Problem Set 3

Applied Stats/Quant Methods 1

Due: November 20, 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

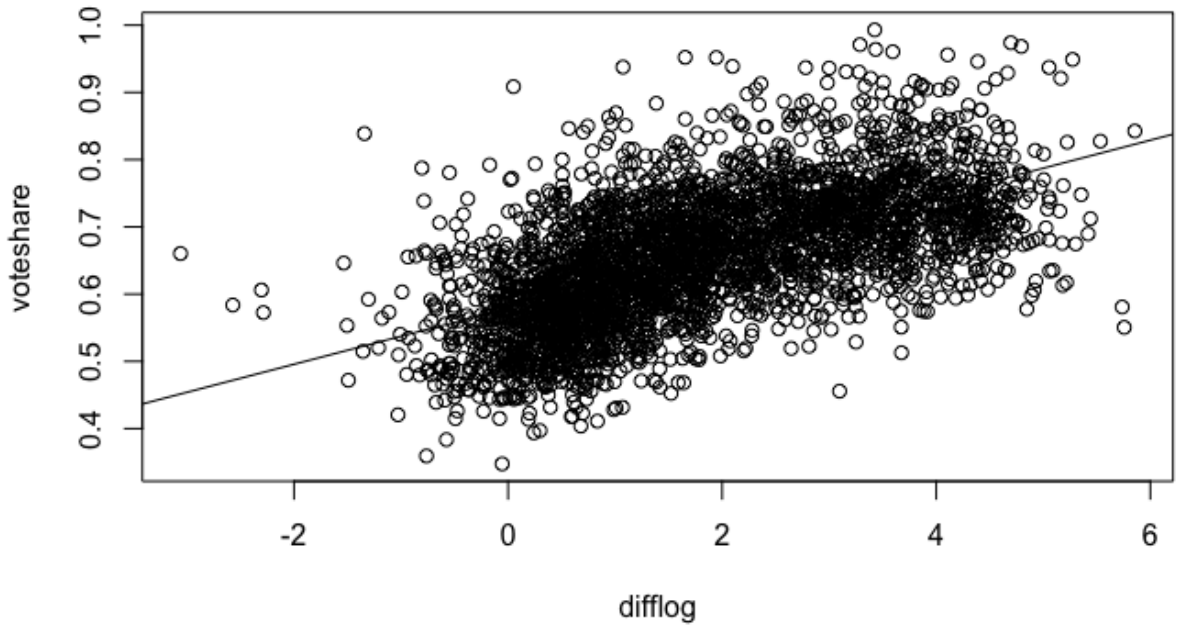
We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

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

```
# Part 1
        q1 <- lm(voteshare ~ difflog, data=df)
        summary(q1)
```

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

```
plot(voteshare ~ difflog, data=df)
abline(q1)
```



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

```
q1res <- q1$residuals
```

4. Write the prediction equation.

$$y = 0.041666x + 0.579031$$

Question 2

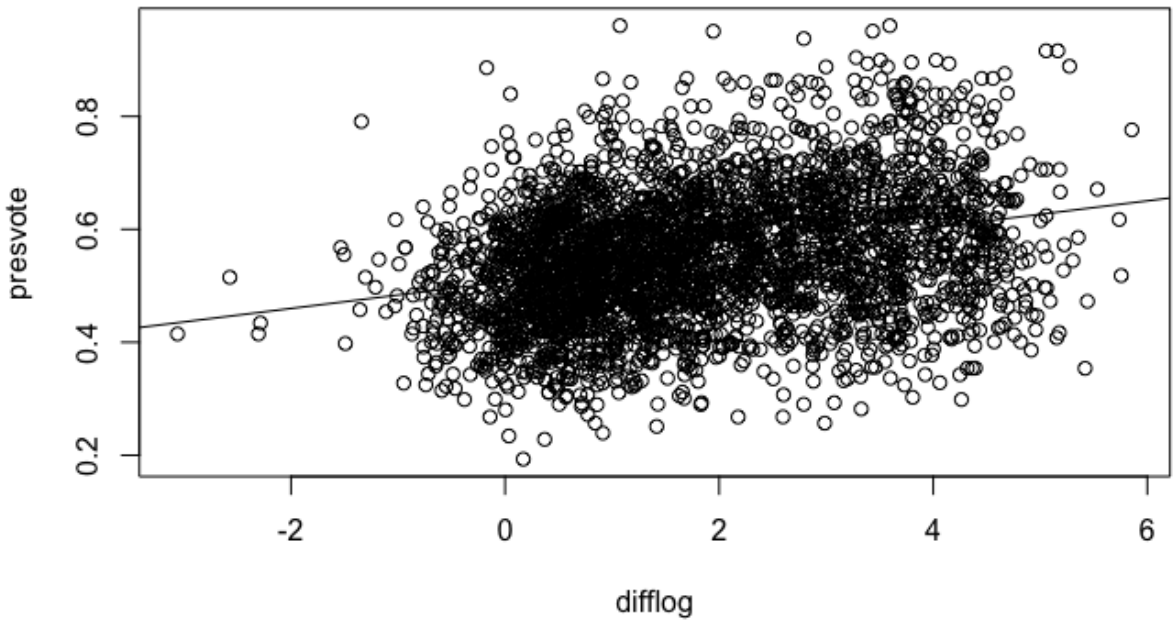
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
q2 <- lm(presvote ~ difflog, data=df)
summary(q2)
```

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

```
plot(presvote ~ difflog, data=df)
abline(q2)
```



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

