

Problem Set 3

Applied Stats/Quant Methods 1

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

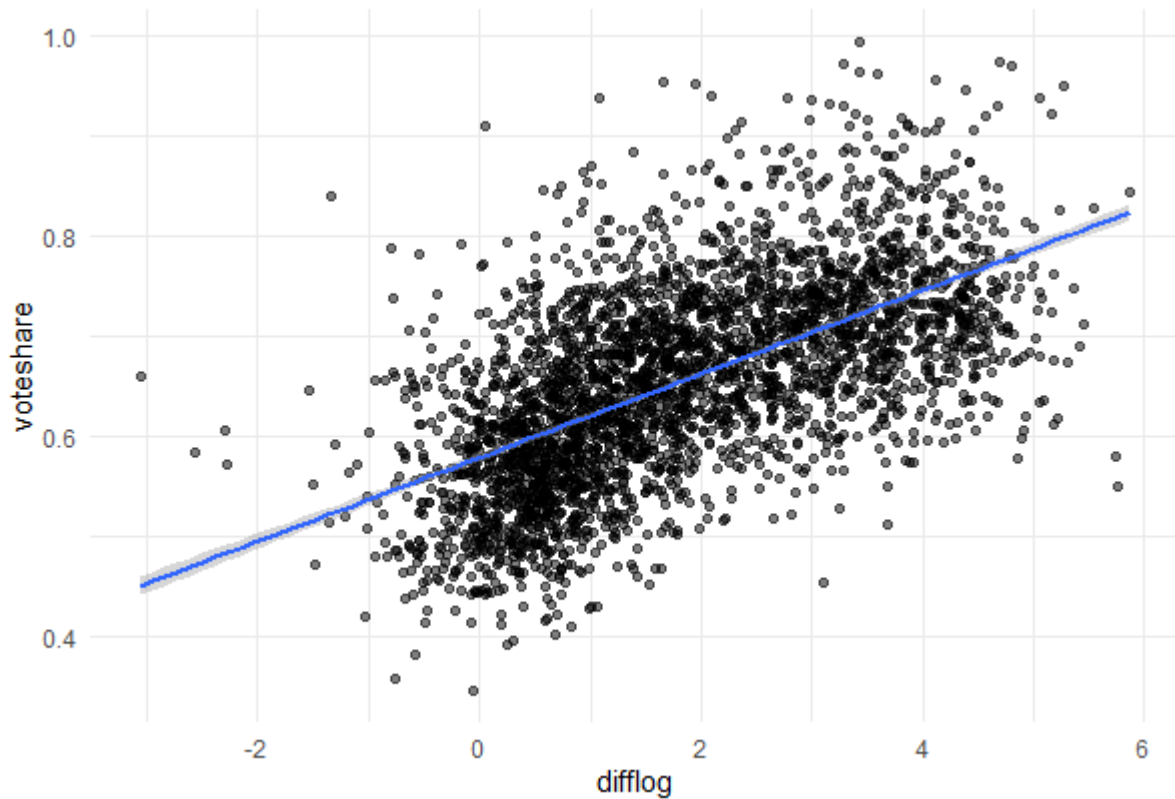
We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

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

```
1 reg1 <- lm (data=df, voteshare ~ difflog)
2 summary(reg1) #checking results of the regression
```

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

```
1 ggplot(data = df, aes(x = difflog, y = voteshare)) +
2   geom_point(alpha = 0.5) + #scatterplot
3   geom_smooth(method = "lm") + #regression line
4   theme_minimal()
```



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

```
1 resid1 <- reg1$residuals
```

4. Write the prediction equation.

$$\text{voteshare}(Y) = 0.579031(\text{Intercept}) + 0.041666 \times \text{difflog}(X1)$$

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**.

