

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`.

The code shows the regression outcomes of incumbent's vote share given the explanatory variable of difference in campaign spending.

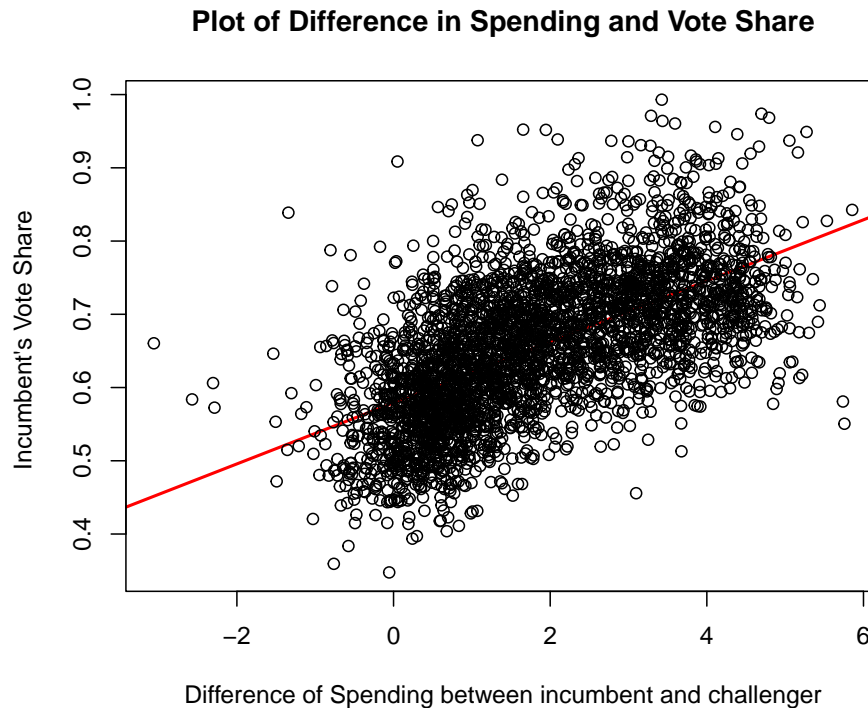
```
1 biregress_1 <- lm(voteshare ~ difflog, data = inc.sub)
2 summary(biregress_1)
3 #Residuals:
4 #   Min       1Q   Median       3Q      Max
5 #-0.26832 -0.05345 -0.00377  0.04780  0.32749
6
7 #Coefficients:
```

```

8 # Estimate Std. Error t value Pr(>|t|)
9 #(Intercept) 0.579031 0.002251 257.19 <2e-16 ***
10 # difflog 0.041666 0.000968 43.04 <2e-16 ***
11
12 #Residual standard error: 0.07867 on 3191 degrees of freedom
13 #Multiple R-squared: 0.3673, Adjusted R-squared: 0.3671
14 #F-statistic: 1853 on 1 and 3191 DF, p-value: < 2.2e-16

```

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



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 prediction 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.

```

1 res_1 <- biregress_1$residuals

```

This code shows the assignment of the residuals to a new object in R studio. The dollar sign pulls just the residuals from the regression model to put into the object.

4. Write the prediction equation.

$$Y_i = 0.579031 + 0.041666X_i + \epsilon_i \quad (1)$$

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 0.041666. The constant of 0.579031 is the predicted value of the incumbent's vote share when the difference in spending is equal to zero.

