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.

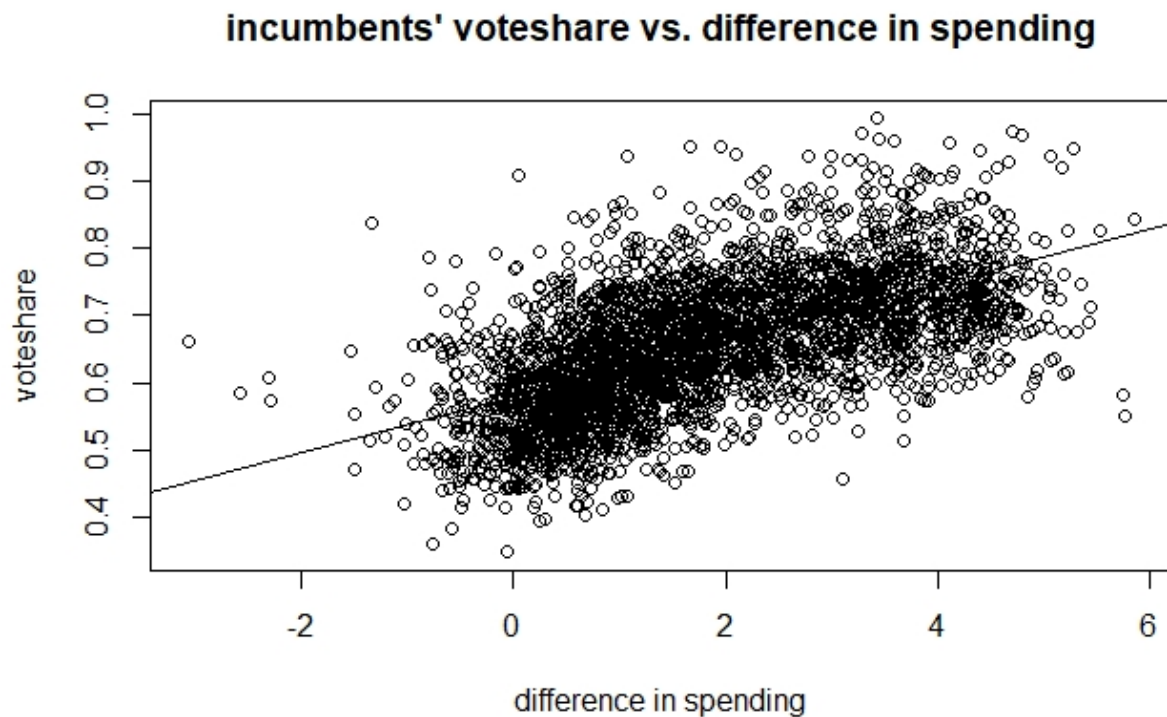1. Run a regression where the outcome variable is voteshare and the explanatory variable is difflog.

```
1  ###############Question 1###############
2  X1 <- incumb$difflog
3  Y1 <- incumb$voteshare
4  #1.regression
5  Reg1 <- lm(Y1~X1)
6  summary(Reg1)
7  #2.plot regression
8  plot(Y1~X1, main = "incumbents' voteshare vs. difference in spending",
       xlab = "difference in spending", ylab = "voteshare")
9  abline(Reg1)
```

```
1  #4.prediction equation
2  #Coefficients:
3    #Estimate Std. Error t value Pr(>|t|)
4  #(Intercept) 0.579031    0.002251   257.19    <2e-16 ***
5    #X1              0.041666    0.000968    43.04    <2e-16 ***
6
7  #Residual standard error: 0.07867 on 3191 degrees of freedom
8  #Multiple R-squared:   0.3673, Adjusted R-squared:   0.3671
9  #F-statistic:   1853 on 1 and 3191 DF,   p-value: < 2.2e-16
10
11 #############Question 2###############
```

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

incumbents' voteshare vs. difference in spending

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

```
1  #3.residual analysis
2  residual1 <- residuals(Reg1)
```

2

4. Write the prediction equation.

$$\bar{y} = \bar{\beta}_0 + \bar{\beta}_1 x_1 + \bar{\epsilon}_1$$

Where $\epsilon \sim N(0, \bar{\sigma}^2)$

We get from results in Question 1:

$$\bar{y} = 0.579 + 0.042 x_1 + \bar{\sigma}_1$$

Where $\epsilon \sim N(0, 0.\bar{0}79^2)$

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 ##############Question 2###############
2 X2 <- incumb$difflog
3 Y2 <- incumb$presvote
4 #1.regression
5 Reg2 <- lm(Y2~X2)
6 summary(Reg2)
```