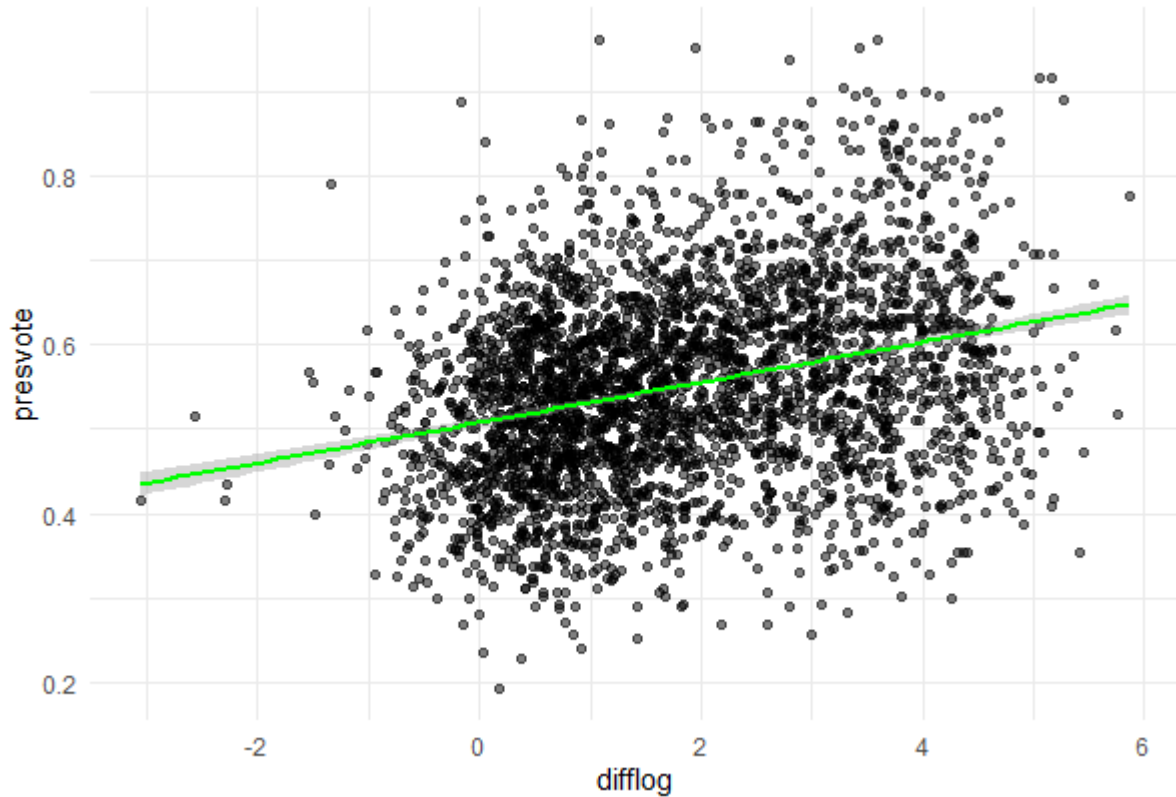
```
1 reg2 <- lm (data=df, presvote ~ difflog)
2 summary(reg2) #checking results of the regression
```

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

```

1 ggplot(data = df, aes(x = difflog, y = presvote)) +
2   geom_point(alpha = 0.5) + #scatterplot
3   geom_smooth(method = "lm", colour = "green") + #regression line
4   theme_minimal()

```



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

```

1 resid2 <- reg2$residuals

```

4. Write the prediction equation.

$$\text{presvote}(Y) = 0.507583(\text{Intercept}) + 0.023837 \times \text{difflog}(X1)$$

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**.