Question 2

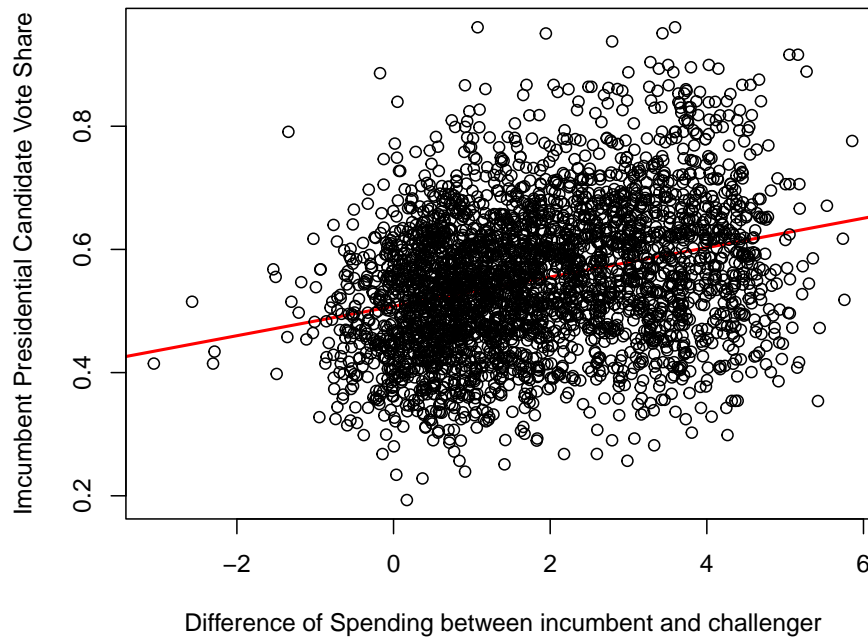
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
1 biregress_2 <- lm(presvote ~ difflog, data = inc.sub)
2 summary(biregress_2)
3 #Residuals:
4 #   Min       1Q   Median       3Q      Max
5 #-0.32196 -0.07407 -0.00102  0.07151  0.42743
6
7 #Coefficients:
8 #   Estimate Std. Error t value Pr(>|t|)
9 #(Intercept)  0.507583   0.003161  160.60  <2e-16 ***
10 #   difflog    0.023837   0.001359   17.54  <2e-16 ***
11
12 #Residual standard error: 0.1104 on 3191 degrees of freedom
13 #Multiple R-squared:  0.08795, Adjusted R-squared:  0.08767
14 #F-statistic: 307.7 on 1 and 3191 DF, p-value: < 2.2e-16
```

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 prediction 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.

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

4. Write the prediction equation.

$$Y_i = 0.507583 + 0.023837X_i + \epsilon_i \quad (2)$$

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 0.023837. The constant of 0.507583 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.

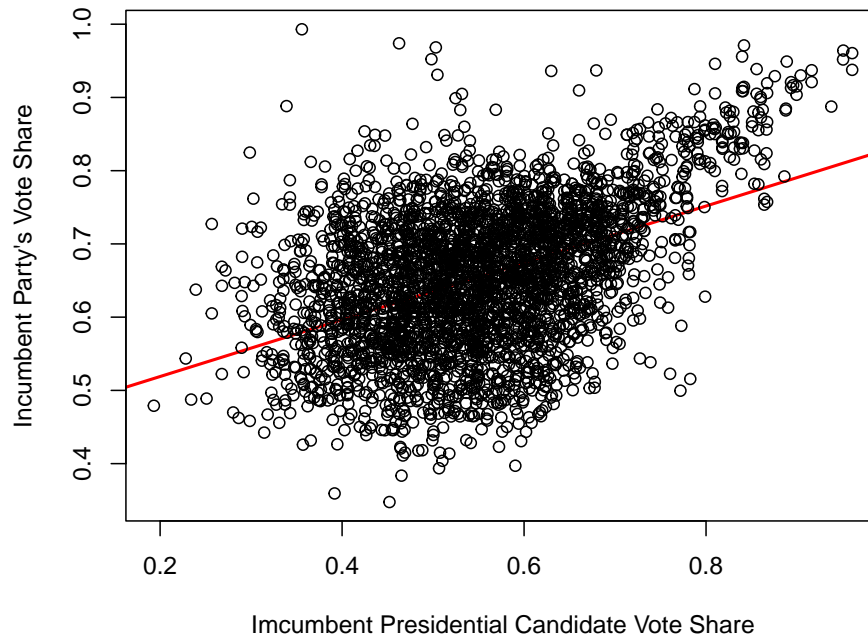
1. Run a regression where the outcome variable is **voteshare** and the explanatory variable is **presvote**.