```
1 #Coefficients:
2   #Estimate Std. Error t value Pr(>|t|)
3 #(Intercept) 0.507583   0.003161  160.60   <2e-16 ***
4   #X2         0.023837   0.001359   17.54   <2e-16 ***
5
6 #Residual standard error: 0.1104 on 3191 degrees of freedom
7 #Multiple R-squared:  0.08795,   Adjusted R-squared:   0.08767
8 #F-statistic: 307.7 on 1 and 3191 DF,  p-value: < 2.2e-16
9
10 ##############Question 3###############
```

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

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

```
1 #3.residual analysis
2 residual2 <- residuals(Reg2)
```

4. Write the prediction equation.

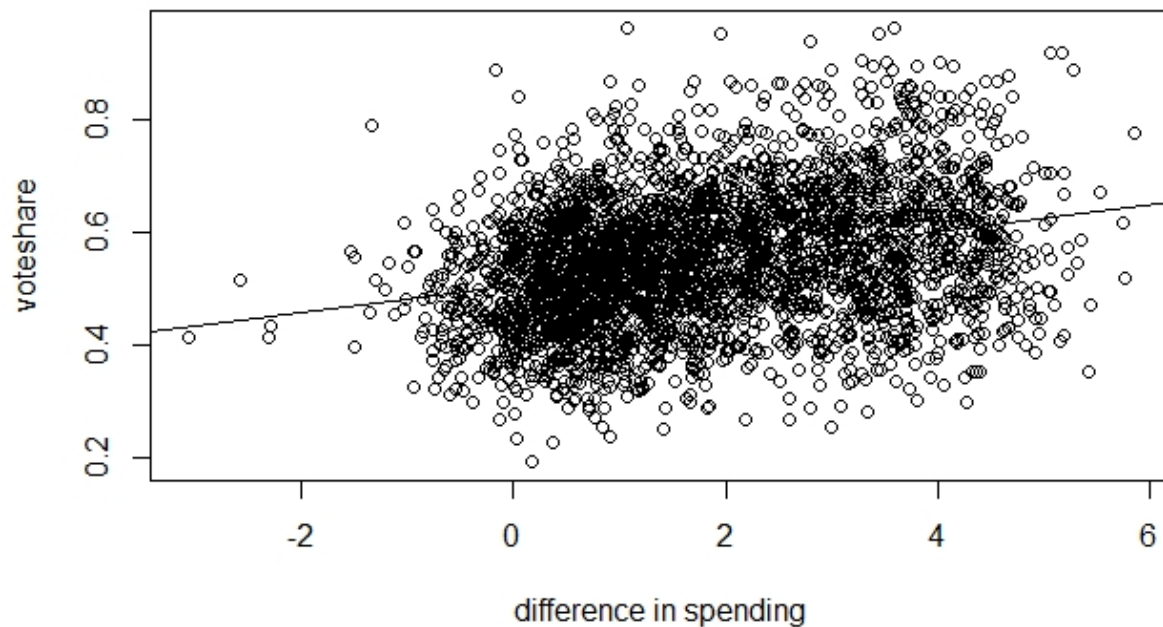$$\bar{y} = \bar{\beta}_0 + \bar{\beta}_1 x_1 + \bar{\epsilon}_1$$

Where $\epsilon \sim N(0, \bar{\sigma}^2)$

We get from results in Question 1:

$$\bar{y} = 0.507 + 0.023x_1 + \bar{\sigma}_1$$

Where $\epsilon \sim N(0, 0.\bar{1}104^2)$

Incumbents' presidential vote vs. difference in spending
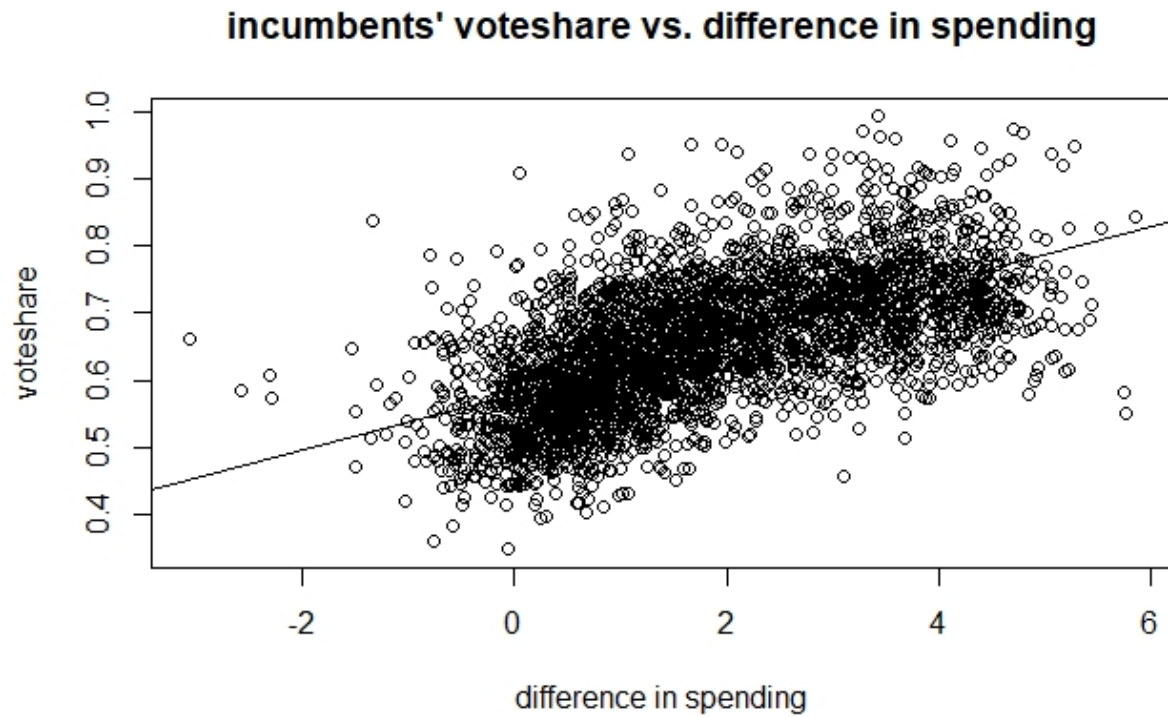

Question 3 (20 points)

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 `presvote` and the explanatory variable is `voteshare`.

