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.

```
inc.sub <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/incumbents_subset.csv")
summary(inc.sub)
```

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

```
model1 <- lm(voteshare~difflog, data = inc.sub)
stargazer(model1, type = "latex")
```

Table 1:

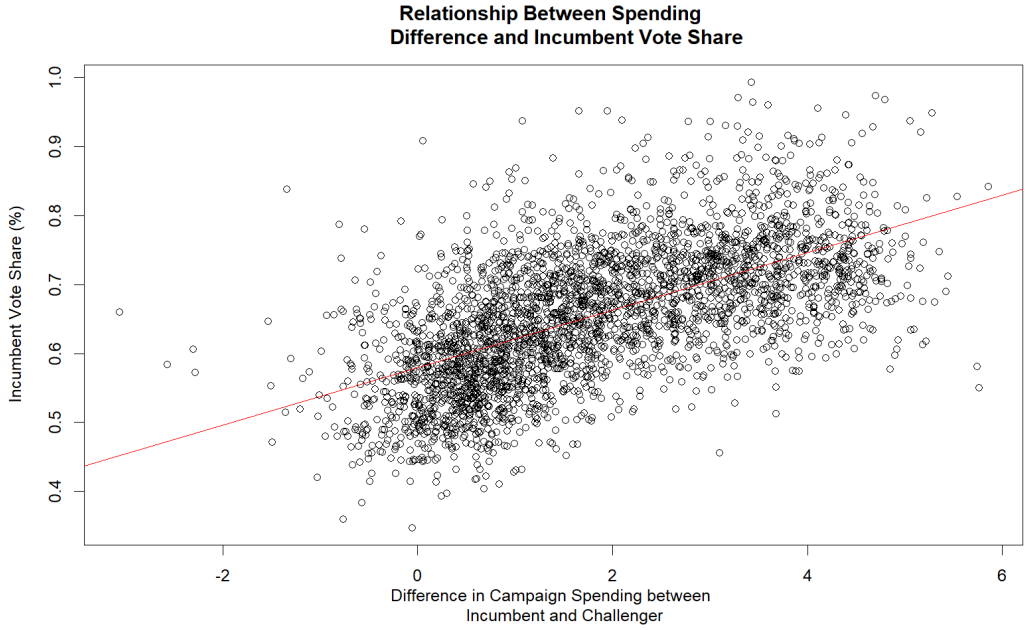
	Dependent variable:
	voteshare
difflog	0.042***
	(0.001)
Constant	0.579***
	(0.002)
Observations	3,193
R^2	0.367
Adjusted R^2	0.367
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,852.791*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

On average, a one unit increase in the logged difference in campaign spending is associated with a 4.2% increase in the incumbent vote share. The positive correlation between the two variables is statistically reliable.

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

```
png("Plot1.png",width = 1600, height = 1000, res = 150)
plot(y=inc.sub$voteshare, x=inc.sub$difflog,
     main = "Relationship Between Spending
     Difference and Incumbent Vote Share",
     ylab = "Incumbent Vote Share (%)",
     xlab = "Difference in Campaign Spending between
     Incumbent and Challenger")
abline(model1, col = "red")
dev.off()
```

Figure 1:



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

```
# Save residuals into new variable
residuals_model1 <- model1$residuals
residuals_model1
```

