Problem Set 3

Valeriia Babaian

November 19, 2023

Question 1

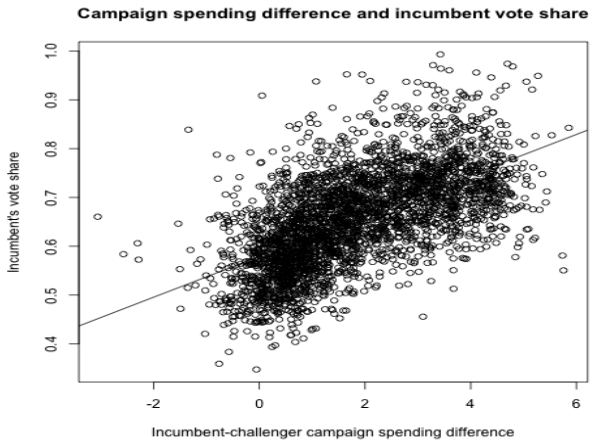
We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
q1 <- lm(voteshare ~ difflog, data = inc.sub)
```

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

```
png("Babaian-plot1.png")
plot(voteshare ~ difflog, data = inc.sub,
     xlab="Incumbent-challenger campaign spending difference",
     ylab="Incumbent's vote share",
     main="Campaign spending difference and incumbent vote share") #
   plotting relationship
abline(q1) # adding the regression line
dev.off()
```



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

```
q1_res <- q1$residuals ## saving residuals
```

4. Write the prediction equation.

```
summary(q1)
stargazer(q1, out = "q1.tex")
```

Table 1:

	Dependent variable:
	voteshare
difflog	0.042***
	(0.001)
Constant	0.580***
	(0.002)
Observations	3,193
R^2	0.370
Adjusted R^2	0.370
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,853.000*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

Prediction equation:
$\widehat{voteshare} = 0.58 + 0.042 * difflog$

Question 2

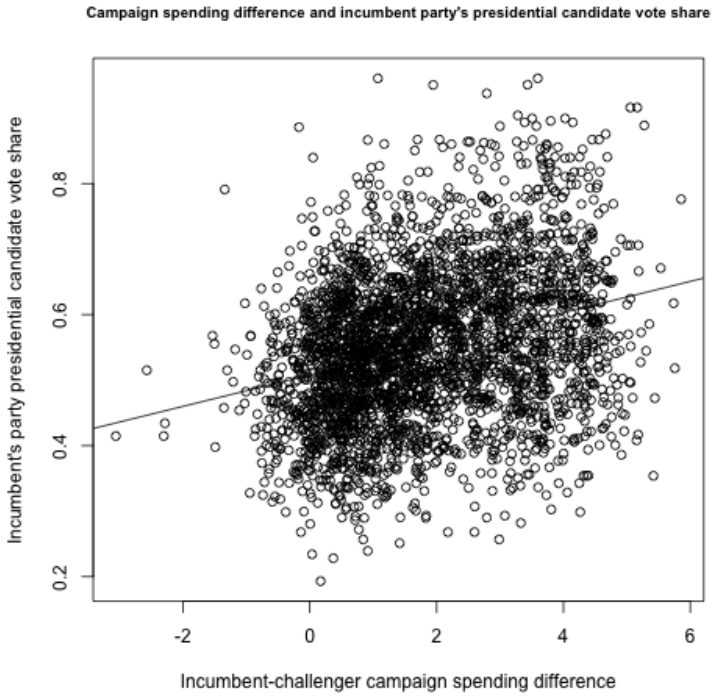
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
q2 <- lm(presvote ~ difflog, data = inc.sub)
```

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

```
png("Babaian-plot2.png")
plot(presvote ~ difflog, data = inc.sub,
     xlab="Incumbent-challenger campaign spending difference",
     ylab="Incumbent's party presidential candidate vote share",
     main="Campaign spending difference and incumbent party's
    presidential candidate vote share",
     cex.main = 0.8)# plotting relationship
abline(q2) # adding the regression line
dev.off()
```



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

```
q2_res <- q2$residuals ## saving residuals
```

4. Write the prediction equation.

```
summary(q2)
stargazer(q2, out = "q2.tex")
```

Table 2:

	Dependent variable:
	presvote
difflog	0.024***
	(0.001)
Constant	0.510***
	(0.003)
Observations	3,193
R^2	0.088
Adjusted R^2	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	308.000*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

Prediction equation:
$\widehat{presvote} = 0.51 + 0.024 * difflog$

Question 3

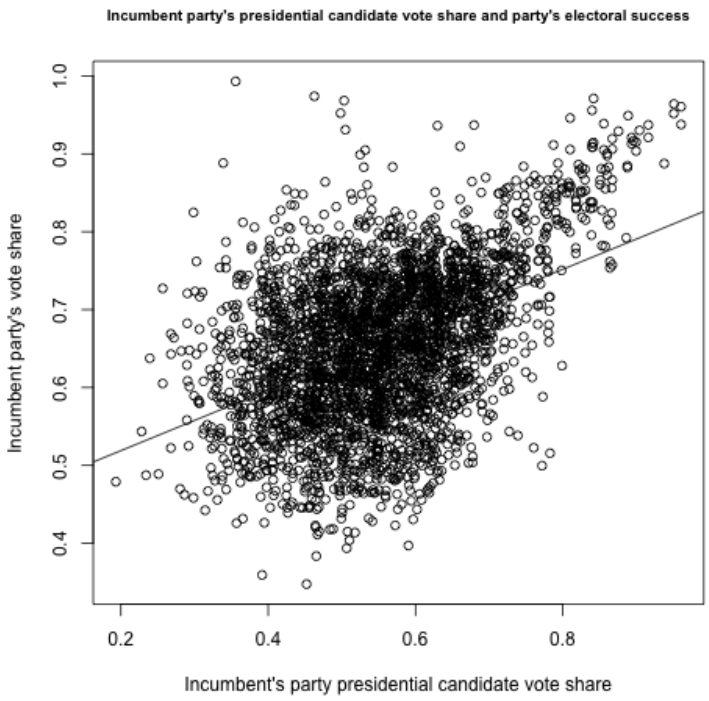
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
q3 <- lm(voteshare ~ presvote, data = inc.sub)
```

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

