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`.

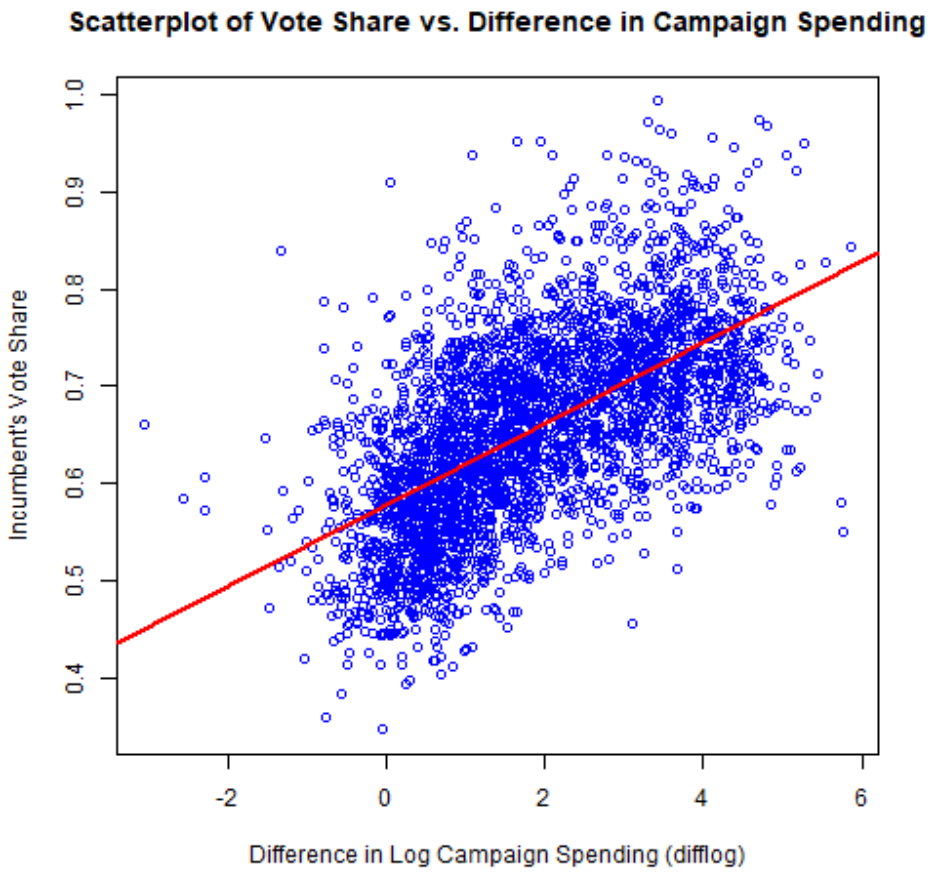
```
reg1 <- lm(voteshare ~ difflog, data = inc.sub)

# display the summary of the regression results
summary(reg1)
```

Table 1: Q1 Regression

	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

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



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

```
res1 <- reg1$residuals
head(res1,12)
```

Table 2: Q1 Residuals

-0.0004	-0.032	-0.005	0.039	0.036	0.032
-0.013	0.014	-0.118	-0.086	-0.012	-0.035

4. Write the prediction equation.

```
y_hat1 <- reg1$coefficients[1] + reg1$coefficients[2] * inc.sub$difflog
head(y_hat1,12)
```

Table 3: Q1 Prediction Equation

0.603	0.615	0.597	0.661	0.606	0.593
0.644	0.578	0.622	0.581	0.601	0.604

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`.

