

# 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`.

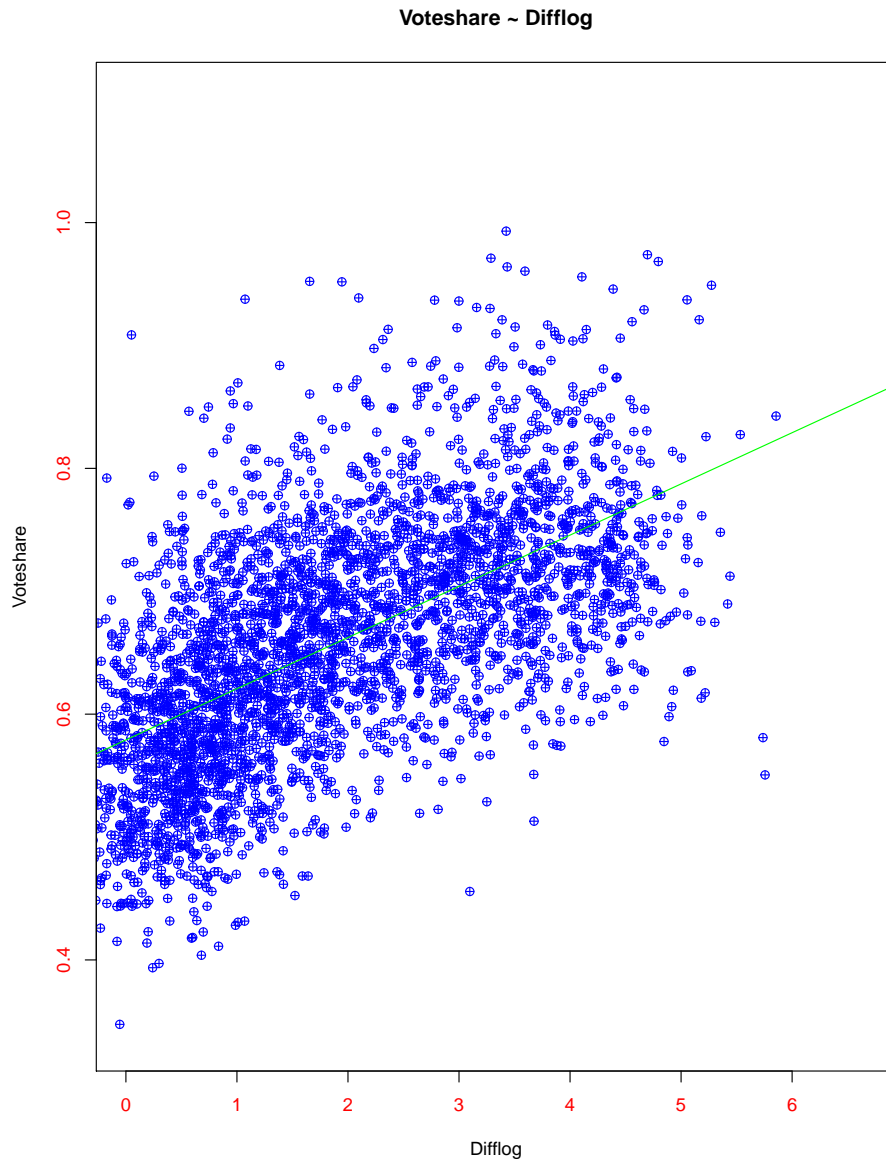
```
1 #1
2 data<-read.csv("https://raw.githubusercontent.com/zieglerjef/
   QTM200Spring2020/master/problem_sets/PS3/incumbents_subset.csv")
3
4 linearMod1 <- lm(voteshare ~ difflog, data=data)
5 linearMod1
```

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

```

1 #2)
2 plot(data$difflog , data$voteshare ,
3       xlab="Difflog" ,
4       ylab="Voteshare" ,
5       main="Voteshare ~ Difflog" , pch=10, col='blue' ,
6       col.axis='red' , xlim=c(0,6.66666) , ylim=c(0.34,1.1))
7
8 abline(linearMod1 , col="green")

```



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

```

1 #3)

```

```
2 residual1 <- resid(linearMod1)
3 plot(density(residual1))
```

4. Write the prediction equation.

$$\text{voteshare} = 0.57903 + 0.04167 * (\text{difflog})$$

## 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 #1)
2 linearMod2 <- lm(presvote ~ difflog , data=data)
3 linearMod2
```

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