```
1 biregress_3 <- lm(voteshare ~ presvote, data = inc.sub)
2 summary(biregress_3)
3 #Residuals:
4 #Min      1Q   Median      3Q      Max
5 #-0.27330 -0.05888  0.00394  0.06148  0.41365
6
7 #Coefficients:
8 #   Estimate Std. Error t value Pr(>|t|)
9 #(Intercept) 0.441330    0.007599   58.08  <2e-16 ***
10 #   presvote   0.388018    0.013493   28.76  <2e-16 ***
11
12 #Residual standard error: 0.08815 on 3191 degrees of freedom
13 #Multiple R-squared:  0.2058, Adjusted R-squared:  0.2056
14 #F-statistic:    827 on 1 and 3191 DF,  p-value: < 2.2e-16
```

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

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 prediction 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.

Relationship Between Presidential and Incumbent Party Vote Shar



3. Write the prediction equation.

$$Y_i = 0.441330 + 0.388018X_i + \epsilon_i \quad (3)$$

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 0.388018. The constant of 0.441330 is the predicted value of the incumbent's vote share when the vote share of the presidential candidate is 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.

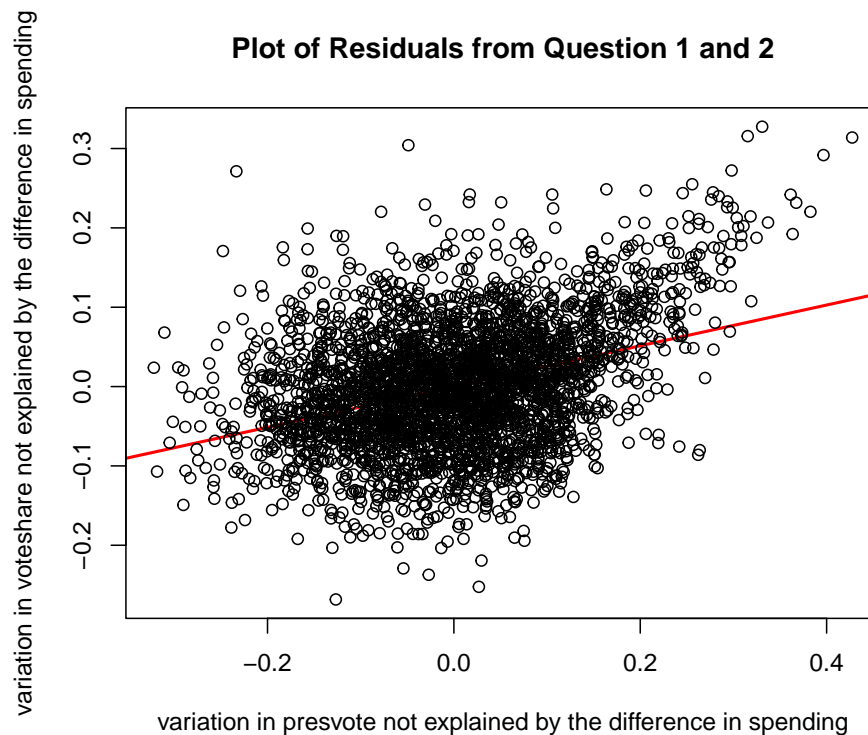
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```

1 resid_regress <- lm(res_1 ~ res_2)
2 summary(resid_regress)
3 #Residuals:
4 #   Min       1Q   Median       3Q      Max
5 #-0.25928 -0.04737 -0.00121  0.04618  0.33126
6
7 #Coefficients:
8 #   Estimate Std. Error t value Pr(>|t|)
9 #(Intercept) -1.942e-18  1.299e-03    0.00      1
10 # res_2       2.569e-01  1.176e-02   21.84 <2e-16 ***
11 #   ---
12
13 #Residual standard error: 0.07338 on 3191 degrees of freedom
14 #Multiple R-squared:  0.13, Adjusted R-squared:  0.1298
15 #F-statistic:  477 on 1 and 3191 DF, p-value: < 2.2e-16

```

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



- Write the prediction equation.

$$Y_i = -1.942e^{-18} + 0.2569X_i + \epsilon_i \quad (4)$$

The constant value $-1.942e^{-18}$ is the predicted 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 0.2569.

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?

- Run a regression where the outcome variable is the incumbent's `voteshare` and the explanatory variables are `difflog` and `presvote`.

```
1 multi_regress <- lm(voteshare ~ difflog + presvote, data = inc.sub)
2 summary(multi_regress)
3 #Residuals:
4 #   Min       1Q   Median       3Q      Max
5 #-0.25928 -0.04737 -0.00121  0.04618  0.33126
6
7 #Coefficients:
8 #   Estimate Std. Error t value Pr(>|t|)
9 #(Intercept)  0.4486442   0.0063297   70.88  <2e-16
10 #difflog      0.0355431   0.0009455    37.59  <2e-16
11 #presvote     0.2568770   0.0117637    21.84  <2e-16
12
13 #Residual standard error: 0.07339 on 3190 degrees of freedom
14 #Multiple R-squared:  0.4496, Adjusted R-squared:  0.4493
15 #F-statistic: 1303 on 2 and 3190 DF, p-value: < 2.2e-16
```

- Write the prediction equation.

$$Y_i = 0.4486442 + 0.0355431X_{1i} + 0.2568770X_{2i} + \epsilon_i \quad (5)$$

The constant 0.4486442 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 0.0355431 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 0.2568770 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 candidate. In this model, we are using the explanatory values of difference of spending and vote share of the presidential candidate to explain the outcome variable of incumbent's vote share. Therefore, the value for vote share of the presidential candidate needs to be a unique contribution and separate from any variation it shares with difference in spending.