

Problem Set 3

Applied Stats/Quant Methods 1 - Ariana Alves Antunes

Date: 20 November, 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 20, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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`.

```
1 # run regression model with voteshare and difflog to understand its
  challenges
2 regression_model_problem1 <- lm(voteshare ~ difflog, data=Incumbets_
  subset)
```

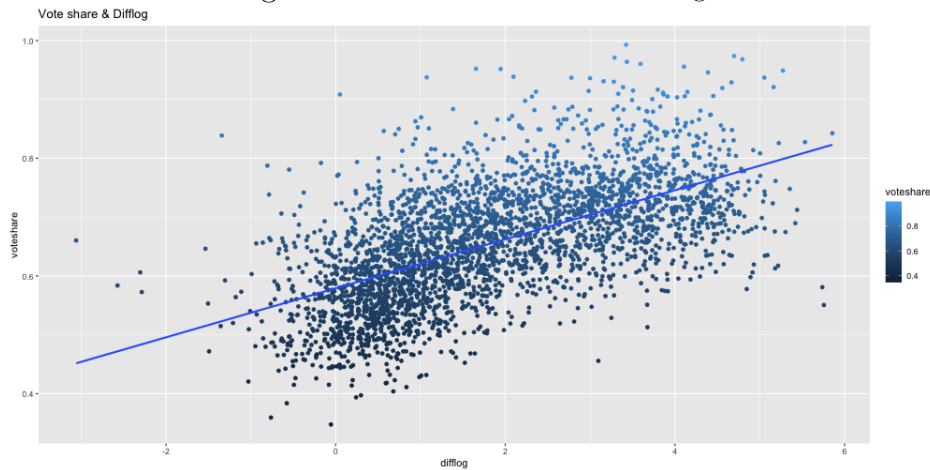
Table 1: Summary table for Difflog and Vote share

	<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 png(filename = "plot_1.png", width = 1000, height = 500)
2 ggplot(data = Incumbets_subset, aes(x = difflog, y = voteshare)) +
3   geom_point(aes(color = voteshare)) +
4   geom_smooth(method = "lm", se = FALSE) +
5   labs(title = "Vote share & Difflog")
6 dev.off()
```

Figure 1: Plot of vote share and difflog.



There is a strong positive trend between the response variable 'vote share' and explana-

tory variable 'difflog'

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

```
1 #Residuals of the model
2 deviance(regression_model_problem1)
3 Inc_residuals <- resid(regression_model_problem1) #Residuals of the model
  in a separate object
```

4. Write the prediction equation.

$$\bar{y} = a + bx$$

$$\bar{y} = 0.579031 + 0.041666x$$

The predicted difference increases by $b = 0.041$ difference between difference in expending between the two candidates.

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**.

```
1 # get summary of model with coefficient estimates
2 summary(regression_model_problem2)
3 #table for the summary with stargazer
4 stargazer(regression_model_problem2, type = "latex", title = "Summary
  table for Difflog & Presvote")
```

Table 2: Summary table for Difflog and Presvote

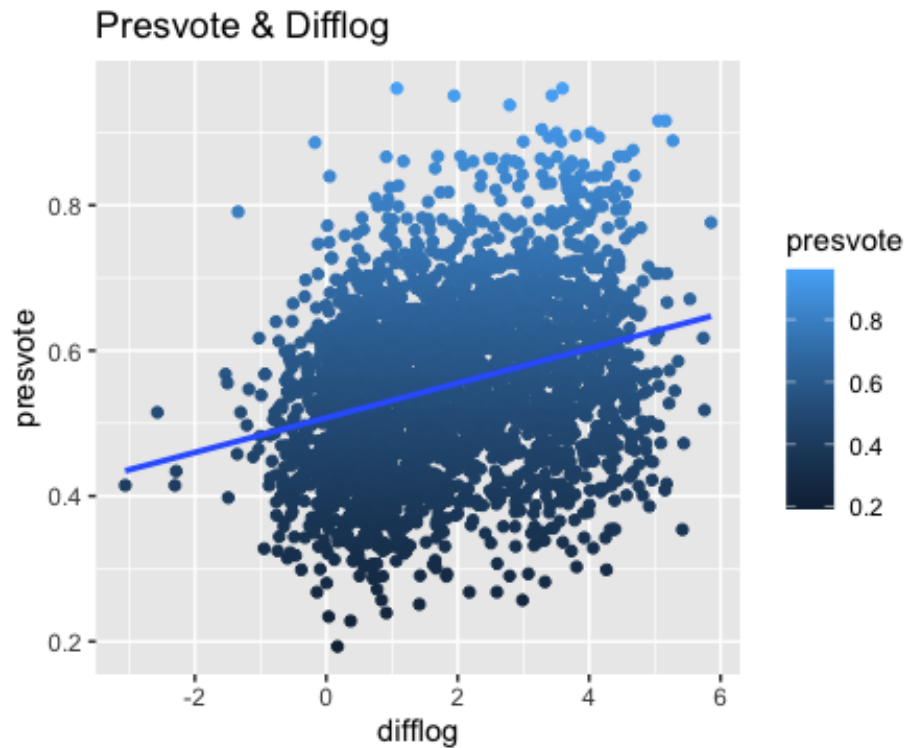
	<i>Dependent variable:</i>
	presvote
difflog	0.024*** (0.001)
Constant	0.508*** (0.003)
Observations	3,193
R ²	0.088
Adjusted R ²	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Inference:

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

