

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 lm(voteshare ~ difflog, data = incumbent)
2 lm1 <- lm(voteshare ~ difflog, data = incumbent)

```

Coefficients: Intercept - 0.57903 difflog - 0.04167

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

```

1 plot(incumbent$voteshare, incumbent$difflog,
2       ylab="Incumbent Voter Share",
3       xlab="Incumbent/Challenger Spending",
4       main="Incumbent vs Challenger Spending")
5
6 abline(lm(incumbent$voteshare ~ incumbent$difflog), col = "red")

```

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

```

1 incumbent$Resids <- residuals(lm1)
2 incumbent$Resids

```

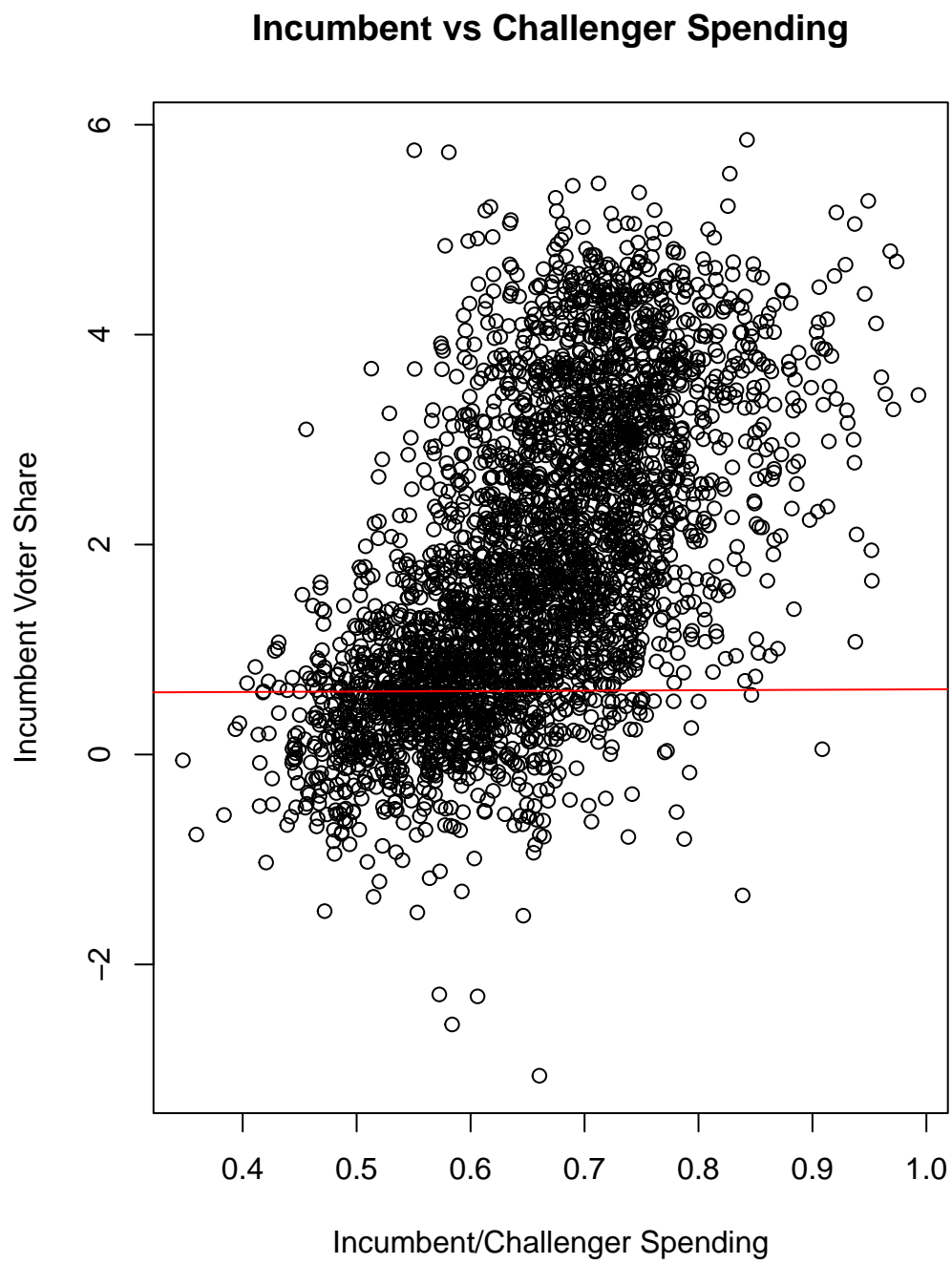


Figure 1: Scatterplot Vote Share Incumbent and Challenger Spend

4. Write the prediction equation.

