

Problem Set 3

Jacqueline Bouvier Applied Stats/Quant Methods 1

Due: November 19, 2022

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in **R**, please include the code you used to get your answers. Please also include the **.R** file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub.
- This problem set is due before 23:59 on Sunday November 19, 2023. No late assignments will be accepted.

In this problem set, you will run several regressions and create an add variable plot (see the lecture slides) in **R** using the `incumbents_subset.csv` dataset. Include all of your code.

Question 1

We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

Table 1: Vote Share vs Spending Difference

	<i>Dependent variable:</i>
	voteshare
difflog	0.042*** (0.001)
Constant	0.579*** (0.002)
Observations	3,193
R ²	0.367
Adjusted R ²	0.367
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,852.791*** (df = 1; 3191)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

2. Make a scatterplot of the two variables and add the regression line.

```

1 plot(inc.sub$difflog, inc.sub$voteshare, main = "Figure 1: Difflog to
   Votershare",
2       xlab = "Difflog", ylab = "Voteshare",
3       col = "darkgreen")
4 abline(q1, col = "black")

```

] Please see Figure 1

3. Save the residuals of the model in a separate object.

```

1 resid_1 <- residuals(q1)
2 print(resid_1)

```

]

4. Write the prediction equation.

$y = 0.579 + 0.0417x$ A one unit increase in difflog means a 0.0417 increase in vote share

Question 2

We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

1. Run a regression where the outcome variable is **presvote** and the explanatory variable is **difflog**.

Table 2: Pres Vote vs Spending Difference

	<i>Dependent variable:</i>
	presvote
difflog	0.024*** (0.001)
Constant	0.508*** (0.003)
Observations	3,193
R ²	0.088
Adjusted R ²	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Make a scatterplot of the two variables and add the regression line.

```
1 plot(inc.sub$difflog, inc.sub$presvote, main = "Figure 2: Difflog to  
   Presvote",  
2       xlab = "Difflog", ylab = "Presvote",  
3       col = "cyan")  
4 abline(q2, col = "black")
```

] Please see Figure 2

3. Save the residuals of the model in a separate object.

```
1 resid_2 <- residuals(q2)  
2 print(resid_2)
```

]

4. Write the prediction equation.

$y = 0.508 + 0.024x$ A one unit increase in difflog is a 0.024 unit in Pres Vote

Question 3

We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

1. Run a regression where the outcome variable is **voteshare** and the explanatory variable is **presvote**.

Table 3: Vote Share vs Pres Vote

	<i>Dependent variable:</i>
	voteshare
presvote	0.388*** (0.013)
Constant	0.441*** (0.008)
Observations	3,193
R ²	0.206
Adjusted R ²	0.206
Residual Std. Error	0.088 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

2. Make a scatterplot of the two variables and add the regression line.

```
1 plot(inc.sub$presvote, inc.sub$voteshare, main = "Figure 3: Presvote to  
   Votershare",  
2       xlab = "Presvote", ylab = "Voteshare",  
3       col = "darkorange")  
4 abline(q3, col = "black")
```

] Please see Figure 3

3. Write the prediction equation.

$y = 0.441 + 0.388x$ A one unit increase in Pres Vote is a 0.388 increase in Vote Share

Question 4

The residuals from part (a) tell us how much of the variation in **voteshare** is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in **presvote** is *not* explained by the difference in spending between incumbent and challenger in the district.

1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

Table 4: Vote Share vs Spending Difference

	<i>Dependent variable:</i>
	resid_1
resid_2	0.257*** (0.012)
Constant	-0.000 (0.001)
Observations	3,193
R ²	0.130
Adjusted R ²	0.130
Residual Std. Error	0.073 (df = 3191)
F Statistic	476.975*** (df = 1; 3191)

Note: *p<0.1; **p<0.05; ***p<0.01

2. Make a scatterplot of the two residuals and add the regression line.

```

1 plot(resid_2, resid_1,
2     main = "Figure 4: Residuals Plot",
3     xlab = "Residuals 1",
4     ylab = "Residuals 2",
5     col = "darkorchid")
6 abline(q4, col = "black")

```

] Please see Figure 4

3. Write the prediction equation.

$y = 2.569e-01x - 1.942e-18$ A one unit decrease in the Residuals from residual of question 1 means a decrease of 1.942e-18 of the residuals of question 2

Question 5

What if the incumbent's vote share is affected by both the president's popularity and the difference in spending between incumbent and challenger?

1. Run a regression where the outcome variable is the incumbent's `voteshare` and the explanatory variables are `difflog` and `presvote`.

Table 5: Comparing Difflog and Pres Vote to Vote Share

	<i>Dependent variable:</i>
	voteshare
difflog	0.036*** (0.001)
presvote	0.257*** (0.012)
Constant	0.449*** (0.006)
Observations	3,193
R ²	0.450
Adjusted R ²	0.449
Residual Std. Error	0.073 (df = 3190)
F Statistic	1,302.947*** (df = 2; 3190)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

2. Write the prediction equation.

$y = 0.449 + 0.035x + 0.257x_1$ For every increase of one unit of difflog and pres vote we would see a 0.035 in vote share along with a 0.257 increase of vote share to the regression lines of the respective explanatory variable.

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case?

We can see that the standard error residuals are very close to being the same. This could mean that a few different things. Since the residuals are relatively close we could assume that the line of best fit would be similar. The standard error of residuals is from the residuals from question 1 and 2, which then could mean that the data from these two regressions are connected with the multiregression of difflog and pres vote to vote share. All the data isn't equal but highly correlated with each other. We could assume that the data all follows a particular pattern leading to a possible agreement of accepting the hypothesis.

Figure 1: Figures 1 to 4