```
1 ggplot(data = Incumbets_subset, aes(x = difflog, y = presvote)) +
2   geom_point(aes(color = presvote)) +
3   geom_smooth(method = "lm", se = FALSE) +
4   labs(title = "Presvote & Difflog")
5 dev.off()
```

Figure 2: Plot of Presvote and Difflog.



There is a strong positive trend between the response variable 'presvote' and explanatory variable 'difflog'.

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

```
1 #Residual
2 Inc_residuals2 <- resid(regression_model_problem2) #Residuals of the
  model in a separate object
```

4. Write the prediction equation.

$$\bar{y} = a + bx$$

$$\bar{y} = 0.508 + 0.024x$$

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**.

```
1 #Regression x voteshare and y presvote
2 regression_model_problem3 <- lm(voteshare ~ presvote, data=Incumbets_subset)
```

Table 3: Summary table for Voteshare and Presvote

	<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.

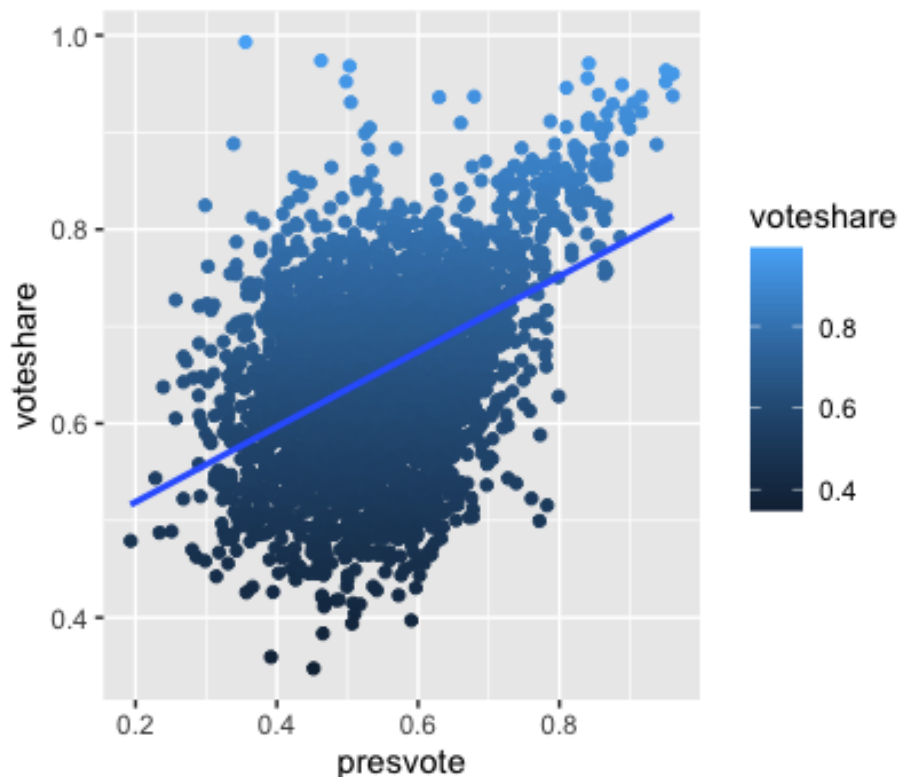
```
1 ggplot(data = Incumbets_subset, aes(x = presvote, y = voteshare)) +
2   geom_point(aes(color = voteshare)) +
3   geom_smooth(method = "lm", se = FALSE)
4 dev.off()
```

3. Write the prediction equation.

$$\bar{y} = a + bx$$

$$\bar{y} = 0.441330 + 0.388018x$$

Figure 3: Plot of Presvote and Difflog.



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.

