Problem Set 3

Applied Stats/Quant Methods 1

Due: November 11, 2024

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 11, 2024. 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`.

```r
1 reg1 <- lm(voteshare ~ difflog, data = inc.sub)
2 summary(reg1)
```

Table 1: Coefficient Table of Model 1

	Dependent variable:
	Incumbent VS
Δ in Campaign Spending	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

The coefficent for Δ in campaign spending is not directy interpretable because the variable is logged. However, a positive relationship between the spending and the incumbent's vote share is depicted in Table 1. Simply, the larger the difference between the incumbent's and the challenger's spending, the higher the electoral outcome of the incumbent is, on average.

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

```r
1 ggplot(inc.sub, aes(x = difflog, y = voteshare)) +
2   geom_point(alpha = 0.5, size = 0.8, color = viridis(1)) +
3   geom_smooth(method='lm', color = viridis(2)[2]) +
4   labs(title = "Relationship between Vote Share and Difference in \n
     Campaign Spending (logged)",
5        x = "Difference in Campaign Spending (logged)",
6        y = "Vote Share of Incumbent") +
7   theme_minimal()
```

Figure 1: Scatterplot of the difference in campaign spending between the incumbent and the challenger and the vote share of the incumbent.

Figure 1 indicates a positive association between the two variables implying that the higher the difference is (the more the incumbent spends on campaigning), the higher her vote share is.

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

```
1 reg1_resid <- reg1$residuals
```

4. Write the prediction equation.

$$\text{Vote Share} = 0.579 + 0.042 \cdot \text{difflog} + \epsilon$$

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`.

   We see a similar relationship as in Table 1. There is a positive and statistically significant (on the 0.01 level) association between the difference in campaign spending with the vote share of the presidential candidate, on average. However, the coefficient is not directly interpretable because it is logged.

```
1 reg2 <- lm( presvote ~ difflog , inc.sub)
2 summary(reg2)
```

Table 2: Coefficient Table of Model 2

	Dependent variable:
	Presidential Cand. VS
Δ in Campaign Spending	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.

```
1 ggplot(inc.sub, aes(x = difflog , y = presvote)) +
2   geom_point(alpha = 0.5 , size = 0.8 , color = viridis(1)) +
3   geom_smooth(method='lm', color = viridis(2)[2]) +
4   labs(title = "Relationship between Vote Share and Difference in \n
     Campaign Spending (logged)",
5         x = "Difference in Campaign Spending (logged)",
6         y = "Vote Share of Presidential Candidate") +
7   theme_minimal()
```

Relationship between Vote Share and Difference in
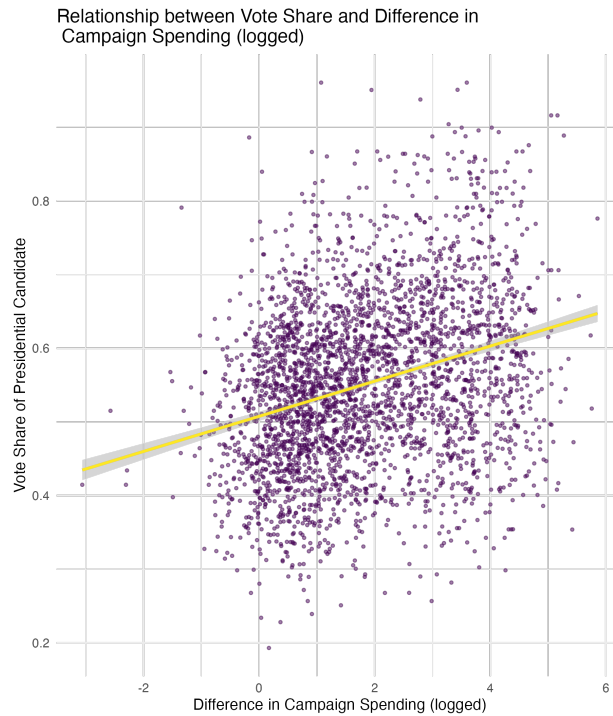Campaign Spending (logged)

Figure 2: Scatterplot of the difference in campaign spending between the incumbent and the challenger and the vote share of the presidential candidate.

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

```
1 reg2_resid <- reg2$residuals
```

4. Write the prediction equation.

$$\text{PresVote} = 0.508 + 0.024 \cdot \text{difflog} + \epsilon$$

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`.

```
reg3 <- lm(voteshare ~ presvote, inc.sub)
summary(reg3)
```

Table 3: Coefficient Table of Model 2

	Dependent variable:
	Incumbent VS
Presidential Cand. VS	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.

```
ggplot(inc.sub, aes(x = presvote, y = voteshare)) +
  geom_point(alpha = 0.5, size = 0.8, color = viridis(1)) +
  geom_smooth(method='lm', color = viridis(2)[2]) +
  labs(title = "Relationship between the Incumbent's and the Presidential
    \n Candidate's Vote Shares",
      x = "Vote Share of Presidential Candidate",
      y = "Vote Share of Incumbent") +
  theme_minimal()
```
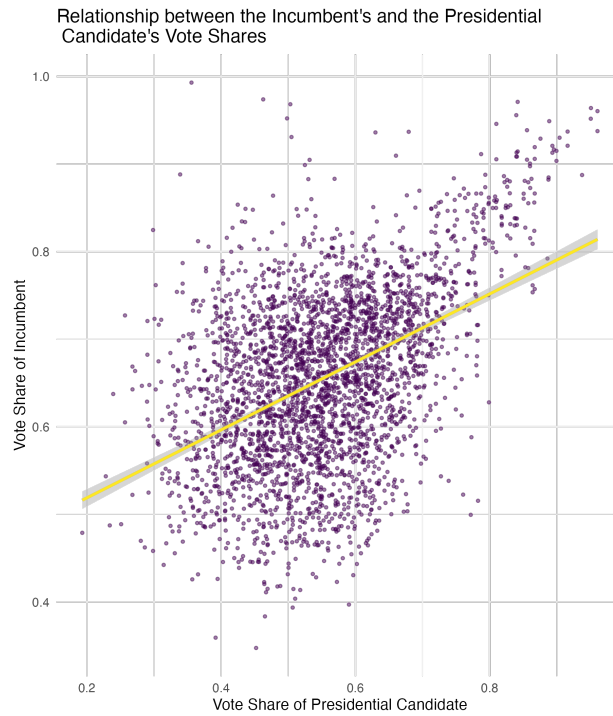
Figure 3: Scatterplot of the vote shares of the presidential candidate of the incumbent party and the vote share of the incumbent.

