

Problem Set 3

Applied Stats/Quant Methods 1

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`.

```
1 # Creating model using lm (I have used lm a lot in the past):  
2 modell <- lm(voteshare ~ difflog, inc.sub)
```

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

```
1 # Saving residuals separately:
2 resid1 <- model1$residuals
3
4 # And creating the scatterplot using ggplot:
5 plot1 <- ggplot(inc.sub, aes(x = difflog, y = voteshare)) +
6   geom_point(color = "darkslategray4", alpha = 0.7) + # Editing point
   color and transparency
7   geom_smooth(method = "lm", se = FALSE, color = "grey") + # Adding
   regression line (no SE)
8   theme_minimal() + # Applying theme
9   labs(x = "Difference in campaign spending between incumbent and
   challenger",
```

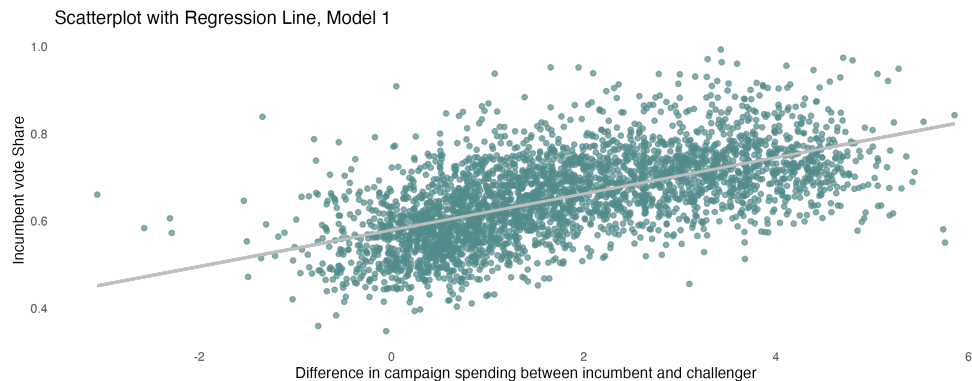


Figure 1: Model 1 Plot

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

```
1 # Exploring model
2 summary(model1)
```

4. Write the prediction equation.

$$\text{VoteSh} = \beta_0 + \beta_1 \cdot \text{DiffLog} + \varepsilon$$

where:

- *VoteSh*: Incumbent's vote share
- *DiffLog*: Logarithm of the difference between incumbent and challenger's spending

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  
2 #Saving plot:
```

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

```
1 # Creating and exploring model:  
2 model2 <- lm(presvote ~ difflog, inc.sub)  
3 summary(model2)  
4  
5 # Saving residuals separately:  
6 resid2 <- model2$residuals  
7  
8 # And creating the scatterplot using ggplot:  
9 plot2 <- ggplot(inc.sub, aes(x = difflog, y = presvote)) +
```

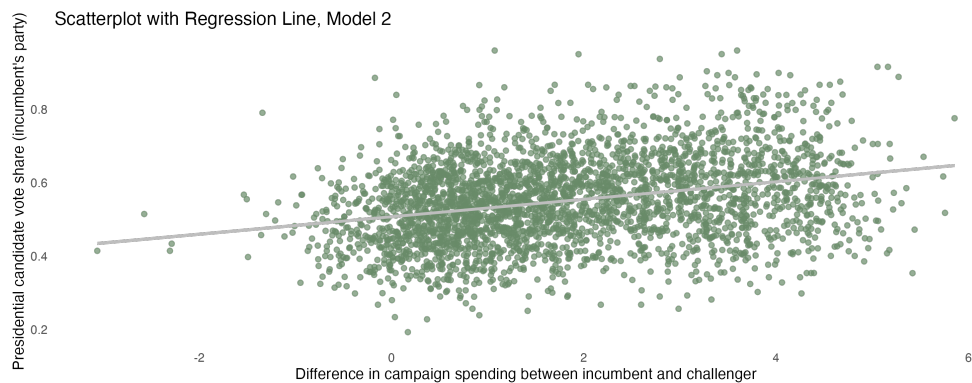


Figure 2: Model 2 Plot

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

