Problem Set 3

Vismante Dringelyte 17341481

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`.

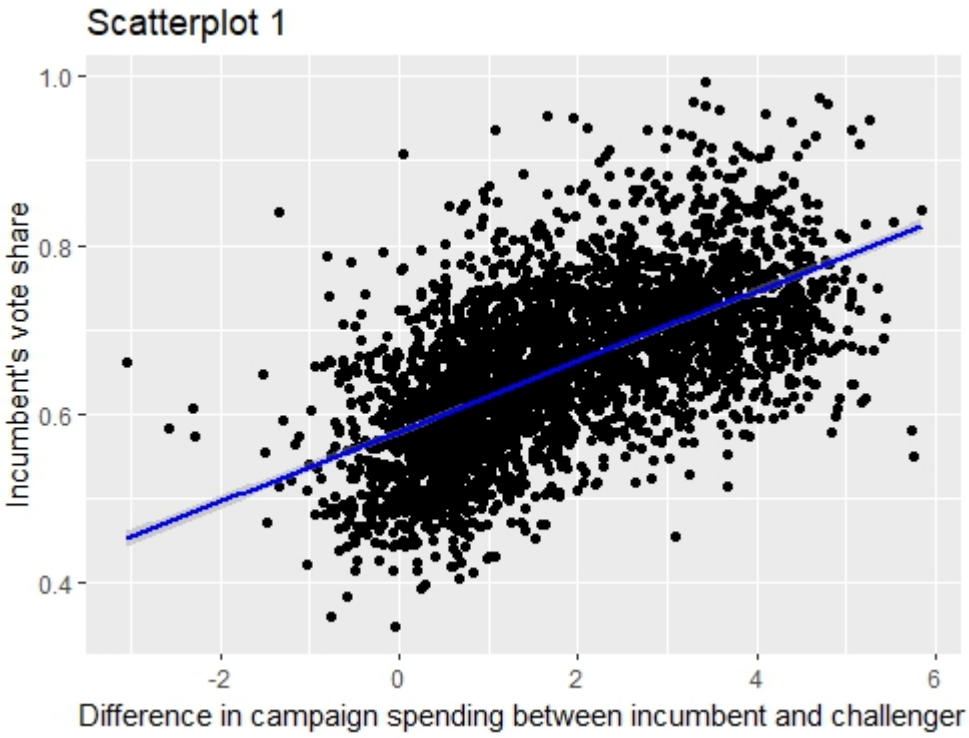
   I use the code we learnt in tutorials to run the regression

```
lm1 <- lm(inc.sub$voteshare ~ inc.sub$difflog)
summary(lm1)
```

Table 1:

	Dependent variable:
	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

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



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

```
residuals1 <- lm1$residuals
```

4. Write the prediction equation.

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 0.579 + 0.042 \cdot \text{difflog}$$

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.

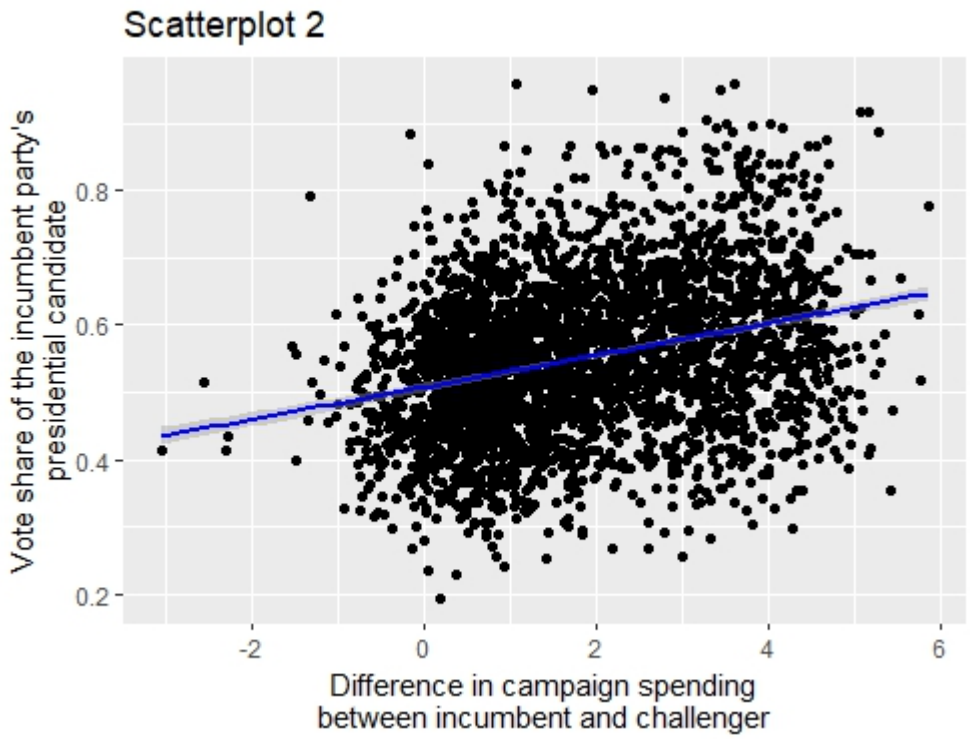
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```
lm2 <- lm(inc.sub$presvote ~ inc.sub$difflog)
summary(lm2)
```

Table 2:

	Dependent variable:
	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.



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

