

Problem Set 3

QTM 200: Applied Regression Analysis

Due: February 17, 2020

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 the course GitHub page in .pdf form.
- This problem set is due at the beginning of class on Monday, February 17, 2020. No late assignments will be accepted.
- Total available points for this homework is 100.

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 (20 points)

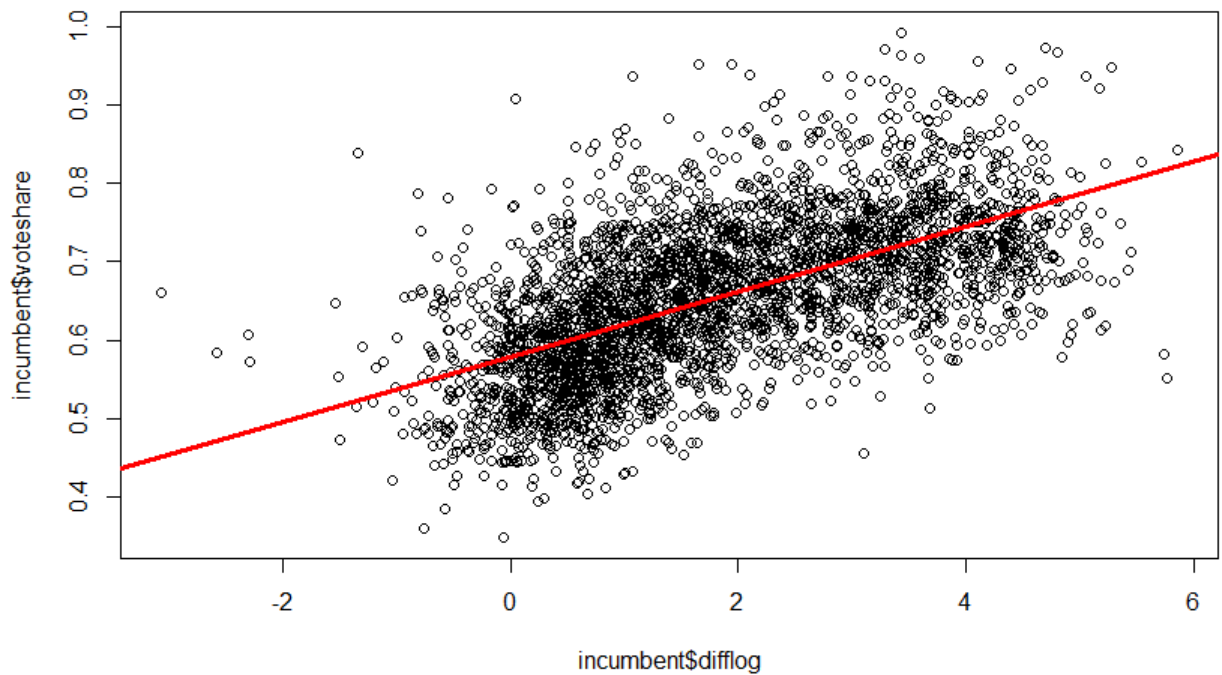
We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

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

```
1 #Run regression
2 regresults1<-lm(voteshare~difflog , data=incumbent)
3 summary(regresults1)
```

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

```
1 #Make scatterplot with regression line
2 plot(incumbent$difflog ,incumbent$voteshare)
3 abline(regresults1 ,col="red",lwd=3)
```



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

```
1 #save residuals
2 residuals1<-resid(regresults1)
3 residuals1
```

4. Write the prediction equation.

```
1 #prediction equation  
2 #y=.579 + .042x
```

Question 2 (20 points)

