Problem Set 3

Applied Stats/Quant Methods 1

Due: November 13, 2025

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 Thursday November 13, 2025. 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`.

Table 1: Incumbent Vote Share explained by Difference of Spending

	Dependent variable:
	voteshare
	Coefficients (biregress1)
difflog	0.042***
	(0.001)
Constant	0.579***
	(0.002)
Observations	3,193
R^2	0.367
Adjusted R^2	0.367
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,852.791*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

Plot of Difference in Spending and Vote Share
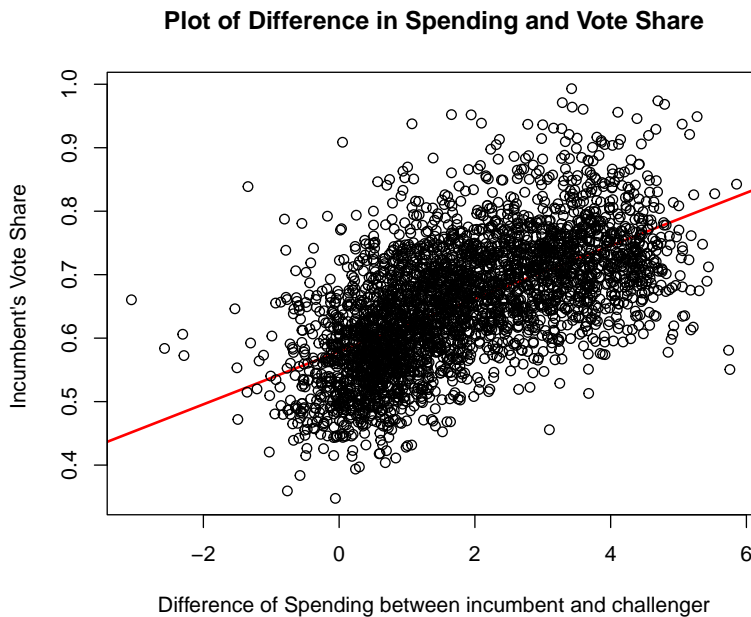
The model shows a moderate positive association between difference in spending and incumbent's vote share. The R squared value is 0.3673, therefore, the difference of spending isn't a strong predicition on the outcome of the incumbent's vote share. This means 36.73 percent of variation in the incumbent's vote share is explained by the difference in spending. The remaining 63.27 percent of variation is due to other factors that are not included in the model.

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

```
res_1 <- biregress_1$residuals
```

4. Write the prediction equation.

$$\widehat{\text{voteshare}} = 0.579 + 0.042 \times \text{difflog} \tag{1}$$

The slope of this equation means that for every 1 unit increase in difference of spending, the incumbent's vote share is predicted to increase 0.042. The constant of 0.579 is the predicted value of the incumbent's vote share when the difference in spending is equal to zero.

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`.

3

Table 2: Presidential Candidate Vote Share explained by Difference of Spending

	Dependent variable:
	presvote
	Coefficients (biregress 2)
difflog	0.024***
	(0.001)
Constant	0.508***
	(0.003)
Observations	3,193
R^2	0.088
Adjusted R^2	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

Difference in Spending and Vote Share of the Incumbent Presidential
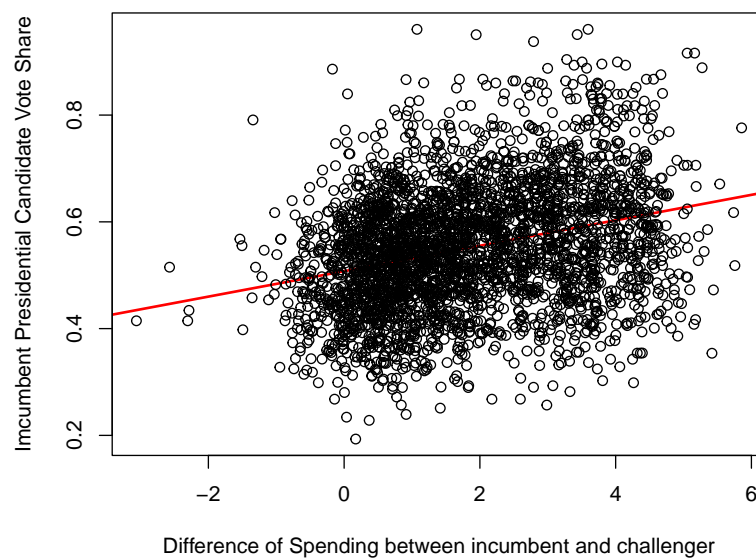
The model shows a weak positive association, because this model is more clustered, between difference in spending and the vote share of the presidential candidate of the incumbent's party. The R squared value is 0.08795, therefore, the difference of spending isn't a strong predicition on the outcome of the vote share of the presidential candidate. This means 8.795 percent of variation in the incumbent's vote share is explained by the difference in spending. The remaining 91.205 percent of variation is due to other factors that are not included in the model.