```
1 #2)
2 plot(data$difflog , data$presvote ,
3       xlab="Difflog" ,
4       ylab="Presvote" ,
5       main="Presvote ~ Difflog" , pch=6 , col='blue' ,
6       col.axis='red' , xlim=c(0 , 6.666666) , ylim=c(0.34 , 1.1))
7
8 abline(linearMod2 , col="green")
```

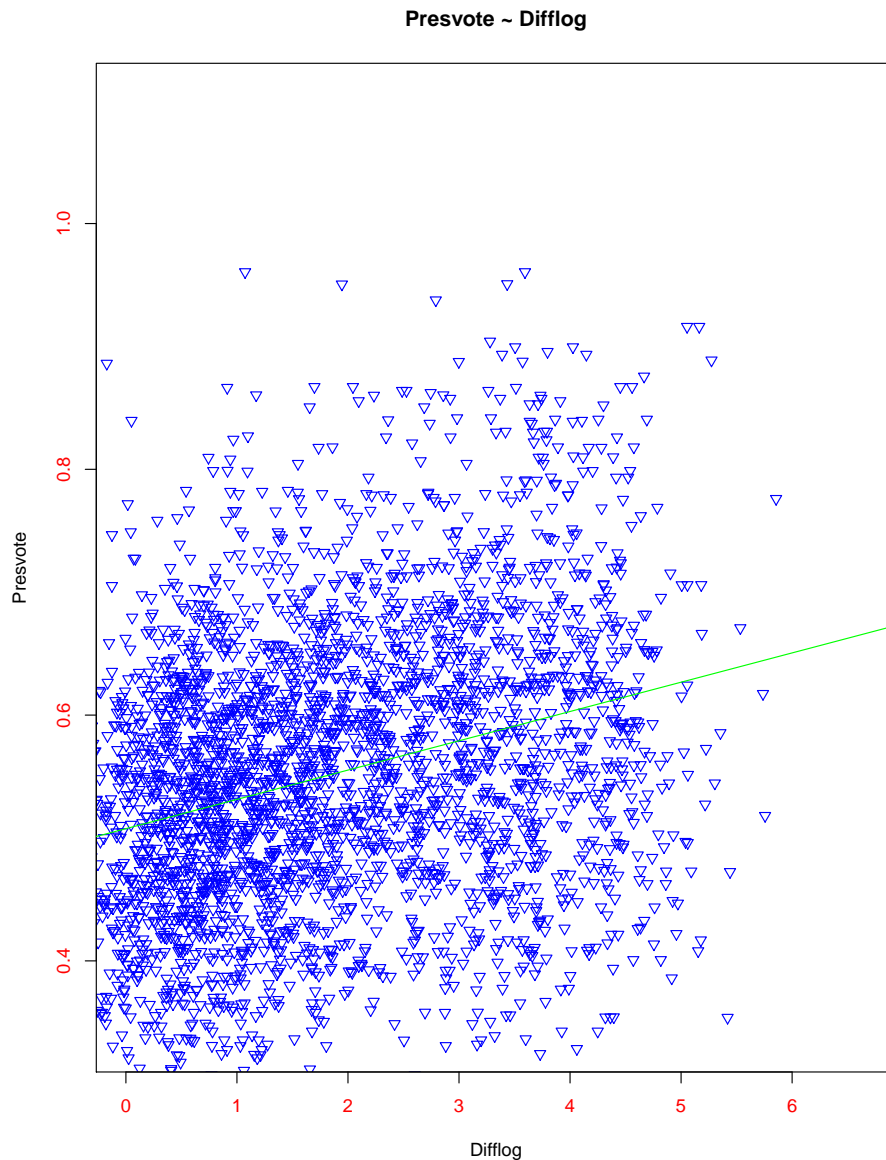
3. Save the residuals of the model in a separate object.