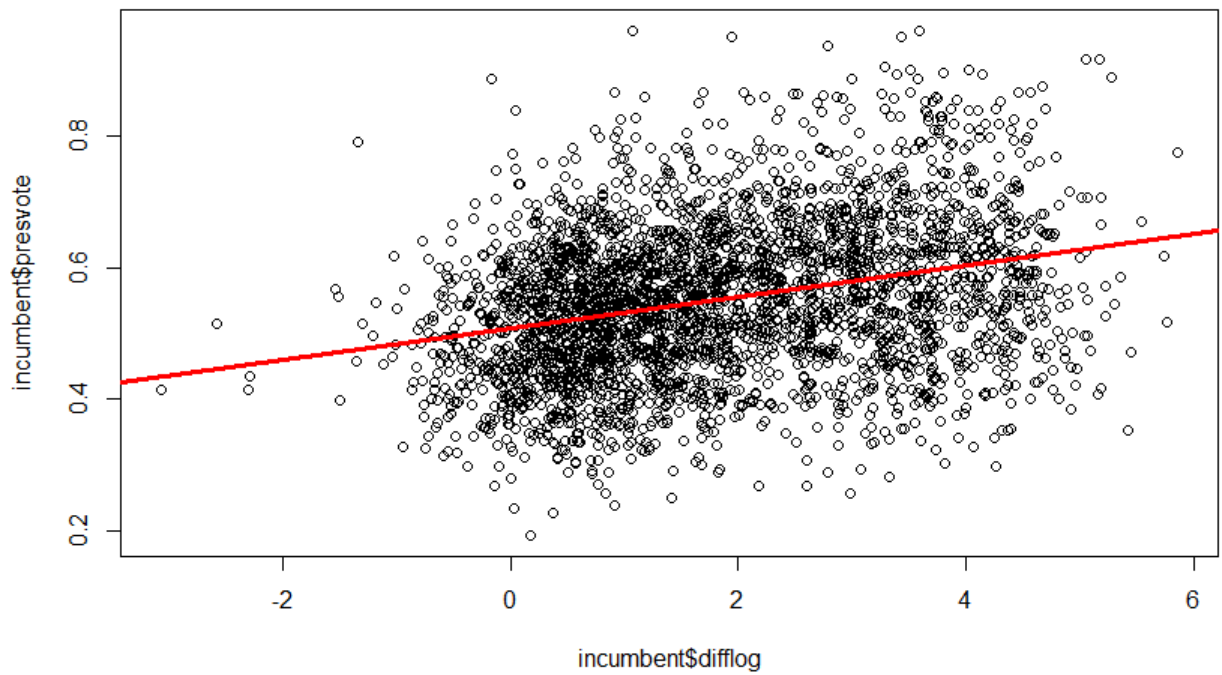
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
1 #run regression
2 regresults2<-lm(presvote~difflog , data=incumbent)
3 summary(regresults2)
```

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

```
1 #make scatterplot with regression line
2 plot(incumbent$difflog ,incumbent$presvote)
3 abline(regresults2 , col="red" ,lwd=3)
```



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

```
1 #save residuals  
2 residuals2<-resid(regresults2)  
3 residuals2
```

4. Write the prediction equation.

```
1 #prediction equation  
2 #y=.508 + .024x
```

Question 3 (20 points)

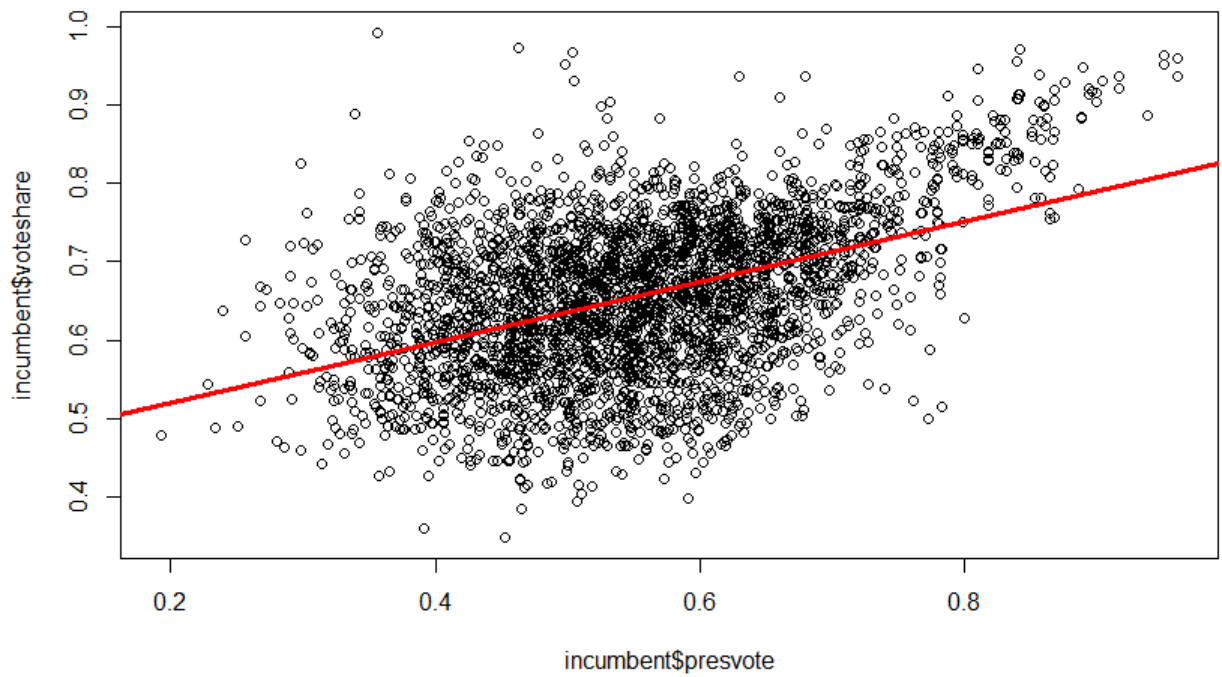
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
1 #run regression
2 regresults3<-lm(voteshare~presvote , data=incumbent)
3 summary(regresults3)
```

