Applied Stats/Quant Methods 1: Problem Set 3

Jamie McClenaghan

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`.

```
model1 <- lm(voteshare ~ difflog, data = inc.sub) #Running regression - voteshare as outcome variable, difflog as exploratory variable

summary(model1) #Summary of results for regression
```

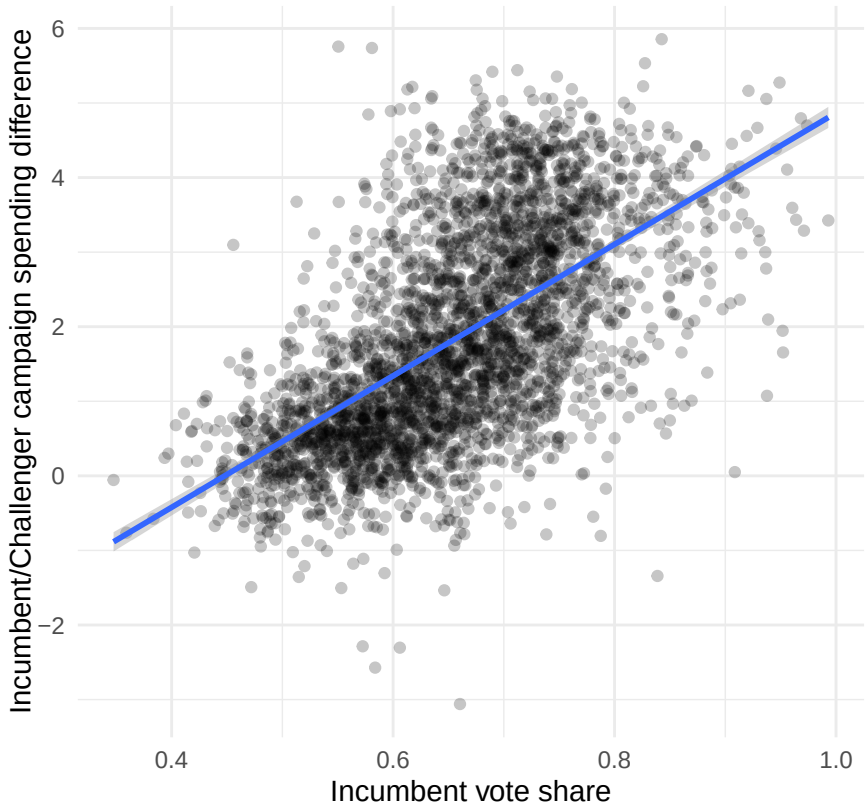
	Model 1
(Intercept)	0.58***
	(0.00)
difflog	0.04***
	(0.00)
R^2	0.37
Adj. R^2	0.37
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Statistical models

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

```
ggplot(inc.sub, aes(x = voteshare, y = difflog)) +
  geom_point(alpha = 0.225) + #Adjusting density of observations to a lower value to avoid overplotting
  geom_smooth(method = "lm", se = TRUE) + #Adding linear regression and including bands for standard errors
  labs(x = "Incumbent vote share", y = "Incumbent/Challenger campaign spending difference") + #Added labels to clearly highlight X and Y axis variables
  theme_minimal() #Added minimal theme for aesthetic purposes
```



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

```
residuals_model1 <- model1$residuals #Added regression model residuals to
    new object
residuals_model1
```

4. Write the prediction equation.

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

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`.

```
model2 <- lm(presvote ~ difflog, data = inc.sub) #Running regression - presvote as outcome variable, difflog as exploratory variable

summary(model2) #Summary of results for regression
```

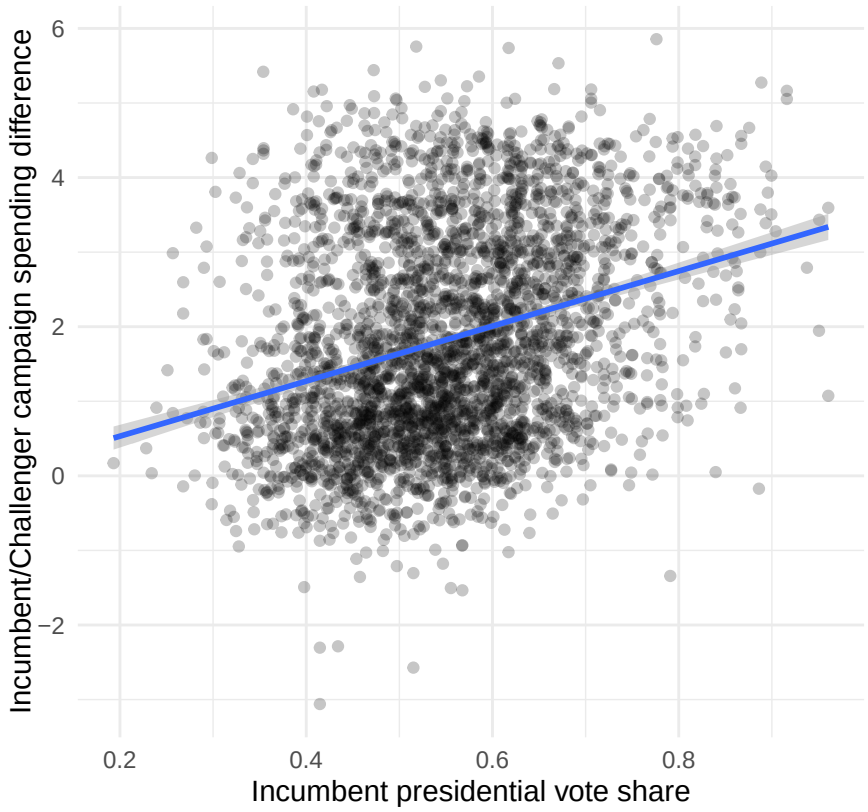
	Model 1
(Intercept)	0.51***
	(0.00)
difflog	0.02***
	(0.00)
R^2	0.09
Adj. R^2	0.09
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Statistical models

