

Problem Set 3: Applied Stats/Quant Methods 1

Luna Goldstein

Due: November 12, 2021

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 in .pdf form.
- This problem set is due before class on Friday November 12, 2021. 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.

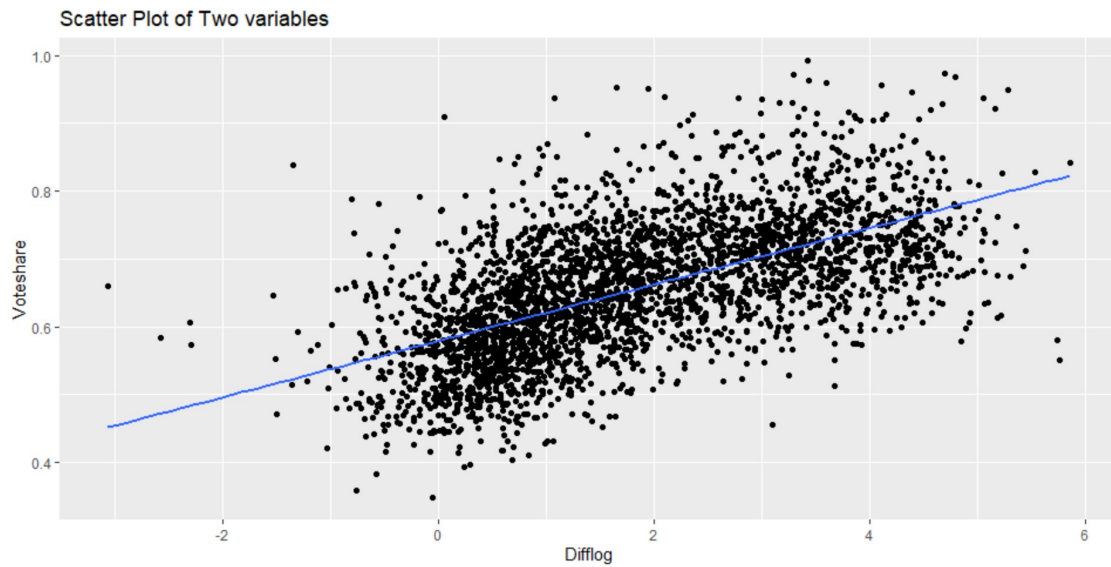
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
1 campaign <- lm(incumbents$voteshare ~ incumbents$difflog)
2 summary(campaign)
```

The output of the Regression :

Y-intercept = 0.579031,
Coefficient estimate = 0.041666,

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



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

```
1 campaign_resid <- resid(campaign)
2 campaign_resid
```

4. Write the prediction equation.

$$\text{voteshare} = 0,579 + 0,042 * \text{difflog}$$

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.

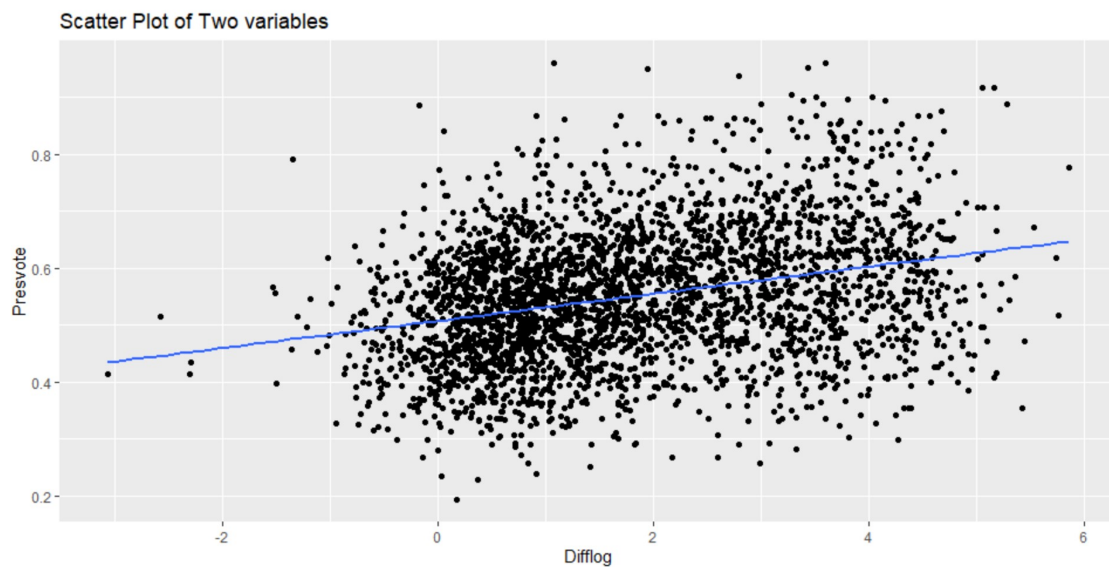
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```
1 spending <- lm(incumbents$presvote ~ incumbents$difflog)
2 summary(spending)
```

The output of the Regression :

Y-intercept = 0.507583,
Coefficient estimate = 0.023837,

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



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

```
1 spending_resid <- resid(spending)
2 spending_resid
```

4. Write the prediction equation.

$$\text{presvote} = 0,508 + 0,024 * \text{difflog}$$

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.