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

```
1 #make scatterplot with regression line
2 plot(incumbent$presvote , incumbent$voteshare)
3 abline(regresults3 , col="red" , lwd=3)
```



3. Write the prediction equation.

```
1 #prediction equation  
2 #y=.441 + .388x
```


Question 4 (20 points)

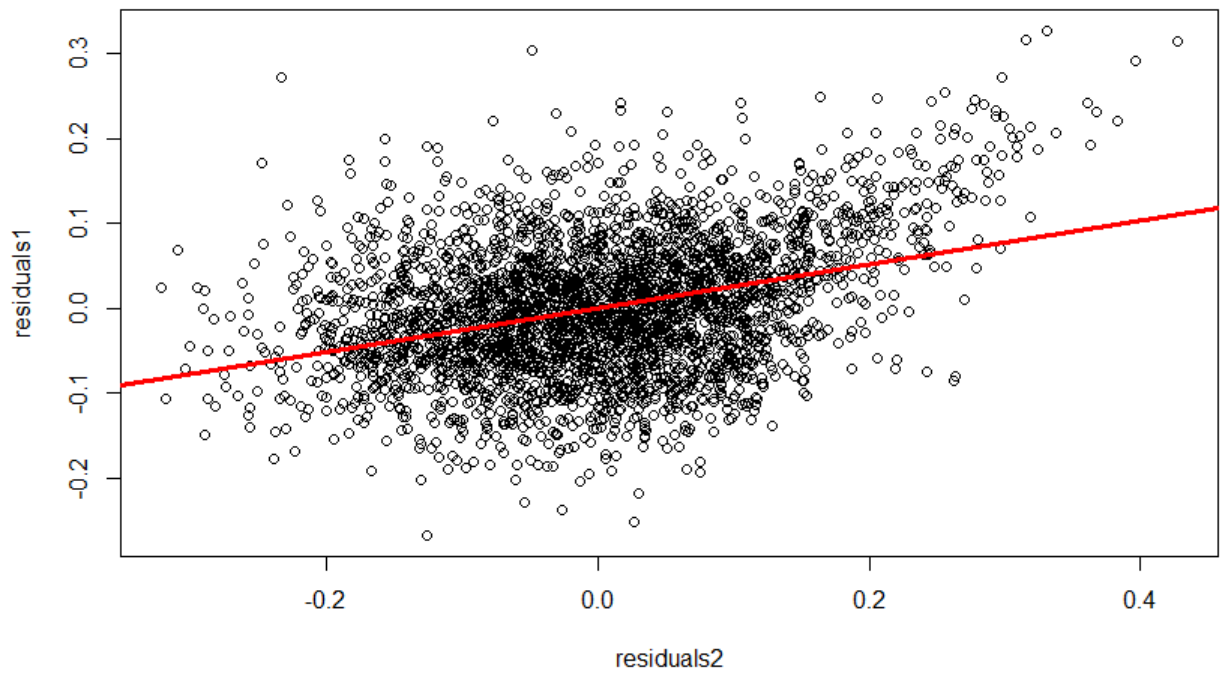
The residuals from part (a) tell us how much of the variation in **voteshare** is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in **presvote** is *not* explained by the difference in spending between incumbent 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 #run regression
2 regresults4<-lm(residuals1 ~ residuals2)
3 summary(regresults4)
```

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

```
1 #make scatterplot with regression line
2 plot(residuals2 , residuals1)
3 abline(regresults4 , col="red" , lwd=3)
```



3. Write the prediction equation.

```
1 #prediction equation  
2 #y=-4.860x10^-18 + .257x
```

Question 5 (20 points)

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 #run regression
2 regresults5<-lm(voteshare~difflog + presvote ,data=incumbent)
3 summary(regresults5)
```

2. Write the prediction equation.

```
1 #write prediction equation
2 #y=.449 + .036x1 + .257x2 (where x1 is difflog and x2 is presvote)
```

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case?

- 1 #What is it in this output that is identical to the output in Question 4?
Why do you think this is the case?
- 2 #The coefficient value for the residuals from the first model in Question 4 is the same as the coefficient value for presvote in this question. Since these residuals tell us how much variation in voteshare is not explained by difflog, it makes sense that this unexplained variation is coming from presvote, which is included in this model but not the earlier one that generated the residuals.