```
png("Babaian-plot3.png")
plot(voteshare ~ presvote, data = inc.sub,
     xlab="Incumbent's party presidential candidate vote share",
     ylab="Incumbent party's vote share",
     main="Incumbent party's presidential candidate vote share and party's electoral success",
     cex.main = 0.8) # plotting relationship
abline(q3)
dev.off()
```



3. Write the prediction equation.

```r
summary(q3)
stargazer(q3, out = "q3.tex")
```

Table 3:

	Dependent variable:
	voteshare
presvote	0.390***
	(0.013)
Constant	0.440***
	(0.008)
Observations	3,193
R^2	0.210
Adjusted R^2	0.210
Residual Std. Error	0.088 (df = 3191)
F Statistic	827.000*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

Prediction equation:
$\widehat{voteshare} = 0.44 + 0.39 * presvote$

Question 4

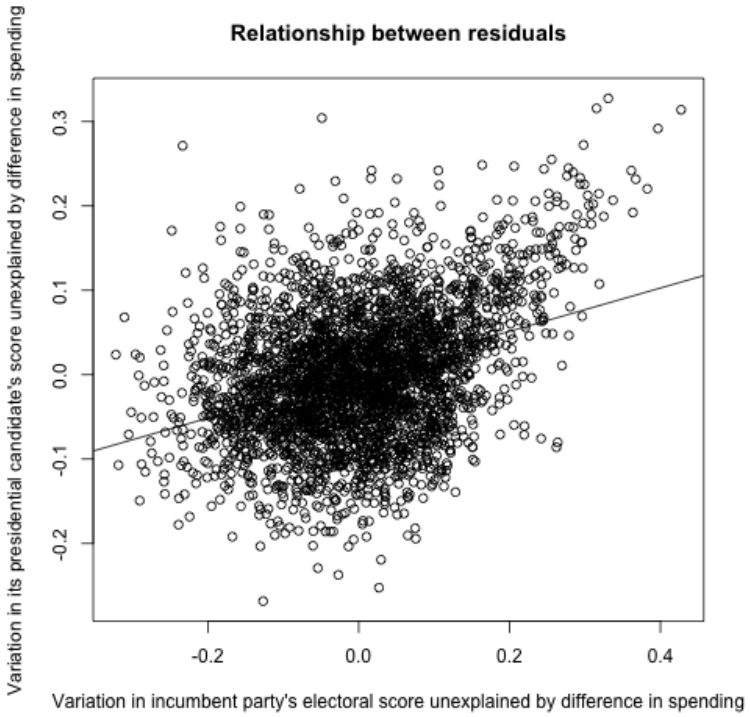
The residuals from part (a) tell us how much of the variation in `voteshare` is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in `presvote` is *not* explained by the difference in spending between incumbent 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.

```
q4 <- lm(q1_res~q2_res)
```

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

```
png("Babaian-plot4.png")
plot(q1_res~q2_res,
     xlab="Variation in incumbent party's electoral score unexplained by
    difference in spending",
     ylab="Variation in its presidential candidate's score unexplained by
     difference in spending",
     main="Relationship between residuals") # plotting relationship
abline(q4)
dev.off()
```



3. Write the prediction equation.

```r
summary(q4)
stargazer(q4, out = "q4.tex")
```

Table 4:

	Dependent variable:
	q1_res
q2_res	0.260^{***}
	(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	477.000^{***} (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

Prediction equation:
$\widehat{voteshare\ residuals} = 0.26 * presvote\ residuals - 4.9 * 10^{-18}$
(A more precise value of constant is derived from R console output.)

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`.

```
q5 <- lm(voteshare ~ difflog + presvote, data = inc.sub)
```

2. Write the prediction equation.

```
summary(q5)
stargazer(q5, out = "q5.tex")
```

Table 5:

	Dependent variable:
	voteshare
difflog	0.036***
	(0.001)
presvote	0.260***
	(0.012)
Constant	0.450***
	(0.006)
Observations	3,193
R^2	0.450
Adjusted R^2	0.450
Residual Std. Error	0.073 (df = 3190)
F Statistic	1,303.000*** (df = 2; 3190)
Note:	*p<0.1; **p<0.05; ***p<0.01

Prediction equation:
$\widehat{voteshare} = 0.45 + 0.26 * presvote + 0.036 * difflog$

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 coefficient estimate for `presvote` is identical to the estimate of its residuals unexplained by `difflog` for predicting `voteshare` unexplained by `difflog`. In this

multivariate model (and in any other OLS model), it occurs because total variance is the variance explained by the regression line added residuals. As the residulas of `voteshare` and `presvote` were those left after accounting for `difflog`, `difflog` was already included in the calculation, therefore, the same effect remains true for the effects of observed `voteshare` and `presvote` that can be obtained from the sums of their values predicted by `difflog` and their residuals.