```
q2res <- q2$residuals
```

4. Write the prediction equation.

$$y = 0.023837 * x + 0.507583$$

Question 3

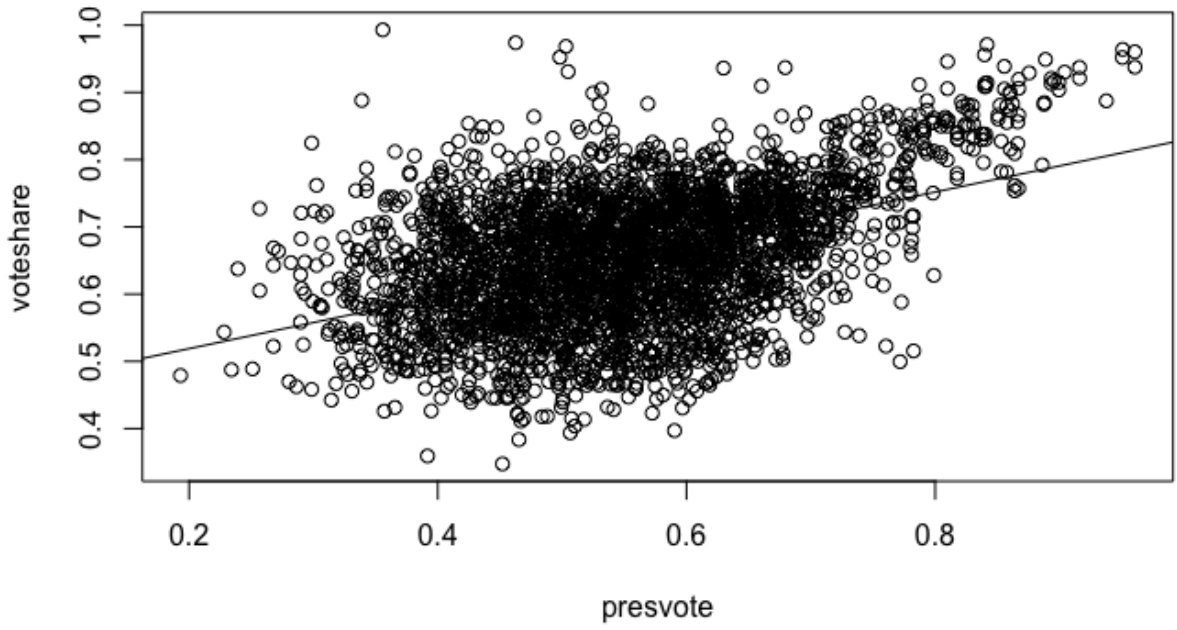
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
q3 <- lm(voteshare ~ presvote, data=df)
summary(q2)
```

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

```
plot(voteshare ~ presvote, data=df)
abline(q3)
```



3. Write the prediction equation.

$$y = 0.388018 * x + 0.441330$$

Question 4

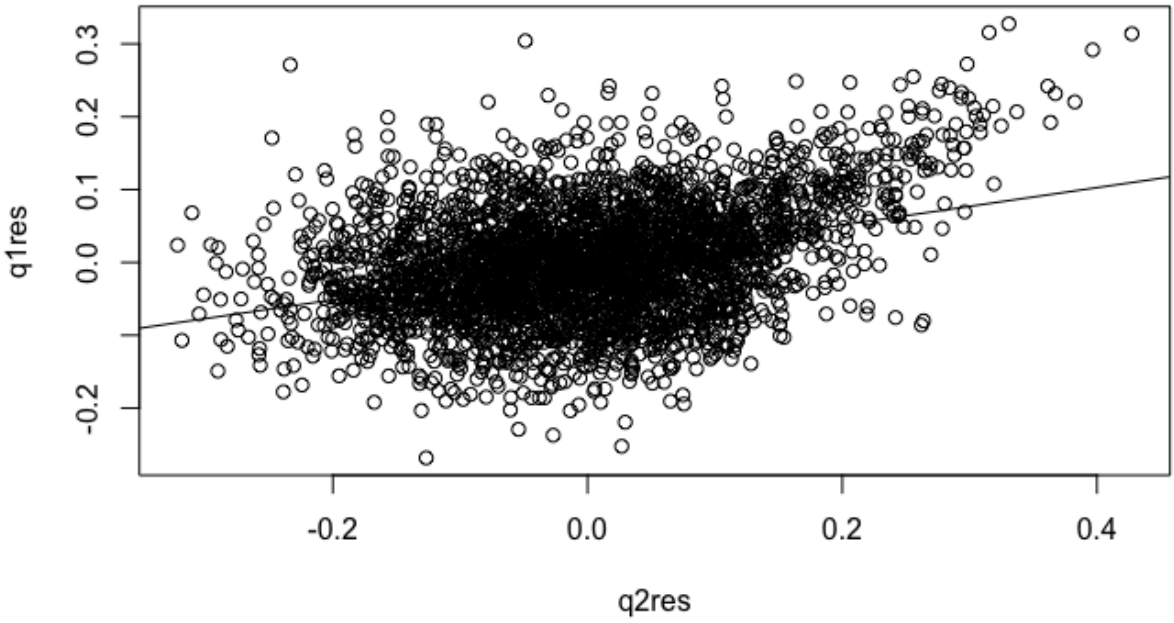
The residuals from part (a) tell us how much of the variation in `voteshare` is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in `presvote` is *not* explained by the difference in spending between incumbent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

```
q4 <- lm(q1res ~ q2res)
summary(q4)
```

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

```
plot(q1res ~ q2res)
abline(q4)
```



3. Write the prediction equation.

$$y = 2.569e - 01 * x + -4.860e - 18$$

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`.

   ```
   q5 <- lm( voteshare ~ difflog + presvote, data=df )
   summary(q5)
   ```

2. Write the prediction equation.

   $$y = 0.0355431 * x1 + 0.2568770 * x2 + 0.4486442$$

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