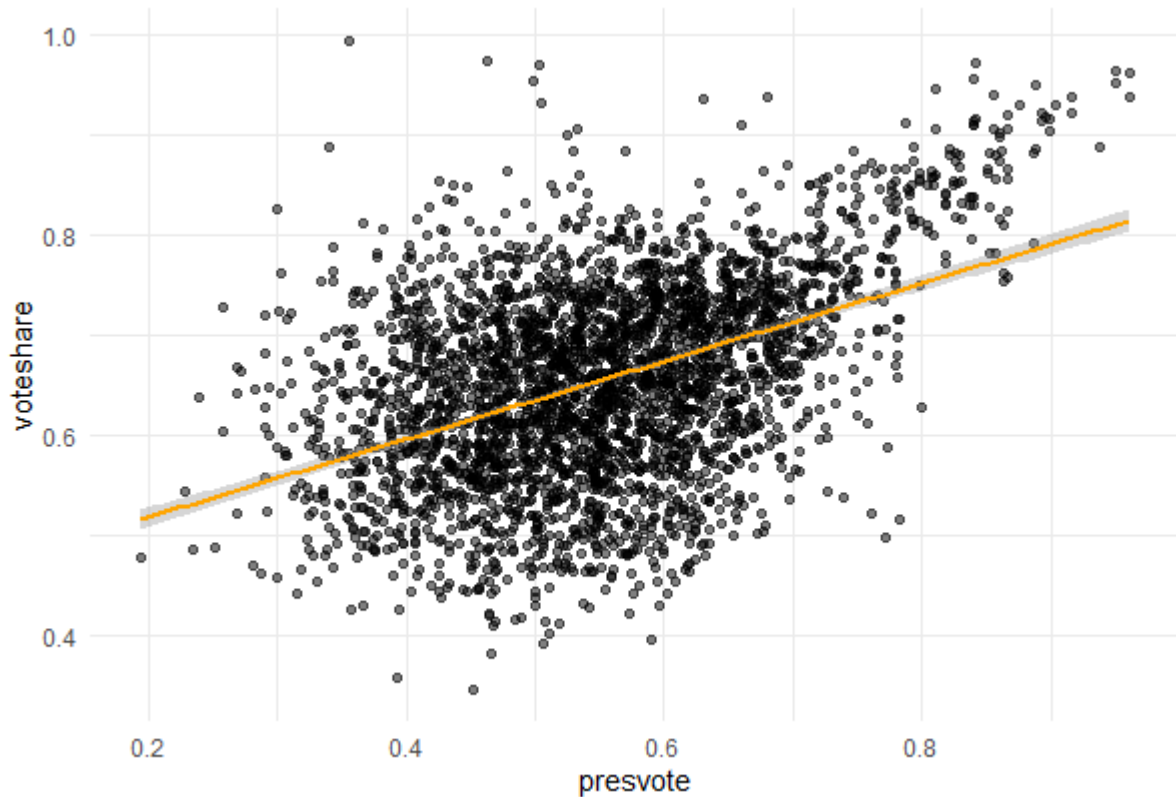
```

1 reg3 <- lm (data=df, voteshare ~ presvote)
2 summary(reg3) #checking results of the regression

```

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

```
1 ggplot(data = df, aes(x = presvote, y = voteshare)) +
2 geom_point(alpha = 0.5) + #scatterplot
3   geom_smooth(method = "lm", colour = "orange") + #regression line
4   theme_minimal()
```



3. Write the prediction equation.

$$\text{voteshare (Y)} = 0.441330 \text{ (Intercept)} + 0.388018 \times \text{presvote(X1)}$$