```
1 coef(lm1) # I use this to get the slope of 0.041
```

$$\mu_y = B_0 + B_1 X_1$$

$$\text{Voteshare} = 0.57903 + 0.04166 * \text{Difflog}$$

Data from code Intercept - 0.57903071 difflog - 0.04166632

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 lm(presvote ~ difflog, data = incumbent)
2 lm2 <- lm(presvote ~ difflog, data = incumbent)
```

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

```
1 plot(incumbent$presvote, incumbent$difflog,
2       xlab="Incumbent/Challenger spending",
3       ylab="Presidential Vote Share",
4       main="Incumbent vs Challenger Spending")
5
6 abline(lm(incumbent$presvote ~ incumbent$difflog), col = "Orange")
```

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

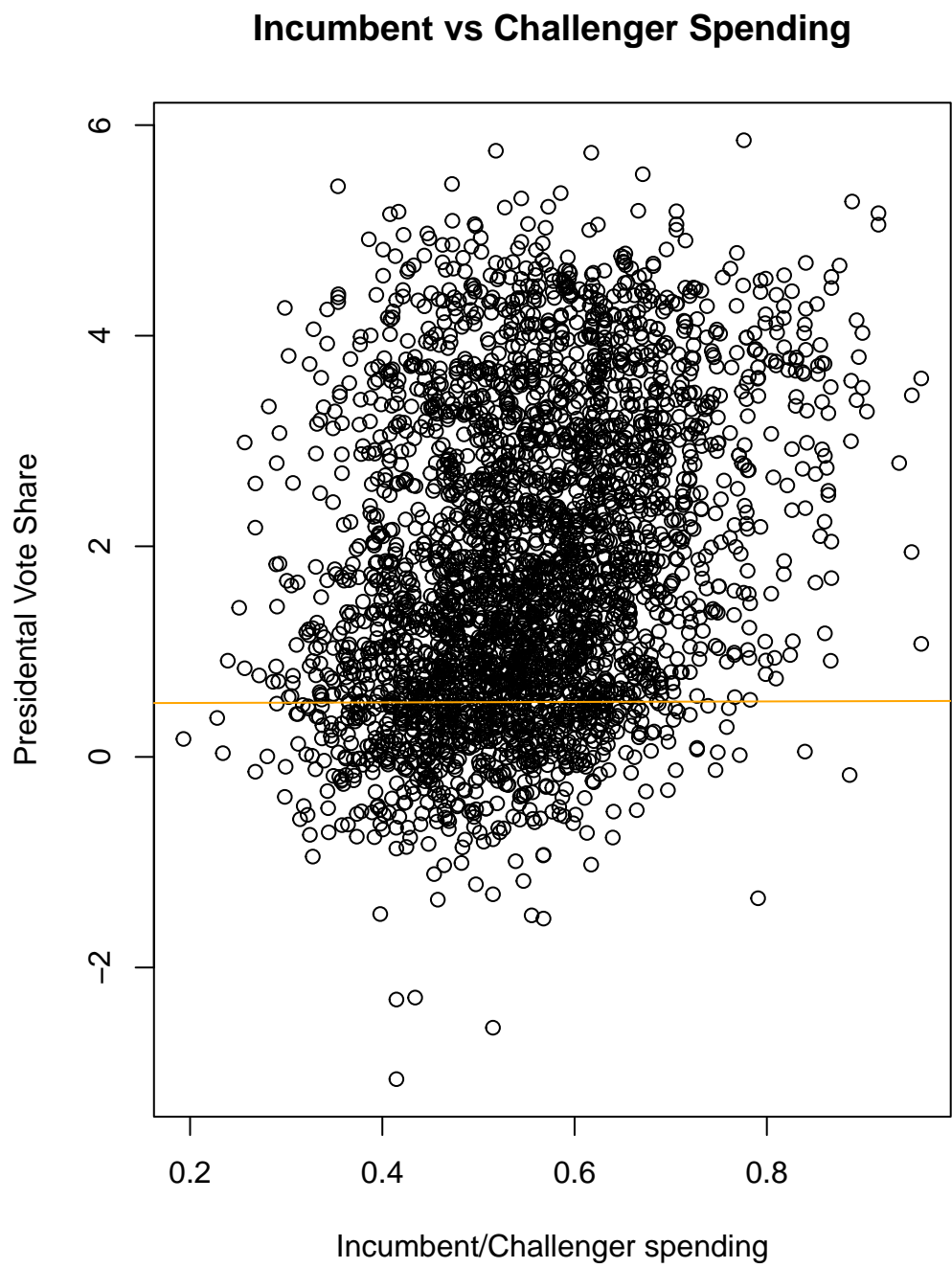


Figure 2: Scatterplot Presidential Vote Incumbent and Challenger Spend

```
1 incumbent$Resids2 <- residuals(lm2)
2 incumbent$Resids2
```

4. Write the prediction equation.

```
1 coef(lm2) # I use this to get the slope of 0.0238
```

$$\mu_y = B_0 + B_1X_1$$

$$\text{Presvote} = 0.507583 + 0.023837 * \text{Difflog}$$

Data from code Intercept - 0.50758333 difflog - 0.02383723

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 lm(voteshare ~ presvote, data = incumbent)
2 lm3 <- lm(voteshare ~ presvote, data = incumbent)
```

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

```
1 plot(incumbent$voteshare, incumbent$presvote,
2       ylab="Incumbent Voter Share",
3       xlab="Presidential Vote Share",
4       main="Incumbent vs Challenger Spending")
5
6 abline(lm(incumbent$voteshare ~ incumbent$presvote), col = "Blue")
```

3. Write the prediction equation.