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

```
ggplot(inc.sub, aes(x = presvote, y = difflog)) +
  geom_point(alpha = 0.225) + #Adjusting density of observations to a lower value to avoid overplotting
  geom_smooth(method = "lm", se = TRUE) + #Adding linear regression and including bands for standard errors
  labs(x = "Incumbent presidential vote share", y = "Incumbent/Challenger campaign spending difference") + #Added labels to clearly highlight X and Y axis variables
  theme_minimal() #Added minimal theme for aesthetic purposes
```



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

```
residuals_model2 <- model2$residuals #Added regression model residuals to
    new object
residuals_model2
```

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`.

```
model3 <- lm(voteshare ~ presvote, data = inc.sub) #Running regression - presvote as outcome variable, difflog as exploratory variable

summary(model3) #Summary of results for regression
```

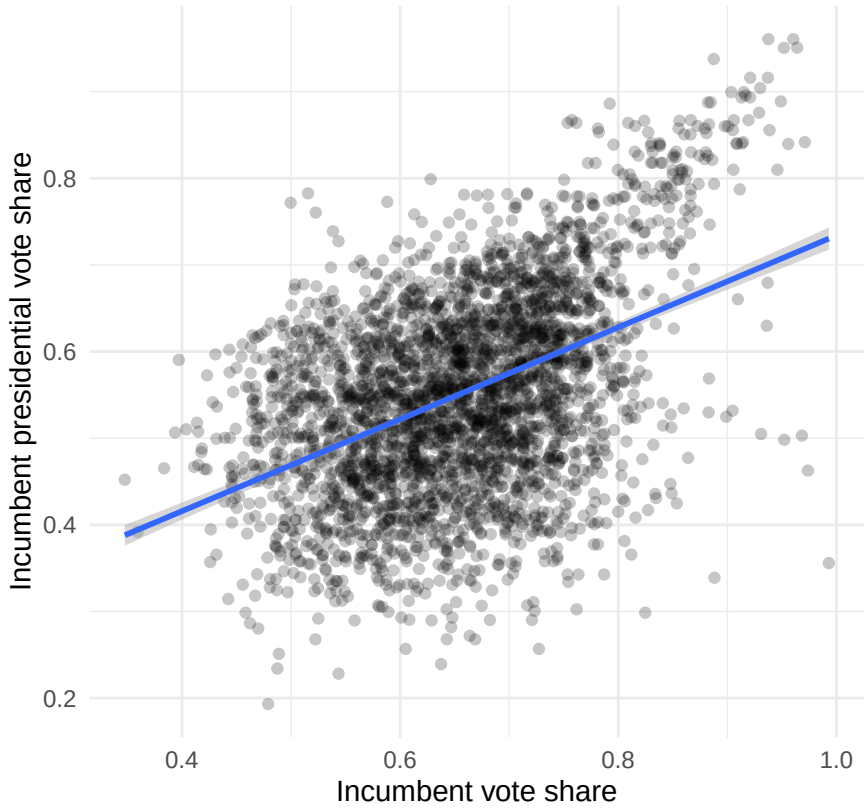
	Model 1
(Intercept)	0.44***
	(0.01)
presvote	0.39***
	(0.01)
R^2	0.21
Adj. R^2	0.21
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Statistical models

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

```
ggplot(inc.sub, aes(x = voteshare, y = presvote)) +
  geom_point(alpha = 0.225) + #Adjusting density of observations to a lower value to avoid overplotting
  geom_smooth(method = "lm", se = TRUE) + #Adding linear regression and including bands for standard errors
  labs(x = "Incumbent vote share", y = "Incumbent presidential vote share") + #Added labels to clearly highlight X and Y axis variables
  theme_minimal() #Added minimal theme for aesthetic purposes
```