Question 4

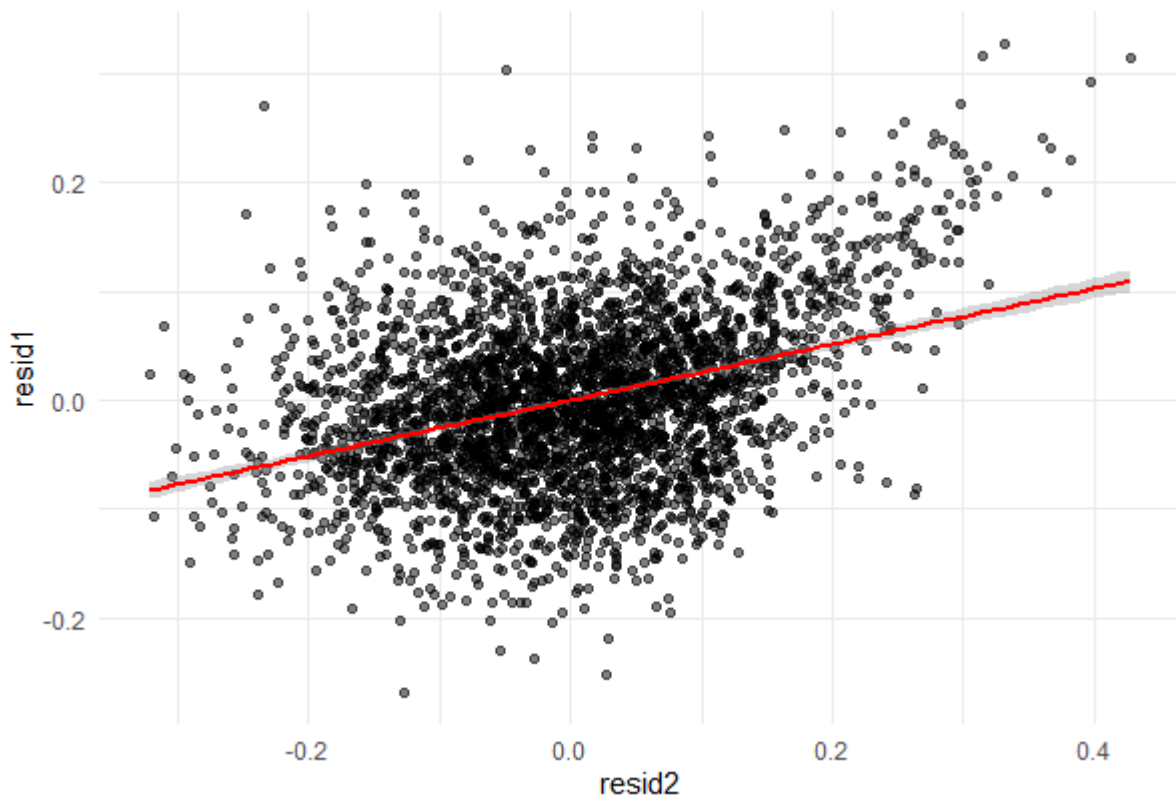
The residuals from part (a) tell us how much of the variation in **voteshare** is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in **presvote** is *not* explained by the difference in spending between incumbent 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 reg4 <- lm(data=df, resid1 ~ resid2)
2 summary(reg4) #checking results of the regression
```

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

```
1 df_resid <- as.data.frame(cbind(resid1, resid2)) # data frame of
  residuals from both regressions
2
3 ggplot(data = df_resid, aes(x = resid2, y = resid1)) +
4   geom_point(alpha = 0.5) + #scatterplot
5   geom_smooth(method = "lm", colour = "red") + #regression line
6   theme_minimal()
```



3. Write the prediction equation.

Residuals from regression 1(Y) = -5.207e-18 (Intercept) + 2.569e-01 x Residuals from regression 2(X1)

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 reg5 <- lm (data=df, voteshare ~ difflog + presvote)
2 summary(reg5) #checking results of the regression
```

2. Write the prediction equation.

$$\text{Incumbemnt's vote share (Y)} = 0.4486442 \text{ (Interecept)} + 0.0355431 \times \text{difflog (X1)} + 0.2568770 \times \text{presvote (X2)}$$

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 from the regressions in questions 4 and 5 are the same (0.07338 in Q4 and 0.07339 in Q5). The only difference is that in Q5 there is one less degree of freedom because the equation has one more coefficient (the second covariant, X2). The same residual standard error in this case occurs because in both cases we calculated the variation of incumbent's vote share that is not explained by the variation in the difference in spending between incumbent and challenger nor by the variation in vote share of the presidential candidate of the incumbent's party.