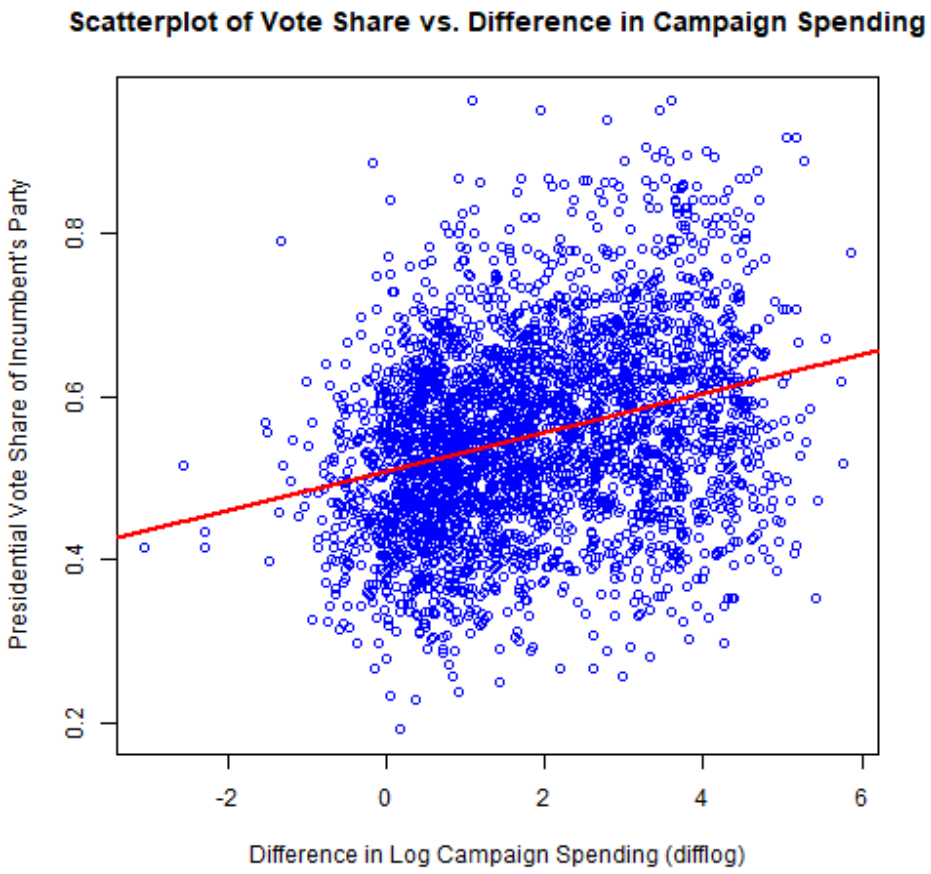
```
reg2 <- lm(presvote ~ difflog, data = inc.sub)

# display the summary of the regression results
summary(reg2)
```

Table 4: Q2 Regression

	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

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



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

```
res2 <- reg2$residuals
head(res2,12)
```

Table 5: Q2 Residuals

0.006	0.038	-0.053	-0.053	-0.046	0.074
-0.004	-0.060	-0.146	0.056	-0.127	0.081

4. Write the prediction equation.

```
# y_hat = B0 + B1 * X
y_hat2 <- reg2$coefficients[1] + reg2$coefficients[2] * inc.sub$difflog
head(y_hat2,12)
```

Table 6: Q2 Prediction Equation

0.521	0.528	0.518	0.554	0.523	0.516
0.544	0.507	0.532	0.509	0.520	0.522

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`.