4. Write the prediction equation.

   voteshare = 0.579 + 0.042difflog

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)
stargazer(model2, type = "latex")
```

Table 2:

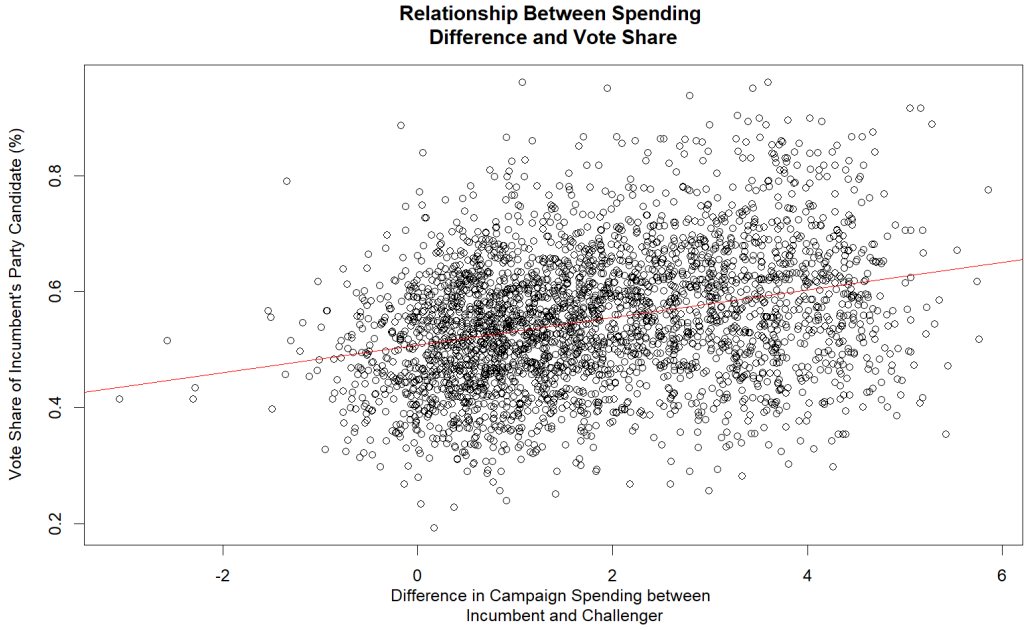
	Dependent variable:
	presvote
difflog	0.024***
	(0.001)
Constant	0.508***
	(0.003)
Observations	3,193
R^2	0.088
Adjusted R^2	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

On average, a 1 unit increase in the logged difference in campaign spending is associated with a 2.4% increase in the incumbent party candidates vote share. The positive correlation between the variables is statistically reliable.

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

```
png("Plot2.png",width = 1600, height = 1000, res = 150)
plot(y = inc.sub$presvote, x = inc.sub$difflog,
     main = "Relationship Between Spending
Difference and Vote Share",
     ylab = "Vote Share of Incumbent's Party Candidate (%)",
     xlab = "Difference in Campaign Spending between
     Incumbent and Challenger")
abline(model2, col = "red")
dev.off()
```

Figure 2:



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

```
# Save Residuals in Object Called residuals_model2
residuals_model2 <- model2$residuals
```

4. Write the prediction equation.

   presvote = 0.508 + 0.024difflog

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)
stargazer(model3, type = "latex")
```

.

Table 3:

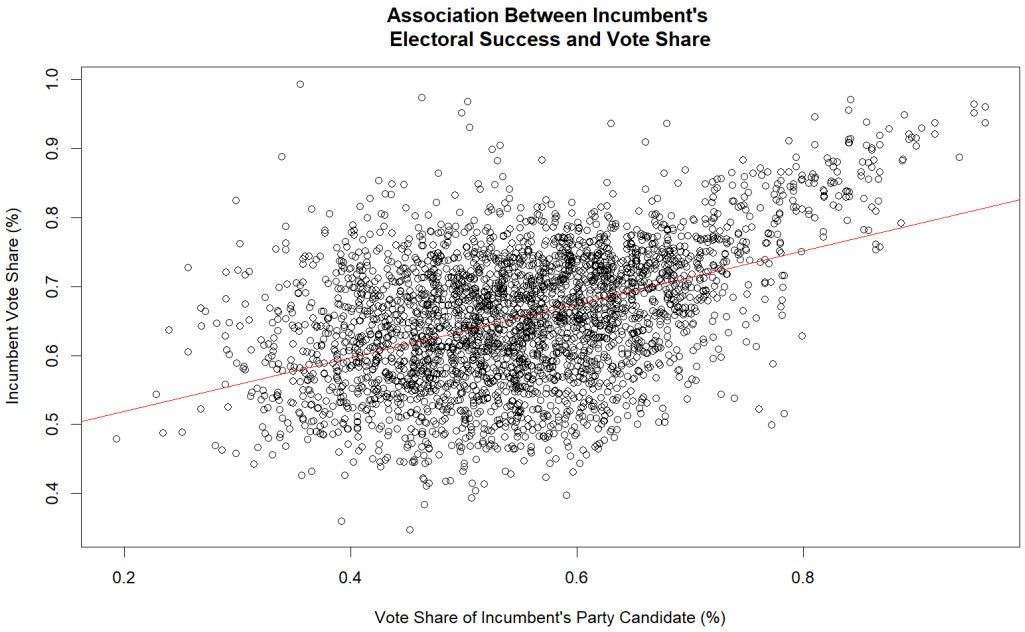
	Dependent variable:
	voteshare
presvote	0.388***
	(0.013)
Constant	0.441***
	(0.008)
Observations	3,193
R^2	0.206
Adjusted R^2	0.206
Residual Std. Error	0.088 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

On average, a one unit increase in the incumbent party candidates vote share is associated with a 38.8% increase in the incumbent vote share. The positive correlation between the two variables is statistically reliable.

