

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.

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

(i) Heteroskedasticity is correct.

All random variables have the same error term or finite variance or can be called "Homoskedasticity" as one of the assumptions, but if the data does not hold or "Heteroskedasticity" (the random variable does not have a common error term or variance) 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.

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:

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

$\begin{matrix} \text{s.e.} & (.32) & (.035) & \text{s.e.} & & (.0041) & (.00054) \\ & & & & & \text{s.e.} & \text{s.e.} \end{matrix}$
 $n = 209, R^2 = .283.$

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

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

$$= 0.012$$

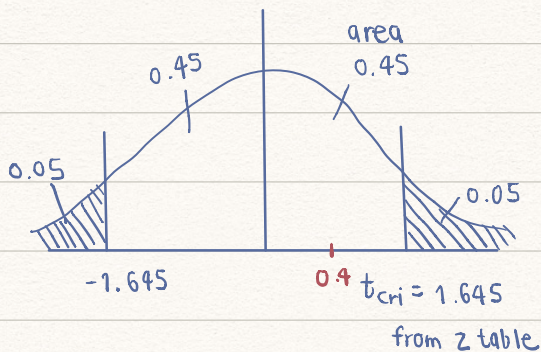
or equal to 1.2 % #

iii) $H_0: \beta_3 = 0$

$$d.f = n - k - 1 = 209 - 3 - 1 = 205$$

$H_a: \beta_3 \neq 0$

significant level = 10% or 0.1



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

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

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

\therefore 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:

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

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).

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 prtystA

| 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] |
|-------------------------|---------------------|-----------|--------|-------|----------------------|
| 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 prtystA | β_3 .1519574 | .0620181 | 2.45 | 0.015 | .0295274 .2743873 |
| intercept _cons | β_0 45.07893 | 3.926305 | 11.48 | 0.000 | 37.32801 52.82985 |

(intercept)

(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

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

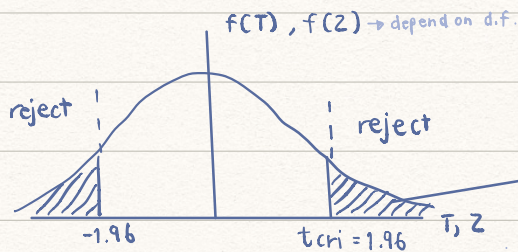
(11.48) (15.92) (-17.46) (2.45)

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

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

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



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

hypothesis / value

\Rightarrow we compute this t -statistic and compare with critical value

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

\therefore 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

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

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?

. regress lwage educ exper tenure

| Source | SS | df | MS | Number of obs | | |
|----------|------------|-----|------------|---------------|---|--------|
| 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] | |
|----------------|-------------|-----------|-------|-------|----------------------|----------|
| x1 educ | b1 .0748638 | .0065124 | 11.50 | 0.000 | .062083 | .0876446 |
| x2 exper | b2 .0153285 | .0033696 | 4.55 | 0.000 | .0087156 | .0219413 |
| x3 tenure | b3 .0133748 | .0025872 | 5.17 | 0.000 | .0082974 | .0184522 |
| intercept_cons | b0 5.496696 | .1105282 | 49.73 | 0.000 | 5.279782 | 5.713609 |

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

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

} two-tailed test

$$t = \frac{(\hat{\beta}_2 - \hat{\beta}_3) - 0}{\text{s.e.}(\hat{\beta}_2 - \hat{\beta}_3)}$$

let $\hat{\Theta}_1 = \hat{\beta}_2 - \hat{\beta}_3$; $H_0 : \Theta_1 = 0$, we have

$H_a : \Theta_1 \neq 0$ $t = \frac{\hat{\Theta}_1 - 0}{\text{s.e.}(\hat{\Theta}_1)}$

If rearrange $\hat{\Theta}_1 = \hat{\beta}_2 - \hat{\beta}_3$, we have $\hat{\beta}_2 = \hat{\Theta}_1 + \hat{\beta}_3$ or $\beta_2 = \Theta_1 + \beta_3$

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

$$= \beta_0 + \beta_1(\text{educ}) + (\Theta_1 + \beta_3)(\text{exper}) + \beta_3(\text{tenure}) + u$$

$$= \beta_0 + \beta_1(\text{educ}) + \Theta_1(\text{exper}) + \beta_3(\text{exper}) + \beta_3(\text{tenure}) + u$$

$$= \beta_0 + \beta_1(\text{educ}) + \Theta_1(\text{exper}) + \beta_3(\text{exper} + \text{tenure}) + u$$

new variable

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 |
| Total | 4565965.05 | 2,016 | 2264.86361 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.1193 |
| | | | | Adj R-squared | = | 0.1185 |
| | | | | 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 |

i)

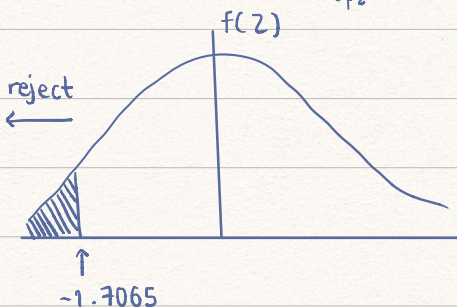
ii) $\hat{\beta}_2 = 0.843$

$H_0 : \beta_2 = 1$

→ 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

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?

* If p-value < 0.01, we reject H_0 at

1% level (or 99%)