```
1  
2 ##### Question 2 #####
```

4. Write the prediction equation.

$$\text{PresSh} = \beta_0 + \beta_1 \cdot \text{DiffLog} + \varepsilon$$

where:

- *PresSh*: Vote share of the presidential candidate of the incumbent's party
- *DiffLog*: Logarithm of the difference between incumbent and challenger's spending

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 theme(panel.grid = element_blank()) # Getting rid of the grid
```

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

```
1 ggsave("plot2.png", plot = plot2, dpi = 300)
2
3
4 ##### Question 3 #####
5
6 # Creating and exploring model:
7 model3 <- lm(voteshare ~ presvote, inc.sub)
8 summary(model3)
```

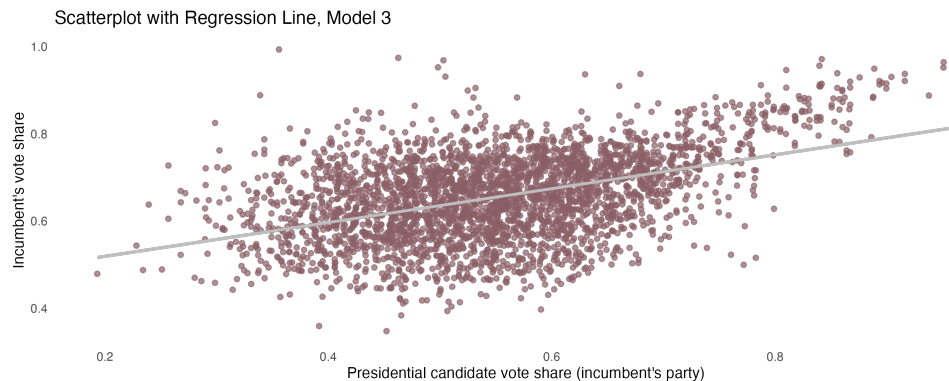


Figure 3: Model 3 Plot

3. Write the prediction equation.

$$\text{VoteSh} = \beta_0 + \beta_1 \cdot \text{PresSh} + \varepsilon$$

where:

- *VoteSh*: Incumbent's vote share
- *PresSh*: Vote share of the presidential candidate of the incumbent's party

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 geom_point(color = "lightpink4", alpha = 0.7) + # Editing point color
  and transparency
2 geom_smooth(method = "lm", se = FALSE, color = "grey") + # Adding
  regression line (no SE)
```

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

```
1 labs(x = "Presidential candidate vote share (incumbent's party)",
2      y = "Incumbent's vote share",
3      title = "Scatterplot with Regression Line, Model 3") + # Fixing
  labels
4 theme(panel.grid = element_blank()) # Getting rid of the grid
5
6 #Saving plot:
7 ggsave("plot3.png", plot = plot3, dpi = 300)
8
9 ##### Question 4 #####
```

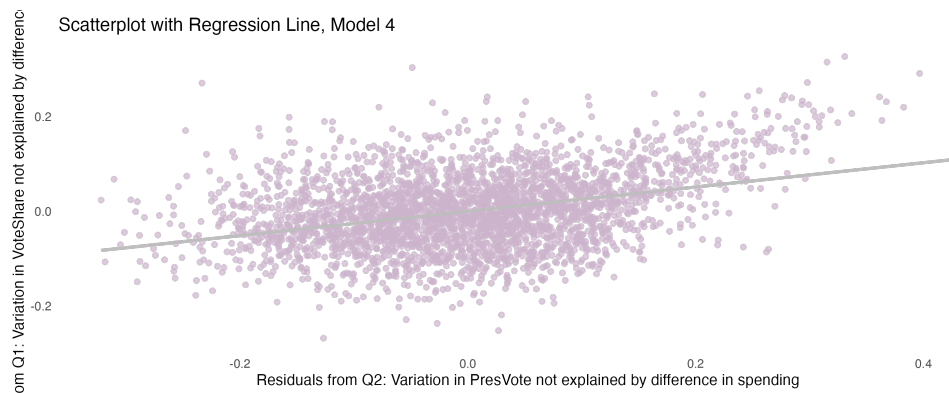


Figure 4: Model 4 Plot

3. Write the prediction equation.

$$\text{Resid1} = \beta_0 + \beta_1 \cdot \text{Resid2} + \varepsilon$$

where:

- *Resid1*: Residuals from first model (variation in incumbent's vote share not explained by difference in spending)
- *Resid2*: Residuals from second model (variation in presidential candidate's vote share not explained by difference in spending)

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