```
1 #3)
2 residual2 <- resid(linearMod2)
3 plot(density(residual2))
```

4. Write the prediction equation.

The graph under this question is actually the graph for the previous question

$$\text{Presvote} = 0.50758 + 0.02384 * (\text{difflog})$$



### Question 3 (20 points)

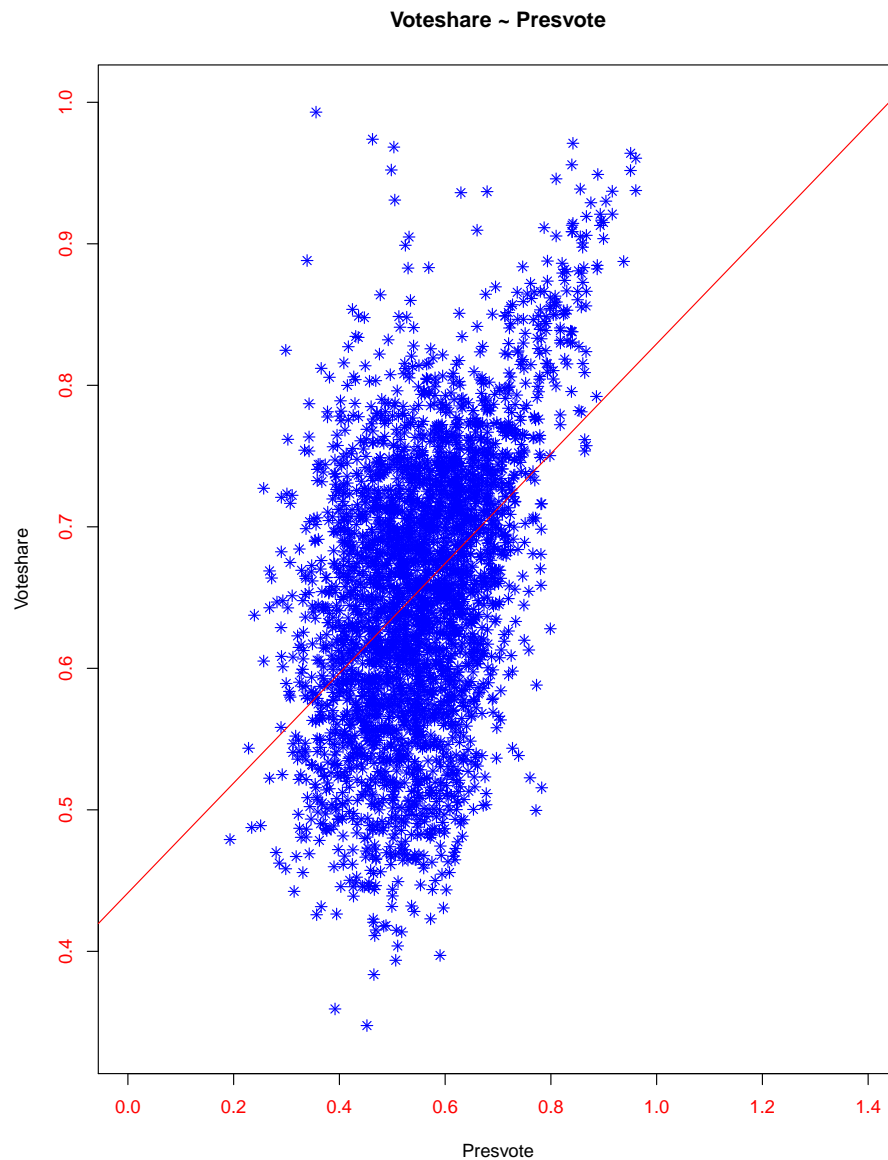
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
1 #1)  
2 linearMod3 <- lm(voteshare ~ presvote, data=data)  
3 linearMod3
```

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

```
1 #2)
2 plot(data$presvote , data$voteshare ,
3       xlab="Presvote" ,
4       ylab="Voteshare" ,
5       main="Voteshare ~ Presvote" , pch=8, col='blue' ,
6       col.axis='red' , xlim=c(0,1.4) , ylim=c(0.34,1))
7
8 abline(linearMod3 , col="red")
```



