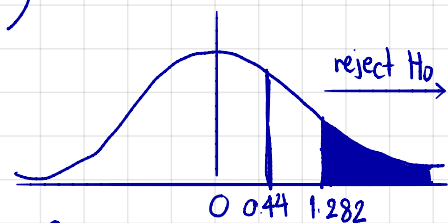
- Test the null hypothesis that ros has no effect on salary against the alternative that ros has a positive effect. Carry out the test at the 10% significance level.

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

$$\text{ii) } 0.00024 \times 50 = 0.012$$

Salary is expected to increase by 1.2%. Therefore ros has a small effect on salary.

iii)



$$\begin{aligned} \text{d.f.} &= n - k - 1 \\ &= 209 - 3 - 1 = 205 \end{aligned}$$

$$t = \frac{\hat{\beta}_3 - a_j}{\text{s.e. } \hat{\beta}_3} = \frac{0.00024}{0.0054}$$

$$= 0.444$$

- Since the critical value is 1.282, we fail to reject H_0 at 10% sig level

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

iv) In order to decide whether we should include ros in the final model, we use t -test. In this case, it shows that t is lower than 1.96 which means that ros does not have a significant impact on salary at 5%. Therefore, we can decide not to include ros in the final model.

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?

i)

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

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

. regress lwage educ 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	.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

Since $n \geq 30$, it has approximately standard normal distribution

$$Z = \frac{(\beta_2 - \beta_3) - 0}{\text{s.e.}(\beta_2 - \beta_3)}$$

$$\hat{\theta}_1 = \hat{\beta}_2 - \hat{\beta}_3$$

$$Z = \frac{\hat{\theta}_1 - 0}{\text{s.e.}(\hat{\theta}_1)} = \frac{\hat{\theta}_1}{\text{s.e.}(\hat{\theta}_1)}$$

Now, $\hat{\beta}_2 = \hat{\theta}_1 + \hat{\beta}_3$
 Or $\beta_2 = \theta_1 + \beta_3$

Sub in the main regression and get

$$y = \beta_0 + \beta_1 \text{educ} + (\theta_1 + \beta_3) \text{exper} + \beta_3 \text{tenure} + u$$

$$= \beta_0 + \beta_1 \text{educ} + \theta_1 \text{exper} + \beta_3 \text{exper} + \beta_3 \text{tenure} + u$$

$$= \beta_0 + \beta_1 \text{educ} + \theta_1 \text{exper} + \beta_3 (\text{exper} + \text{tenure}) + u$$

Now, the explanatory variables are going to be educ, exper, exper + tenure

We can calculate $\frac{\hat{\theta}_1 - 0}{\text{s.e.}(\hat{\theta}_1)}$

```
. gen new = exper+tenure
. reg lwage educ exper new
```

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
new	.0133748	.0025872	5.17	0.000	.0082974 .0184522
_cons	5.496696	.1105282	49.73	0.000	5.279782 5.713609

$$t = \frac{0.0019537}{0.0047434} = 0.4119 < 1.96$$

Therefore H_0 cannot be rejected at 5% level of significance

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).

- How many single-person households are there in the data set?
- Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u_i$$