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

```
png("Plot3.png",width = 1600, height = 1000, res = 150)
plot(y = inc.sub$voteshare, x = inc.sub$presvote,
     main = "Association Between Incumbent's
Electoral Success and Vote Share",
     ylab = "Incumbent Vote Share (%)",
     xlab = "Vote Share of Incumbent's Party Candidate (%)")
abline(model3, col = "red")
dev.off()
```

Figure 3:



3. Write the prediction equation.

   voteshare = 0.441 + 0.388presvote

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.

```
resid_regr <- lm(residuals_model1 ~ residuals_model2)
stargazer(resid_regr, type = "latex")
```

Table 4:

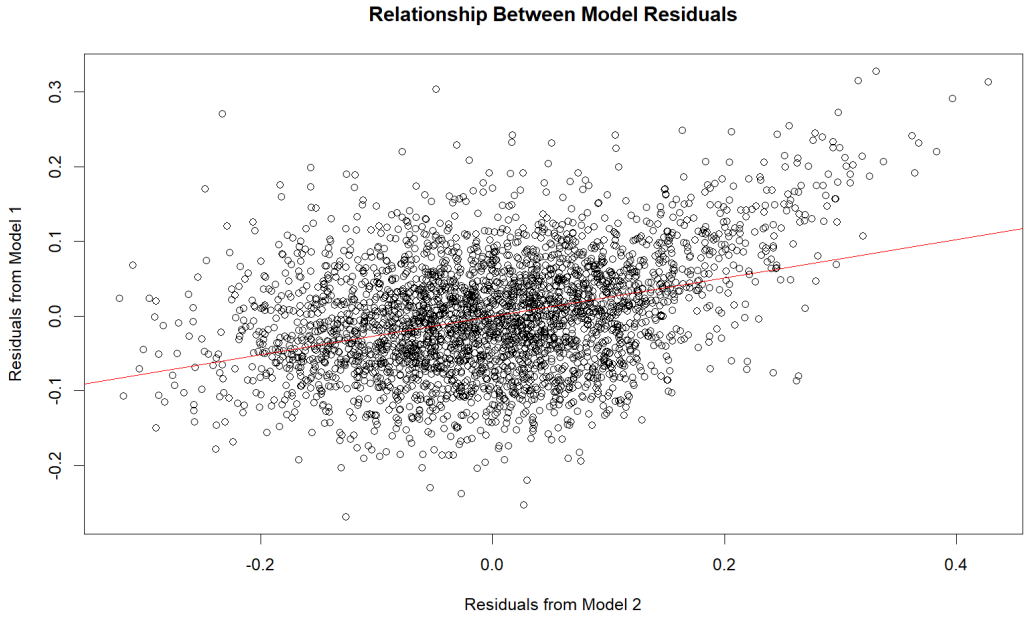
	Dependent variable:
	residuals_model1
residuals_model2	0.257^{***}
	(0.012)
Constant	-0.000
	(0.001)
Observations	3,193
R^2	0.130
Adjusted R^2	0.130
Residual Std. Error	0.073 (df = 3191)
F Statistic	476.975^{***} (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

On average, a one unit increase in the residuals of model 2 is associated with a 0.257 unit increase in the residuals of model 1. The positive correlation between the variables is statistically reliable.

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

```
png("Plot4.png",width = 1600, height = 1000, res = 150)
plot(y = residuals_model1, x = residuals_model2,
     main = "Relationship Between Model Residuals",
     ylab = "Residuals from Model 1",
     xlab = "Residuals from Model 2")
abline(resid_regr, col = "red")
dev.off()
```

Figure 4:



3. Write the prediction equation.

   residuals model 1 = -0.000 + 0.257residualsmodel2

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`.

```
mlr <- lm(voteshare~difflog + presvote, data = inc.sub)
stargazer(mlr, type = "latex")
```

Table 5:

	Dependent variable:
	voteshare
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

Holding all other variables constant, a one unit increase in the logged spending difference is associated with a 3.6% increase in the incumbent vote share on average.

Holding all other variables constant, a one unit increase in the incumbent party candidates vote share is associated with a 25.7% increase in the incumbent's vote share on average.

Both explanatory variables have statistically reliable positive correlations with the outcome variable

2. Write the prediction equation.

   voteshare = 0.449 + 0.036difflog + 0.257presvote

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 on the X variable "presvote" is the same as the coefficient of the residuals of model 2 when regressed on the residuals of model 1. The effect of presvote on voteshare is therefore contained within the residuals of model 2. By adding presvote to the regression model, we are therefore removing some portion of omitted variable bias that is present within model 2 - wherein presvote was not included as a regressor.