3. Write the prediction equation.

$$\text{Voteshare} = 0.4413 + 0.3880 * (\text{presvote})$$

## 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 #1)
2 linearMod4 <- lm(residual1 ~ residual2 , data=data)
3 linearMod4
4 summary(linearMod4)
```

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

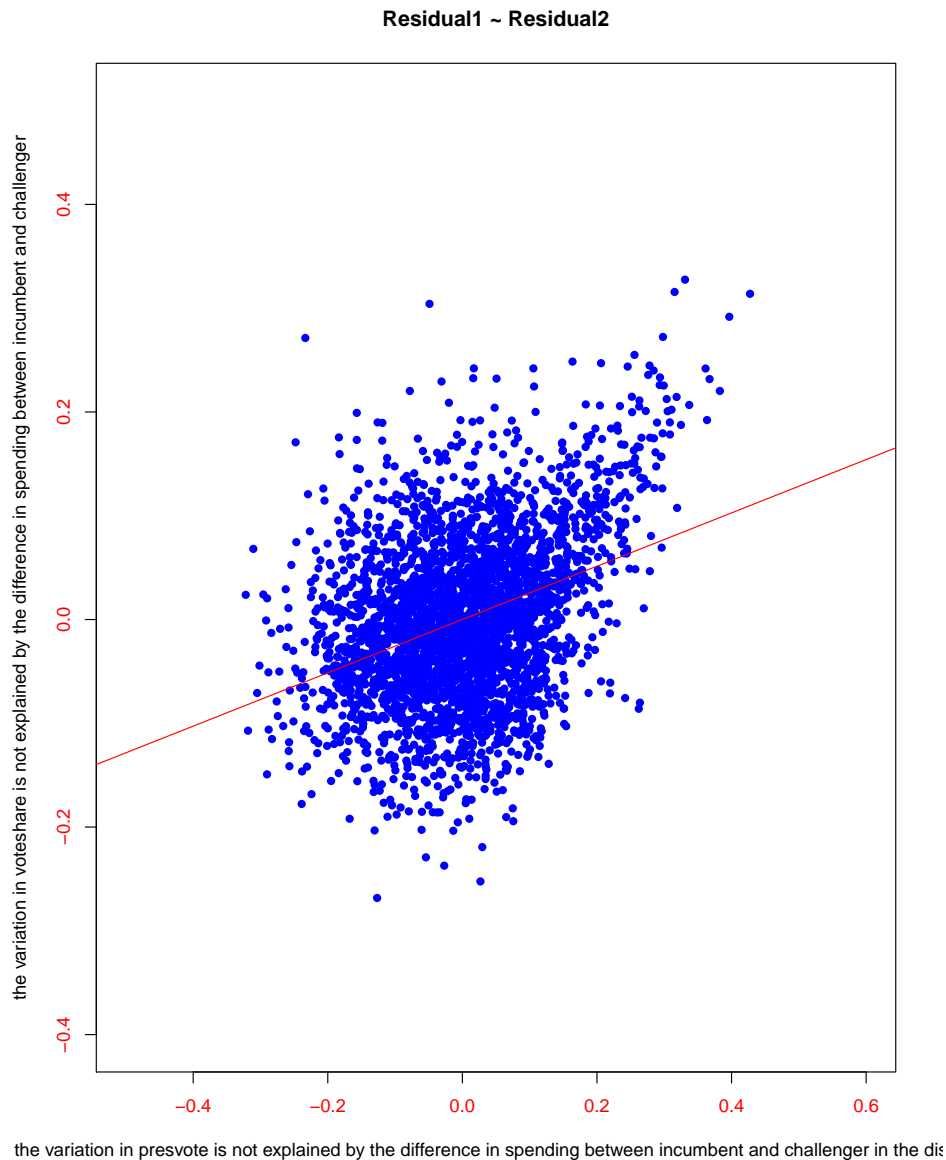
```
1 #2)
2 plot(residual2 , residual1 ,
3       xlab="the variation in presvote is not explained by the difference
4       in spending between incumbent and challenger in the district",
5       ylab="the variation in voteshare is not explained by the difference
6       in spending between incumbent and challenger",
7       main="Residual1 ~ Residual2" , pch=16, col='blue' ,
8       col.axis='red' , xlim=c(-0.5,0.6) , ylim=c(-0.4,0.5))
9 abline(linearMod4 , col="red")
```

3. Write the prediction equation.

The graph under this question is actually the graph for the previous question

$$\text{Residual1} = -4.860\text{e-}18 + 2.569\text{e-}01 * (\text{Residual2})$$





## 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 #1)
2 MultiLinear <- lm(voteshare ~ difflog + presvote, data=data)
3 MultiLinear
4 summary(MultiLinear)
```

2. Write the prediction equation.

$$\text{Voteshare} = 0.44864 + 0.03554 * (\text{difflog}) + 0.25688 * (\text{presvote})$$

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

Residual1 records the unexplained variability within the linear model built to explain voteshare to difflog

Residual2 records the unexplained variability within the linear model built to explain presvote to difflog

It has been proven that Residual1 could be modeled as a linear function by Residual2

The coefficient of Residual2 when trying to model Residual1 is identical to the coefficient of presvote when we are trying to construct a relationship between difflog and voteshare in a multiple regression

This is because during the actual construction of multiple regression, when we are trying to obtain the coefficient for each of the variable wrt the outcome, for example in this case when we are trying to find the relationship between difflog and voteshare in this multiple regression, we are holding every other variable in the regression constant and only study how would this variable change wrt to the outcome. So in this case because we are doing voteshare difflog, we are holding presvote constant. Due to the partial effect, the coefficient of presvote in this case is the "unexplained variability" in the relationship of voteshare to difflog.