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 house holds

ii)

```
. reg 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
Total	4565965.05	2,016	2264.86361	Root MSE	=	44.683

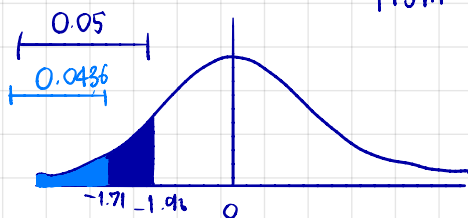
nettfa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inc	.7993167	.0597307	13.38	0.000	.6821762 .9164572
age	.8426563	.0920169	9.16	0.000	.6621982 1.023115
_cons	-43.03981	4.080393	-10.55	0.000	-51.04204 -35.03758

• Assuming *ceteris paribus*, if income increases by 1 unit, net total financial asset is expected to increase by 0.799. Similarly, if age increases by 1, *nettfa* is expected to increase by 0.8426

iii) The intercept found is negative. The estimated *nettfa* is -43 when *inc* = 0 and

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

$$t = \frac{\beta_2 - a_2}{s.e.(\beta_2)} = \frac{0.8426563 - 1}{0.0920169} = -1.7099 \approx -1.71$$



From the z table, the p-value of the value -1.71 is 0.0436

Since the p-value < significance level (0.05), we reject H_0 at 5% level of significance

. reg netffa inc

Source	SS	df	MS	Number of obs	=	9,275
Model	5381009.18	1	5381009.18	F(1, 9273)	=	1532.38
Residual	32562380.3	9,273	3511.52597	Prob > F	=	0.0000
Total	37943389.5	9,274	4091.3726	R-squared	=	0.1418
				Adj R-squared	=	0.1417
				Root MSE	=	59.258

netffa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inc	.9999113	.0255433	39.15	0.000	.9498408 1.049982
_cons	-20.17948	1.176434	-17.15	0.000	-22.48555 -17.87341

. When we do a simple regression of netffa on inc, the estimated β_1 is close to 1. In comparison to the estimate in part ii which is almost close to 0.8, this is not much different.

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

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

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.)

$$\begin{aligned} \text{i)} \quad \Delta \text{Vote A} &= \beta_1 \log(\text{expend A}) \\ &= \frac{\beta_1}{100} (100 \times \Delta \log(\text{expendA})) \\ &= \frac{\beta_1}{100} (\% \Delta \log \text{expendA}) \end{aligned}$$

$\frac{\beta_1}{100}$ is the percentage point change in percentage of vote received when campaign expenditure by candidate A when campaign expenditure increases by 1%

$$\begin{aligned} \text{ii)} \quad H_0 &: -\beta_1 = \beta_2 \\ H_a &: -\beta_1 \neq \beta_2 \end{aligned}$$

reg voteA lexpendA lexpendB prtystra

Source	SS	df	MS	Number of obs	=	173
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

$$y = 45.07 + 6.08 \log(\text{lexpendA}) - 6.62 \log(\text{lexpendB}) + 0.15 (\text{prtystra})$$

The coefficients on $\log(\text{lexpendA})$ and $\log(\text{lexpendB})$ are very significant, as shown by the t-statistics significantly greater than 2.57. The estimates show that a 1000 increase in expenditure by candidate A increases the percentage of votes by 6%. On the other hand, a 1000 increase in expenditure by candidate B reduces vote A by 6%.

Since the coefficients are opposite in sign, we can reject H_0

$$\text{iv) } H_0: -\beta_1 = \beta_2$$

$$H_0: 0 = \beta_1 + \beta_2$$

$$t = \frac{(\beta_1 + \beta_2) - 0}{\text{se}(\beta_1 + \beta_2)} = \frac{\beta_1 - \beta_2}{\text{se}(\beta_1 + \beta_2)}$$

$$\theta_1 - \beta_2 = \beta_1 + \beta_2$$

$$\theta_1 - \beta_2 = \beta_1$$

$$\begin{aligned} \text{Vote A} &= \beta_0 + (\theta_1 - \beta_2) \log(\text{lexpendA}) + \beta_2 \log(\text{lexpendB}) + \beta_3 (\text{prtystra}) + u \\ &= \beta_0 + \theta_1 \log(\text{lexpendA}) - \beta_2 \log(\text{lexpendA}) + \beta_2 \log(\text{lexpendB}) + \beta_3 (\text{prtystra}) + u \\ &= \beta_0 + \theta_1 \log(\text{lexpendA}) + \beta_2 (\log(\text{lexpendB}) - \log(\text{lexpendA})) + \beta_3 (\text{prtystra}) + u \end{aligned}$$

Now, the explanatory variables are $\log(\text{lexpendA})$, $\log(\text{lexpendB}) - \log(\text{lexpendA})$, prtystra

→ $\text{lexpendB} - \text{lexpendA}$

. reg voteA lexpendA new prtystra

Source	SS	df	MS	Number of obs	=	173
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
				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
new	-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

$$t = \frac{\hat{\theta}_1 - 0}{\text{se}(\hat{\theta}_1)}$$

$$= \frac{-0.532}{0.533}$$

$$= -1$$

Therefore, H_0 cannot be rejected at 5% level