```
1 regression_model_problem4 <- lm(Inc_residuals ~ Inc_residuals2)
2 # get summary of model
3 summary(regression_model_problem4)
4 #table for the summary with stargazer
5 stargazer(regression_model_problem4, type = "latex", title = "Summary
6           table for Residuals")
```

Table 4: Summary table for Residuals

	<i>Dependent variable:</i>
	Inc_residuals
Inc_residuals2	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)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

```

1 #Scatterplot
2 png("plot_4.png", width = 600, height = 350)
3 ggplot(data = Incumbets_subset, aes(x = Inc_residuals2, y = Inc_residuals
4   )) +
5   geom_point(aes(color = Inc_residuals2)) +
6   geom_smooth(method = "lm", se = FALSE)
7 dev.off()

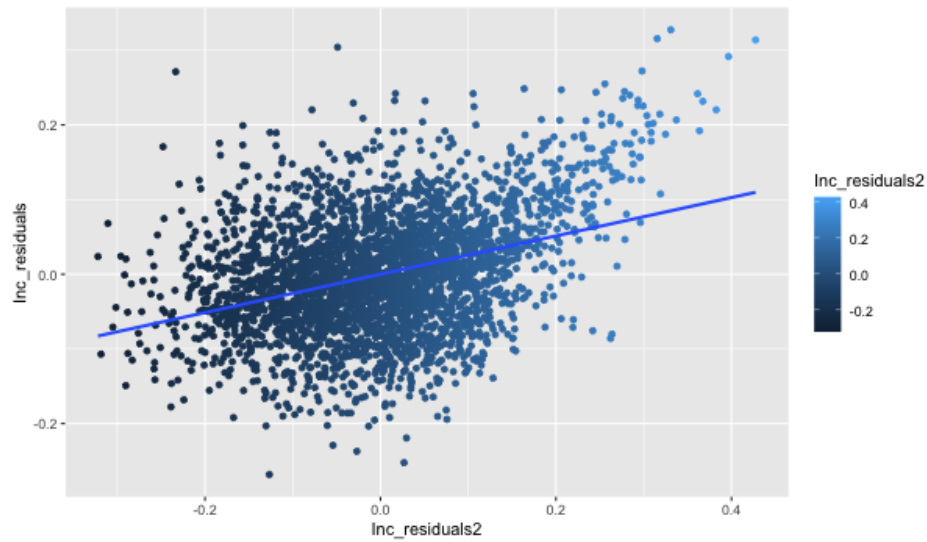
```

3. Write the prediction equation.

$$\bar{y} = a + bx$$

$$\bar{y} = 0 + 0.257x$$

Figure 4: Plot of Residuals



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`.

```
1
2 regression_model_problem5 <- lm(voteshare ~ difflog + presvote, data=
  Incumbets_subset)
3 summary(regression_model_problem5)
4 stargazer(regression_model_problem5, type = "latex", title = "Summary
  table for
5           voteshare, difflog and presvote")
```

2. Write the prediction equation.

$$Y = Xb + e = 0.36 + 0.257 \times \text{presvote} + 0.449 \times \text{constant}$$

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

Table 5: Summary table for voteshare, difflog and presvote

	<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

In this summary the presvote output is 0.257 and in questions 4 the residual for question 2 is also 0.257, since the variables are with the same covariance and variables. ! this is probably wrong.