3. Write the prediction equation.

$$\text{voteshare} = 0.441 + 0.388 \cdot \text{presvote} + \epsilon$$

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.

```
df <- data.frame(Model1Residuals = residuals_model1, Model2Residuals =
    residuals_model2) #Created data frame to more easily run regressions

model4 <- lm(Model1Residuals ~ Model2Residuals, data = df) #Running
    regression - presvote as outcome variable, difflog as exploratory
    variable

summary(model4) #Summary of results for regression
```

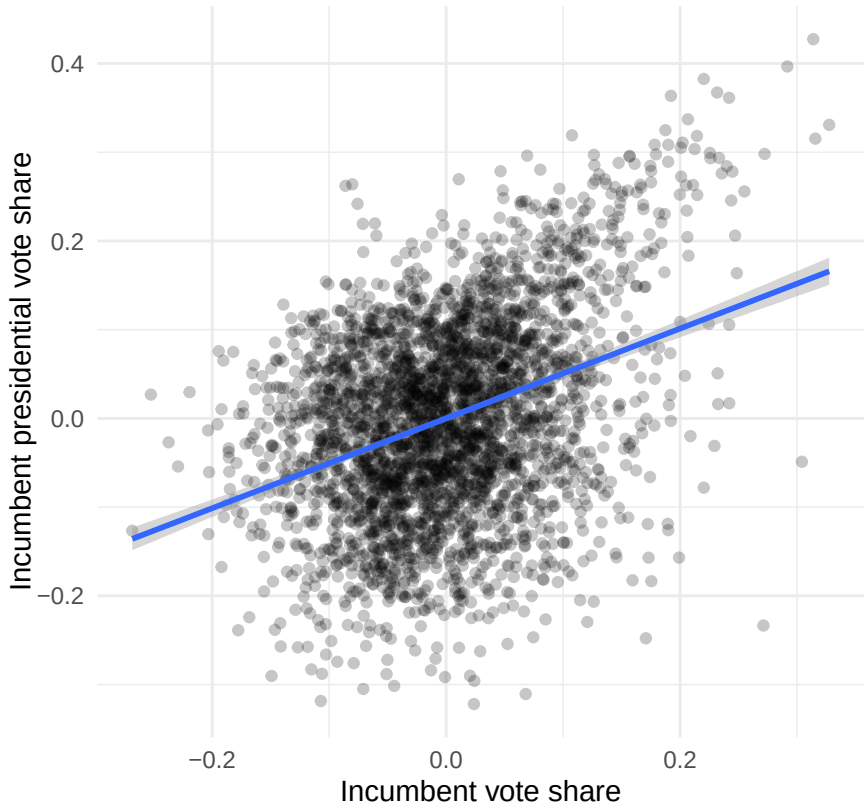
	Model 1
(Intercept)	−0.00
	(0.00)
Model2Residuals	0.26***
	(0.01)
R^2	0.13
Adj. R^2	0.13
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Statistical models

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

```
ggplot(df, aes(x = Model1Residuals, y = Model2Residuals)) +
  geom_point(alpha = 0.225) + #Adjusting density of observations to a
    lower value to avoid overplotting
  geom_smooth(method = "lm", se = TRUE) + #Adding linear regression and
    including bands for standard errors
  labs(x = "Incumbent vote share", y = "Incumbent presidential vote share
    ") + #Added labels to clearly highlight X and Y axis variables
  theme_minimal() #Added minimal theme for aesthetic purposes
```



3. Write the prediction equation.

$$\text{Model1Residuals} = 0.257 \cdot \text{Model2Residuals} + \epsilon$$

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`.

```
model5 <- lm(voteshare ~ difflog + presvote, data = inc.sub) #Running
    regression - presvote as outcome variable, difflog as exploratory
    variable
summary(model5)
```

	Model 1
(Intercept)	0.45***
	(0.01)
difflog	0.04***
	(0.00)
presvote	0.26***
	(0.01)
R^2	0.45
Adj. R^2	0.45
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: Statistical models

2. Write the prediction equation.

$$\text{voteshare} = 0.449 + 0.036 \cdot \text{difflog} + 0.257 \cdot \text{presvote} + \epsilon$$

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 from Question 1 and Question 2 in Table 4 shows a positive and significant coefficient (0.257) for Model2Residuals, this would suggest to us that the unexplained variation in `voteshare` is positively correlated with unexplained variation in `presvote`.

   The coefficient for `presvote` is the same in in Table 5 and this would suggest that the unexplained variation in `voteshare` and `presvote` are closely related, and that the `difflog` control for campaign spending differences did not affect this relationship.

   Overall, the coefficients identical nature suggests that the unexplained variation in voteshare and presvote is highly correlated, and the influence of presvote on voteshare remains the same even after we account for the campaign spending differences.