```
residuals2 <- lm2$residuals
```

4. Write the prediction equation.

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 0.508 + 0.024 \times \text{difflog}$$

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.

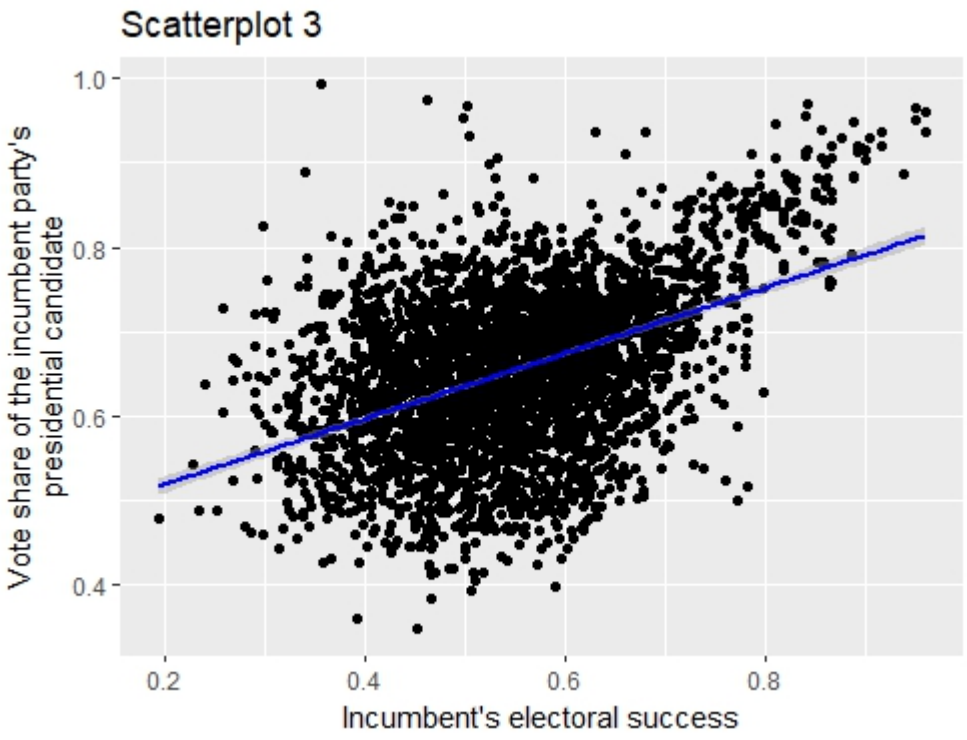
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`.

```
lm3 <- lm(inc.sub$voteshare ~ inc.sub$presvote)
summary(lm3)
```

Table 3:

	Dependent variable:
	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

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



3. Write the prediction equation.

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = 0.441 + 0.388 \times \text{presvote}$$

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.

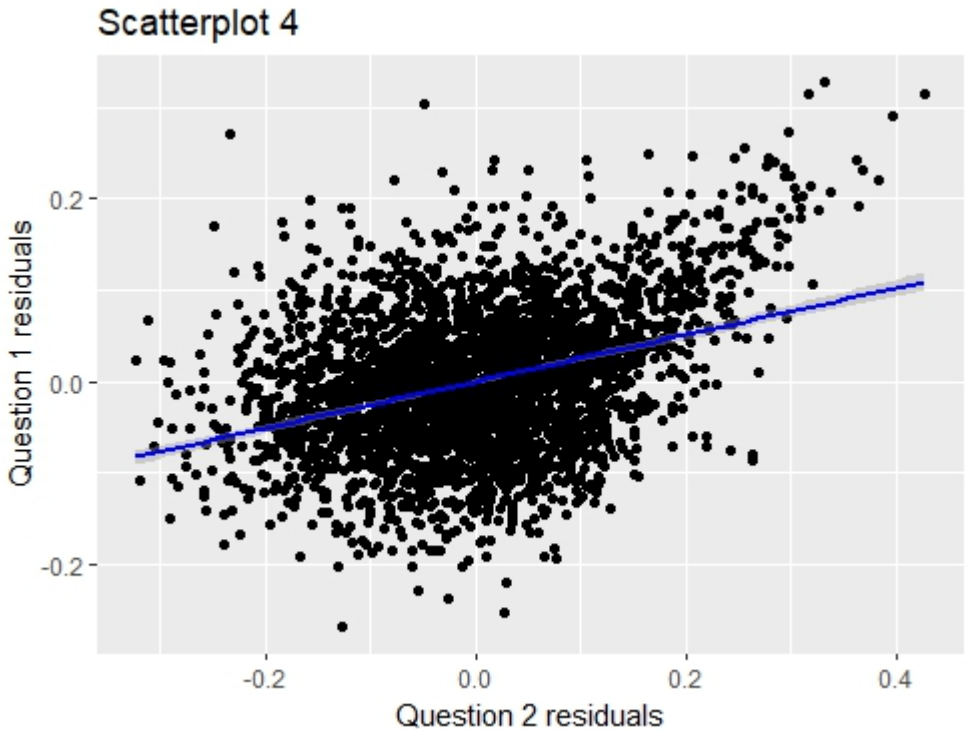
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
lm4 <- lm(residuals1 ~ residuals2)
summary(lm4)
```

Table 4:

	Dependent variable:
	residuals1
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)
Note:	*p<0.1; **p<0.05; ***p<0.01

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



3. Write the prediction equation.

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = -5.934 \times 10^{-18} + 25.69 \times \text{residuals2}$$

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`.

```
lm5 <- lm(inc.sub$voteshare ~ inc.sub$difflog + inc.sub$presvote, data =
    inc.sub)
summary(lm5)
```

Table 5:

	Dependent variable:
	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Write the prediction equation.

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 = 0.449 + 0.036 \times \text{difflog} + 0.257 \times \text{presvote}$$

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

   The residual standard error is the same. This may be because both models show a similar level of variability.