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. ($se\,\hat\beta$)	t ($\frac{\hat\beta - 0}{se\,\hat\beta}$)	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 (intercept) 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 $\rightarrow$ $t = \dfrac{(\hat\beta_2 - \hat\beta_3) - 0}{se.(\hat\beta_2 - \hat\beta_3)}$

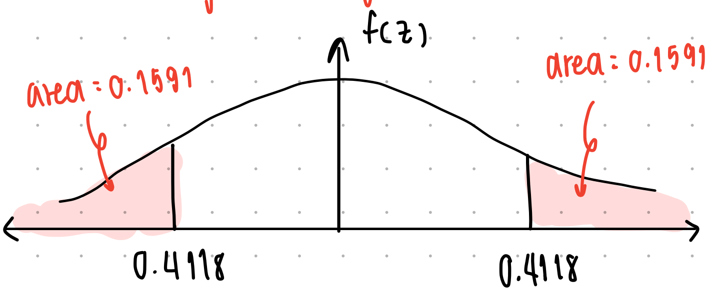
let $\hat\theta_1 = \hat\beta_2 - \hat\beta_3$; $H_0 : \theta_1 = 0$, $t = \dfrac{\hat\theta_1 - 0}{s.e(\hat\theta_1)}$

$H_a : \theta_1 \neq 0$

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

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

do the regression again




```
. regress lwage educ exper newV
```

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.	t	P>|t|	[95% Conf.	Interval]
educ	.0748638	.0065124	11.50	0.000	.062083	.0876446
θ exper	.0019537	.0047434	0.41	0.681	-.0073554	.0112627
newV	.0133748	.0025872	5.17	0.000	.0082974	.0184522
_cons	5.496696	.1105282	49.73	0.000	5.279782	5.713609

$t = \dfrac{0.0019537 - 0}{0.0047434} = 0.4118 \approx 0.41$

at 5% significant, p value is $0.1591 > 0.025$ (significant level)

So, we do not reject H_0 because p value (0.1591) is greater than significant level (0.025)

Therefor, another year of general workforce experience has the same effect on log(wage) 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. 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? at $\alpha = 1\%$.

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 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]
x_1 inc	$\hat{\beta}_1$ .7993167	.0597307	13.38	0.000	.6821762	.9164572
x_2 age	$\hat{\beta}_2$ .8426563	.0920169	9.16	0.000	.6621982	1.023115
intercep _cons	-43.03981	4.080393	-10.55	0.000	-51.04204	-35.03758

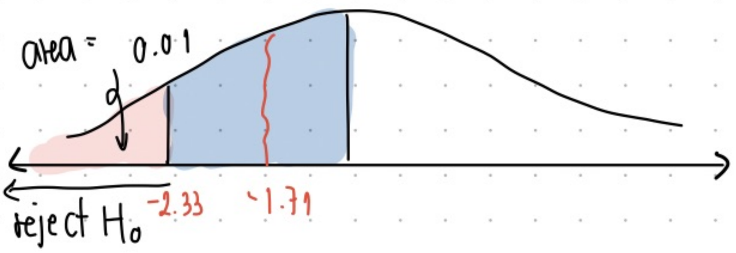
i N = 2017

ii. We can see that if annual family income increase by $1000, on average, net financial wealth will increase by $799.3. For every additional age of survivor respondent, the average net financial wealth go up by about $842.7. There is no surprise for the increase of nettfa that come from each additional ice and age, the more year of working can mean the kore year of saving too.

iii. The interception point is -43.03981, it means that if both inc and Age are zero the net financial wealth would be about negative $43,039.81. Therefore, when the age of the survey respondent and annual family income are zero, the saving of 43,039.81 will occur.

iv. $H_0: \beta_2 = 1$

$H_a: \beta_2 < 1$



area left = 0.049

$z = -2.33$

$$z = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0.8426 - 1}{0.0920} = -1.71$$

$-1.71 > -2.33$, the computed z value did not fall into the rejection region, therefor, we do not reject H_0

$inc = \beta_0 + \beta_1 nettfa + u$

$inc = 28.07666 + 0.100737\, nettfa + u$

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

Source	SS	df	MS		
Model	46335.1731	1	46335.1731	Number of obs =	2,017
Residual	514127.962	2,015	255.150354	F(1, 2015) =	181.60
				Prob > F =	0.0000
				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

we can see from the slope of this equation that if net financial wealth increase by $1,000, on average, the annual family income will increase by $100.737
And it shows the different at the interception point, as the orientation one result is negative but in this equation when the net financial wealth is zero, the remaining amount is positive $28,076.66.

To summarise is the effect of (inc on nettfa) has a higher potential than effect of (nettfa on inc) due to the significance change of the slope from $799.3 to $100.737 when we change to regress inc on nettfa.

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 ?

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)?