```
1 coef(lm3) # I use this to get the slope of 0.388
```

$$\mu_y = B_0 + B_1X_1$$

$$\text{Voteshare} = 0.44132 + 0.38801 * \text{Prevote}$$

Incumbent vs Challenger Spending

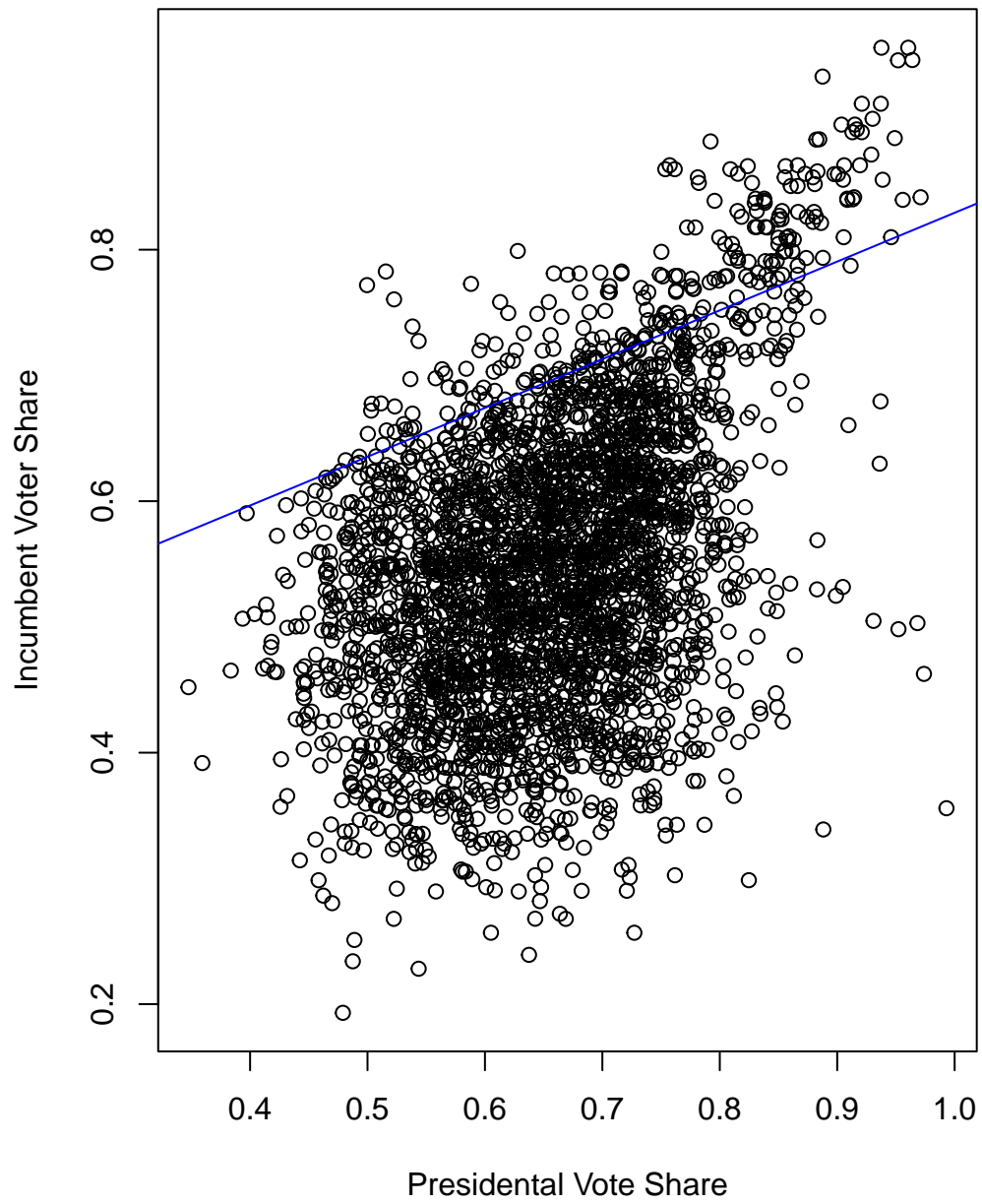


Figure 3: Scatterplot Presidential Vote Incumbent and Challenger Spend

Data from code

Intercept - 0.4413299 presvote - 0.3880184

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 lm(incumbent$Resids ~ incumbent$Resids2, data = incumbent)
```

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

```
1 plot(incumbent$Resids, incumbent$Resids2,
2       ylab="Vote Share / Spending",
3       xlab="Presidential Vote Share / Spending",
4       main="Incumbent vs Challenger Spending")
5 abline(lm(incumbent$Resids ~ incumbent$Resids2), col = "Green")
```