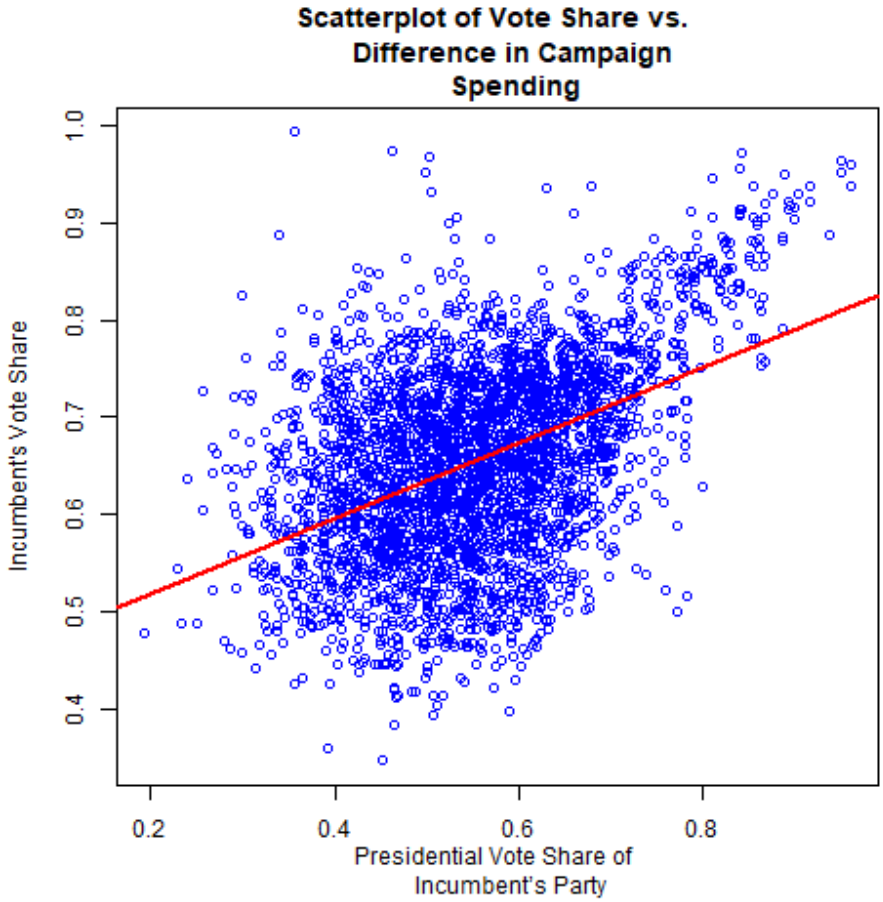
```
reg3 <- lm(voteshare ~ presvote, data = inc.sub)


# display the summary of the regression results
summary(reg3)
```

Table 7: Q3 Regression

	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

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



3. Write the prediction equation.

```
# y_hat = B0 + B1 * X
y_hat3 <- reg3$coefficients[1] + reg3$coefficients[2] * inc.sub$presvote
head(y_hat3,12)
```

Table 8: Q3 Prediction Equation

0.646	0.661	0.622	0.636	0.627	0.670
0.651	0.615	0.591	0.661	0.594	0.675

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.

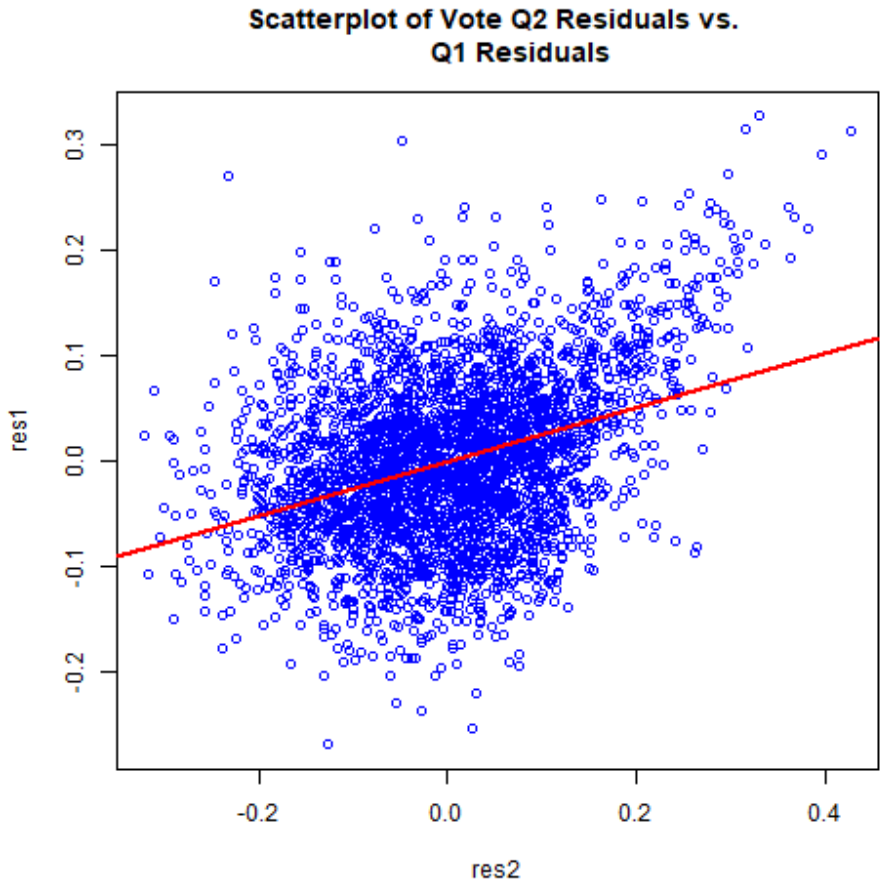
```
reg4 <- lm(res1 ~ res2)

# display the summary of the regression results
summary(reg4)
```