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

```
res_2 <- biregress_2$residuals
```

4. Write the prediction equation.

$$\widehat{\text{presvote}} = 0.508 + 0.024 \times \text{difflog} \tag{2}$$

The slope of this equation means that for every 1 unit increase in difference of spending, the vote share of the presidential candidate is predicted to increase 0.024. The constant of 0.508 is the predicted value of the vote share of the presidential candidate when the difference in spending is equal to zero.

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`.

Table 3: Incumbents Vote Share explained by Incumbent Presidential Candidate Vote Share

	Dependent variable:
	voteshare
	Coefficients (biregress 3)
presvote	0.388***
	(0.013)
Constant	0.441***
	(0.008)
Observations	3,193
R^2	0.206
Adjusted R^2	0.206
Residual Std. Error	0.088 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

Relationship Between Presidential and Incumbent Party Vote Shar
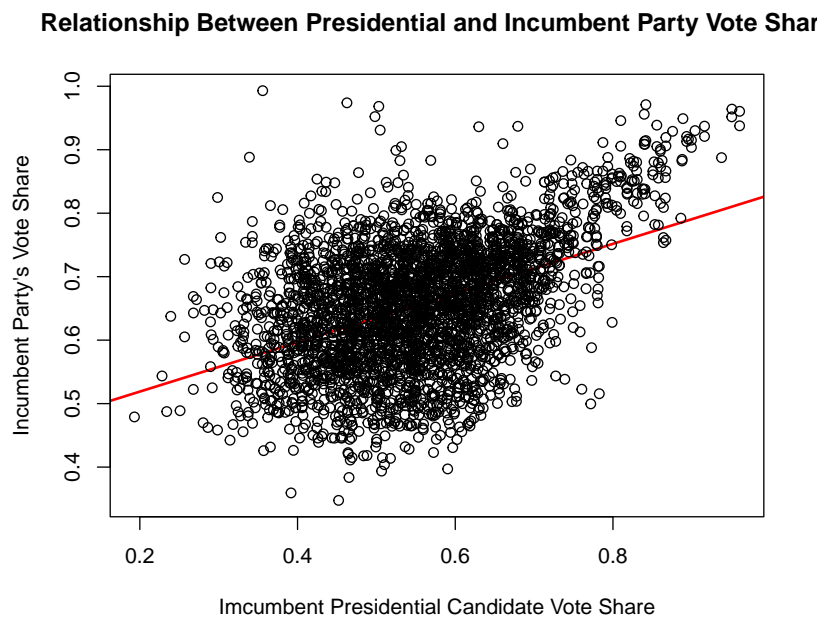
The model shows a moderate positive association between the vote share of the presidential candidate of the incumbent's party and the incumbent's vote share. The R squared value is 0.2058, therefore, the difference of spending isn't a strong predicition on the outcome of the incumbent's vote share. This means 20.58 percent of variation in the incumbent's vote share is explained by the vote share of presidential candidate of the incumbent's party. The remaining 79.42 percent of variation is due to other factors that are not included in the model.

3. Write the prediction equation.

$$\widehat{\text{voteshare}} = 0.388 + 0.441 \times \text{presvote} \tag{3}$$

The slope of this equation means that for every 1 unit the vote share of the presidential candidate, the incumbent's vote share is predicted to increase by 0.388. The constant of 0.441is the predicted value of the incumbent's vote share when the the vote share of the presidential candidate to zero.