3. Write the prediction equation.

Incumbent vs Challenger Spending

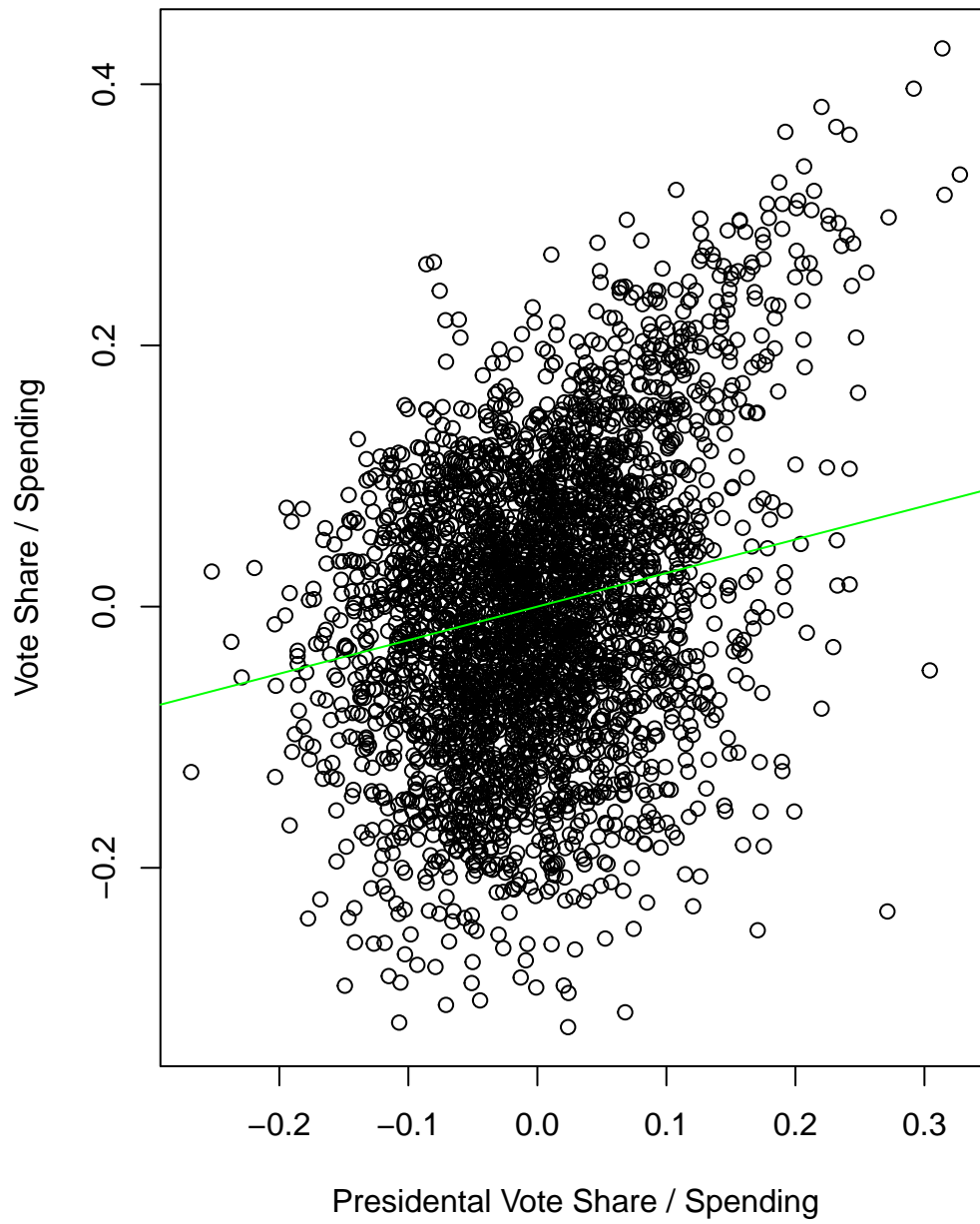


Figure 4: Scatterplot Presidential Vote Incumbent and Challenger Spend

```
1 coef(lm(incumbent$Resids ~ incumbent$Resids2))
```

$\mu y = B0 + B1X1$

RegVoteShareDifflog = -4.8596 + 2.5687 * RegPresVoteDifflog

Data from code

Intercept - (-4.859631e) incumbentResids2 - 2.568770e-01

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 lm(voteshare ~ difflog * presvote, data = incumbent)
```

2. Write the prediction equation.

$$\mu_y = B_0 + B_1X_1 + B_2X_2$$

$$\text{IncumbentVotshare} = 0.46371 + 0.02791 * \text{Difflog} + 0.22815 * \text{Presvote}$$

Data from code

Intercept - 0.46371 difflog - 0.02791 presvote - 0.22815 difflog:presvote - 0.01393

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

Like Question 4, this question is investigating more than one explanatory variable (Difflog and Presvote), where Question 4 also had 2 explanatory variables (Presvote and

Difflog) also known as the second question. As we can see they are both investigating these two factors as explanatory variables. These two outputs will be similar as adding additional controls positively correlated to voteshare will lower the slope.