Table 9: Q4 Regression

	Dependent variable:
	res1
res2	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

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



3. Write the prediction equation.

```
# y_hat = B0 + B1 * X
y_hat4 <- reg4$coefficients[1] + reg4$coefficients[2] * res2
head(y_hat4,12)
```

Table 10: Q4 Prediction Equation

0.001	0.010	-0.014	-0.014	-0.012	0.019
-0.001	-0.015	-0.038	0.014	-0.033	0.021

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`.

```
reg5 <- lm(voteshare ~ difflog + presvote, data = inc.sub)

# display the summary of the regression results
summary(reg5)

# get the latex code
stargazer(reg5)
```

Table 11: Q5 Regression

	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

2. Write the prediction equation.

```r
# y_hat = B0 + B1*X + B2*X
y_hat5 <- reg5$coefficients[1] +
          reg5$coefficients[2] * inc.sub$difflog +
          reg5$coefficients[3] * inc.sub$presvote
head(y_hat5,12)
```

Table 12: Q5 Prediction Equation

0.604	0.625	0.583	0.647	0.595	0.613
0.643	0.563	0.584	0.596	0.568	0.624

3. What is it in this output that is identical to the output in Question 4?

```r
reg4$coefficients  # Coefficient for res2 in reg4
```

Table 13: Coefficients for reg4

(Intercept)	res2
-5.9341e^{-18}	0.257

```r
reg5$coefficients # Coefficient for presvote in reg5
```

Table 14: Coefficients for reg5

(Intercept)	difflog	presvote
0.449	0.036	0.257

Both **reg5$coefficients[3]** and **reg4$coefficients[2]** are coefficients that describe the relationship between the variable **presvote** and the outcome variable **voteshare**, but they do so in different contexts. Specifically, they both quantify how variations in presidential vote share (whether direct or through residuals) affect the vote share of incumbents:

4. Why do you think this is the case?

   The similarity between **reg5$coefficients[3]** and **reg4$coefficients[2]** arises from the interconnected nature of the relationships they represent. Both coefficients reflect how the presidential vote share is associated with the vote share of incumbents: in **reg5**, this relationship is direct and considers all variance, while in **reg4**, it focuses on the variance not explained by **difflog**. The coefficient from **reg4** captures the relationship of the residuals - parts of the votes not explained by **difflog**, while reg5 isolates

the influence of **presvote** on **voteshare**. Additionally, if **presvote** and **difflog** are correlated, the unexplained variations in **presvote** (as captured in **res2**) are likely to reflect similar unexplained patterns in **voteshare** (as captured in **res1**). Consequently, both coefficients capture the same underlying relationship from different perspectives, revealing similar insights about the same data relationships.