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.)

i. **This equation shows that if campaign expenditure by CandidateA increases by 1,on average, the percentage of the vote received by candidateA will increase by β1**

ii. $H_0 : \beta_2 = -\beta_1$, $H_0 : \beta_2 + \beta_1 = 0$

$H_a : \beta_2 \neq -\beta_1$, $H_a : \beta_2 + \beta_1 = 0$

```
. 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.	t	P>|t|	[95% Conf.	Interval]
x_1 lexpendA	6.083316	.38215	15.92	0.000	5.328914	6.837719
x_2 lexpendB	-6.615417	.3788203	-17.46	0.000	-7.363246	-5.867588
x_3 prtystrA	.1519574	.0620181	2.45	0.015	.0295274	.2743873
intercept _cons	45.07893	3.926305	11.48	0.000	37.32801	52.82985

iii. $VoteA = 45.07893 + 6.083316 \log(expendA) - 6.615417 \log(expendB) + 0.1519574\, prtystrA + u$

Both variables have effects on voteA but we can not analyse how different effect from each variable.
Therefor, we can't use these results to test the hypothesis in prat ii

---

iv. $H_0 : \beta_2 = -\beta_1$, $H_0 : \beta_2 + \beta_1 = 0$ $\longrightarrow$ $t = \dfrac{(\hat\beta_2 + \hat\beta_1) - 0}{se.(\hat\beta_2 + \hat\beta_1)}$

$H : \beta_2 \neq -\beta_1$, $H_a : \beta_2 + \beta_1 \neq 0$

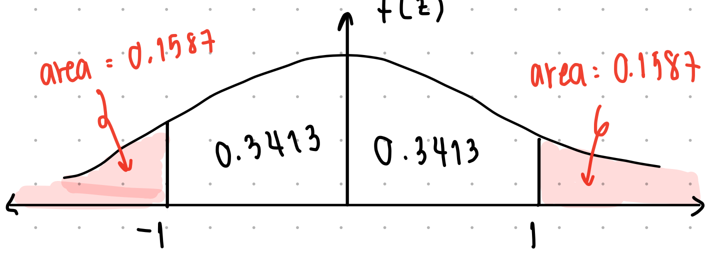
let $\hat\theta_1 = \hat\beta_2 + \hat\beta_1 \longrightarrow H_0 : \theta_1 = 0$, $t = \dfrac{\hat\theta_1 - 0}{se.(\hat\theta_1)}$

$H_a : \theta_1 \neq 0$

if rearrange $\hat\theta_1 = \hat\beta_2 + \hat\beta_1$, we have $\hat\beta_1 = \hat\theta_1 - \hat\beta_2$ or $\beta_1 = \theta_1 - \beta_2$

$$\begin{aligned} voteA &= \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u \\ &= \beta_0 + \theta_1 - \beta_2 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u \\ &= \beta_0 + \theta_1 \log(expendA) - \beta_2 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u \\ &= \beta_0 + \theta_1 \log(expendA) - \beta_2 (\underbrace{\log(expendB) - \log(expendA)}_{newV}) + \beta_3 prtystrA + u \end{aligned}$$

do the regresion again




```
. regress voteA lexpendA newV prtystrA
```

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

$t = \dfrac{\hat\theta_1 - 0}{se\,\hat\theta_1} = \dfrac{-0.532101 - 0}{0.5330858} = -0.99851 \approx -1$

at 1% significant, p value is 0.1587 > 0.005 (significant level)

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.

So, we do not reject H_0 because p value (0.1587) is greater than significant level (0.005)

Therefor, a 1% increase in A's expenditures is offset by a 1% increase in B's expenditure.

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. Due to assumption 5 on MLR refer to homoskedasticity therefore the OLS t statistics will be invalid.

iii. Omitting an important explanatory variable will cause the OLS to invalid as the equation would not find the accurate outcome from the incomplete equation. The coefficient of regression will fluctuate and will eventually cause the equation to be 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):

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

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*?

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.

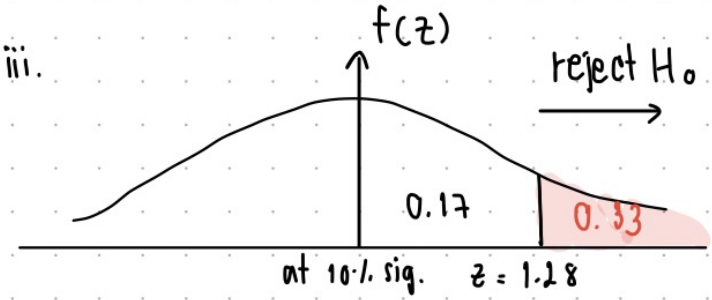
i. $H_0: \beta_3 = 0 \rightarrow$ stock performance has no effect on CEO's salary.

$H_a: \beta_3 > 0 \rightarrow$ better stock market performance increase a CEO's salary

$n > 30$

ii. if Ros increase by 50 points, salary will increase by 0.012 or percentage increase equal to 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 50 point increase in log(sale) and ROE will result in salary to increase by 14 and 0.87 respectively.

iii.



reject H_0 $\longrightarrow$

$$z = \frac{\hat{\beta}_3 - 0}{se\,\hat{\beta}_3} = \frac{0.00024 - 0}{0.00054} = 0.44$$

at 10% significant critical value is 1.28 that is greater than 0.44 $(0.44 < 1.28)$

so we do not reject H_0 at 10% significant level. the computed z value did not fall into the rejection region. Therefor, ros has no effect on CEO's salary.

iv. I would not include Ros into the equation as Ros has no effect on ceo's salary so I have no reason to put it in the final model to calculate for ~~log(salary)~~.