3. Write the prediction equation.

$$\text{Vote Share} = 0.441 + 0.388 \cdot \text{PresVote} + \epsilon$$

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_reg <- lm(reg1_resid ~ reg2_resid, data = resid_df)
2 summary(resid_reg)
```

Table 4: Coefficient Table of Residuals Regression

	Dependent variable:
	Residuals from Reg1
Residuals from Reg2	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

The regression of the residuals from the incumbent vote share model (Model 1) on the residuals from the presidential candidate vote share model (Model 2) results in a significant positive relationship (on the 0.01 α level) indicating that the unexplained variation in the incumbent's vote share is positively correlated with the unexplained variation in the presidential candidate's vote share.

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

```
1 ggplot(resid_df, aes(x = reg2_resid, y = reg1_resid)) +
2   geom_point(alpha = 0.5, size = 0.8, color = viridis(1)) +
3   geom_smooth(method='lm', color = viridis(2)[2]) +
4   labs(title = "Residual Association between Incumbent's and Presidential
       Vote Shares",
5        x = "Residuals of Presidential Vote Share (from Model 1)",
```

```
6          y = "Residuals  of  Incumbent  Vote  Share  (from  Model  2)")  +
7    theme_minimal()
```
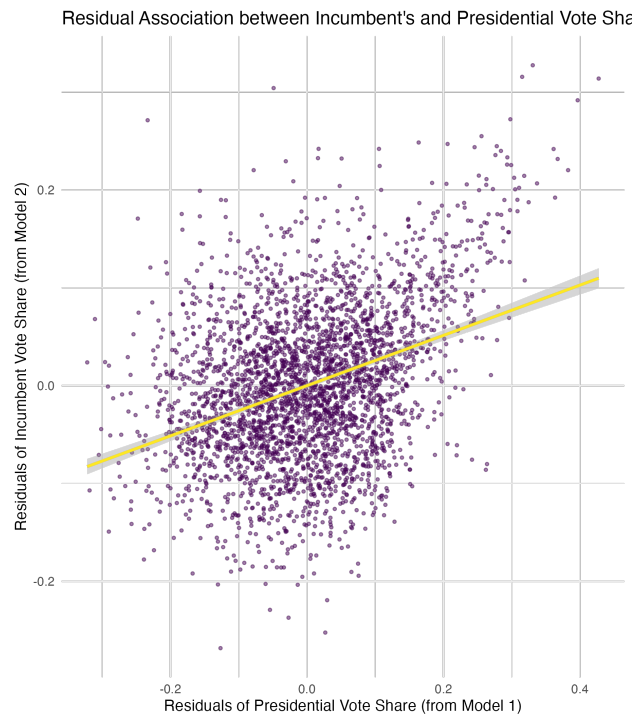

Figure 4: The relationship between the residuals of Model 1 and Model 2.

3. Write the prediction equation.

$$\text{Residuals of Vote Share (reg1)} = 0.257 \cdot \text{Residuals of Presidential VS (reg2)} + \epsilon$$

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 reg4 <- lm(voteshare ~ difflog + presvote, inc.sub)
2 summary(reg4)
```

Table 5: Coefficient Table of Model 3

	Dependent variable:
	Incumbent VS
Δ in Campaign Spending	0.036***
	(0.001)
Presidential Cand. VS	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

The interpretations of the previous models still apply to this multivariate model. The difference in campaign spending and the vote share of the presidential candidate are still positively correlated with the vote share of the incumbent. The coefficients are both statistically significant on the 0.01α level. The difference in spending coefficient cannot directly be interpreted since it is on a logged scale. However, we can interpret the presidential candidate vote share as follows: for every unit increase on the vote share of the presidential candidate of the incumbent's party, the vote share of the incumbent increases by 0.257 percentage points, on average, keeping everything else constant.

2. Write the prediction equation.

$$\text{Vote Share} = 0.449 + 0.036 \cdot \text{difflog} + 0.257 \cdot \text{PresVote} + \epsilon$$

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 regression between the residuals (Table 4) shows a positive and significant coefficient (0.257), meaning that unexplained variation in `voteshare` is positively correlated with unexplained variation in `prescand`. The fact that the coefficient for `prescand` remains the same in Table 5 suggests that the unexplained variation in `voteshare` and `prescand` are closely related, and that controlling for campaign spending differences did not change the relationship between these two variables.

In short, the similarity in the coefficients suggests that the unexplained variation in voteshare and prescand is highly correlated, and the influence of prescand on voteshare remains stable even after accounting for campaign spending differences.