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.

Table 4: Residuals of Incumbents Vote Share explained by Residuals of Incumbent Presidential Candidate Vote Share

	Dependent variable:
	res_1
	Coefficients (residuals)
res_2	0.257***
	(0.012)
Constant	−0.000
	(0.001)
Observations	3,193
R^2	0.130
Adjusted R^2	0.130
Residual Std. Error	0.073 (df = 3191)
F Statistic	476.975*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

Plot of Residuals from Question 1 and 2

8
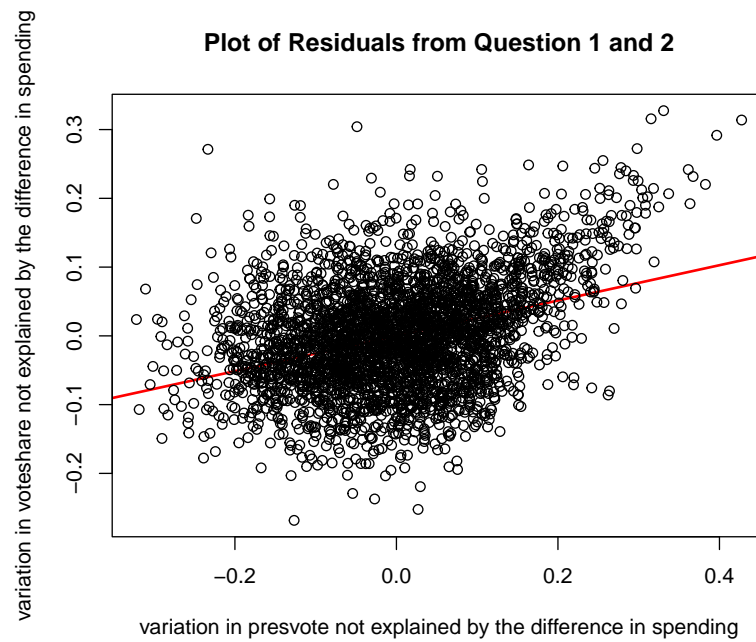
3. Write the prediction equation.

$$\widehat{\text{variation in voteshare}} = -1.942e^{-18} + 0.257 \times \text{variation in presvote} \qquad (4)$$

The constant value $-1.942e^{-18}$ is the predicticted value of variation in voteshare is not explained by the difference in spending when the variation in presvote is not explained by the difference in spending equals zero. For every one unit increase in the variation in presvote not explained by the difference in spending, the variation in voteshare is not explained by the difference in spending is predicted to increase by 0.257.

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`.

Table 5: Incumbents Vote Share explained by Difference in Spending and Incumbent Presidential Candidate Vote Share

	Dependent variable:
	voteshare
	Coefficients (residuals)
difflog	0.036***
	(0.001)
presvote	0.257***
	(0.012)
Constant	0.449***
	(0.006)
Observations	3,193
R^2	0.450
Adjusted R^2	0.449
Residual Std. Error	0.073 (df = 3190)
F Statistic	1,302.947*** (df = 2; 3190)
Note:	*p<0.1; **p<0.05; ***p<0.01

9

2. Write the prediction equation.

$$\widehat{\text{voteshare}} = 0.449 + 0.036 \times \text{difflog} + 0.257 \times \text{presvote} \tag{5}$$

The constant 0.449 is the predicted value of incumbent's vote share given difference in spending and the vote share of the presidential candidate equals zero. On average, a one unit increase in difference of spending leads to a 0.036 increase in incumbent's vote share, holding everything else constant. On average, a one unit increase in the vote share of the presidential candidate leads to a 0.257 increase in incumbent's vote share, holding everything else constant.

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 that's identical to the output in Question 4 is the slope for the vote share of the presidential candidate. In Question 4, the model was looking at the variation not explained by difference in spending for both incumbent's vote share and the vote share of the presidential canidate. In this model, we are using the explanatory values of difference of spending and vote share of the presidential canidate to explain the outcome variable of incumbent's vote share. Therefore, the value for vote share of the presidential canidate needs to be a unique contrabution and seperate from any variation it shares with difference in spending.