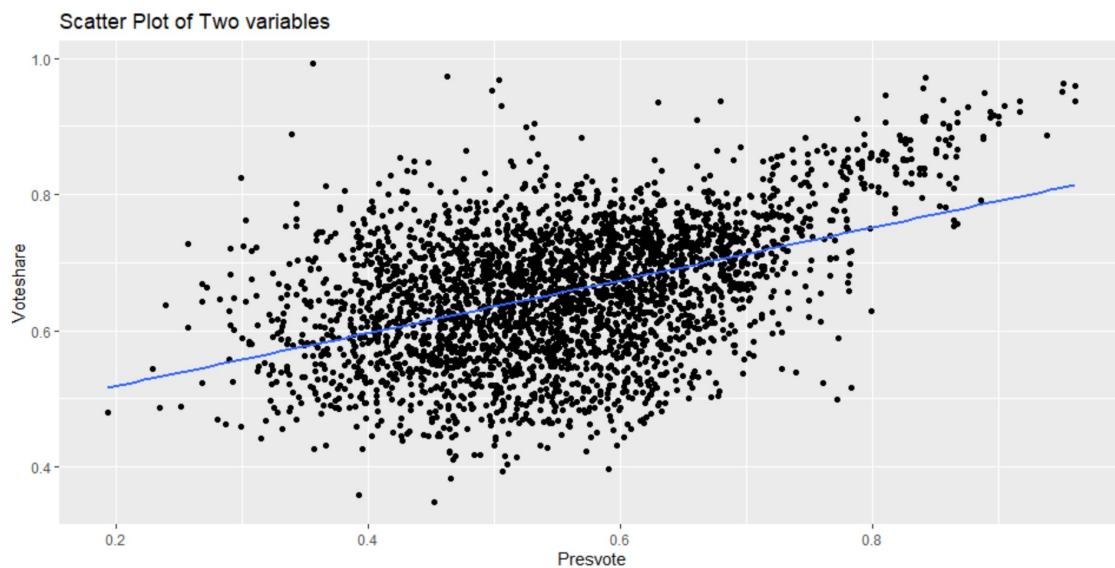
1. Run a regression where the outcome variable is **voteshare** and the explanatory variable is **presvote**.

```
1 pres_vote <- lm(incumbents$voteshare ~ incumbents$presvote)
2 summary(pres_vote)
```

The output of the Regression :

Y-intercept = 0.441330,
Coefficient estimate = 0.3880187,

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



3. Write the prediction equation.

$$\text{voteshare} = 0,508 + 0,024 * \text{Presvote}$$

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.

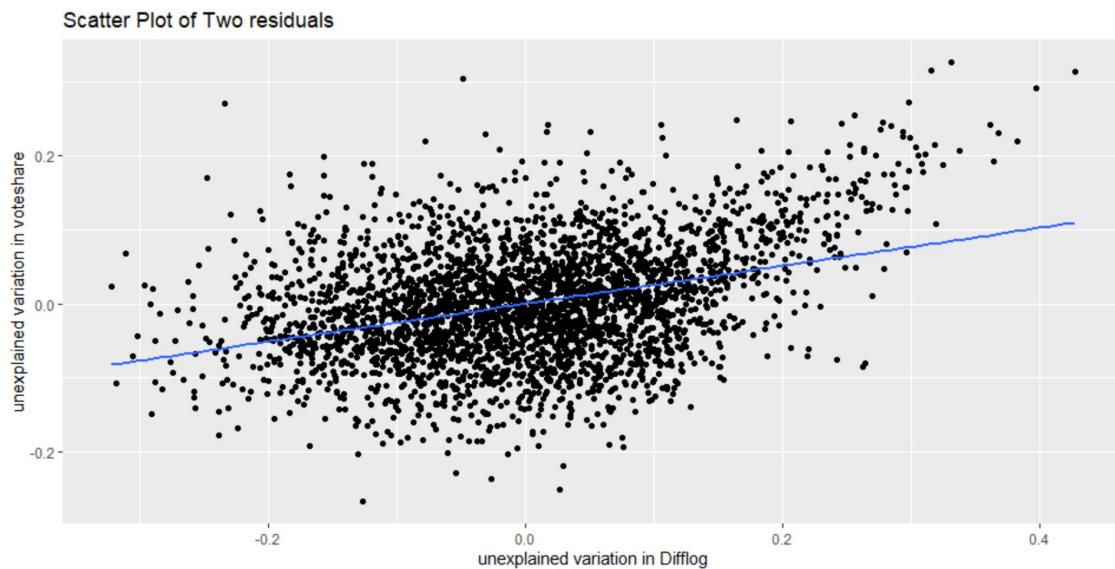
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
1 ques_4 <- lm(campaign_resid ~ spending_resid)
2 summary(ques_4)
```

The output of the Regression :

Y-intercept = -0.00000000000000000486,
Coefficient estimate = 0.25687701270009788423,

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



3. Write the prediction equation.

$$\text{campaignresid} = (-4.860\text{e-}18) + (2.569\text{e-}01) * \text{spendingresid}$$

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 model <- lm(voteshare ~ difflog + presvote, data = incumbents)
2 summary(model)
```

The output of the Regression :

Y-intercept = 0.4486442,
Coefficient estimate = 0.0355431,
Regression coefficient = 0.2568770,

2. Write the prediction equation.

$$\text{voteshare} = 0.44864 + 0.03554 * \text{difflog} + 0.25688 * \text{presvote}$$

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 output is identical to the coefficient estimate of question 4. The coefficient estimate tells us how much of the variation in `presvote` is not explained by the difference in spending between incumbent and challenger in the district. This suggests that when adding another independent variable gives us an identical value.