```
1 ##############Question 3##############
2 X3 <- incumb$presvote
3 Y3 <- incumb$voteshare
4 #1.regression
5 Reg3 <- lm(Y3~X3)
6 summary(Reg3)
```

```
1 #Coefficients:
2   #Estimate Std. Error  t value Pr(>|t|)
3 #(Intercept) 0.441330   0.007599    58.08   <2e-16 ***
4   #X3         0.388018   0.013493    28.76   <2e-16 ***
5
6 #Residual standard error: 0.08815 on 3191 degrees of freedom
7 #Multiple R-squared:  0.2058, Adjusted R-squared:  0.2056
8 #F-statistic:    827 on 1 and 3191 DF,  p-value: < 2.2e-16
```

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

incumbents' voteshare vs. difference in spending

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

```
1 #3. residual analysis
2 residual3 <- residuals(Reg3)
```

4. Write the prediction equation.

$$\bar{y} = \bar{\beta}_0 + \bar{\beta}_1 x_1 + \bar{\epsilon}_1$$

Where $\epsilon \sim N(0, \bar{\sigma}^2)$

We get from results in Question 1:

$$\bar{y} = 0.441 + 0.388 x_i + \bar{\epsilon}_1$$

Where $\epsilon \sim N(0, 0.\bar{0}88^2)$

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  ##############Question 4##############
2  X4 <- residual2
3  Y4 <- residual1
4  #1.regression
5  Reg4 <- lm(Y4~X4)
6  summary(Reg4)
```

```
1  #Coefficients:
2    #Estimate Std. Error t value Pr(>|t|)
3  #(Intercept) -4.860e-18  1.299e-03    0.00        1
4  #X4           2.569e-01  1.176e-02   21.84    <2e-16 ***
5
6  #Residual standard error: 0.07338 on 3191 degrees of freedom
7  #Multiple R-squared:   0.13,   Adjusted R-squared:   0.1298
8  #F-statistic:   477 on 1 and 3191 DF,  p-value: < 2.2e-16
9
10 ##############Question 5##############
```

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

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

```
1  #3.residual analysis
2  residual4 <- residuals(Reg4)
```
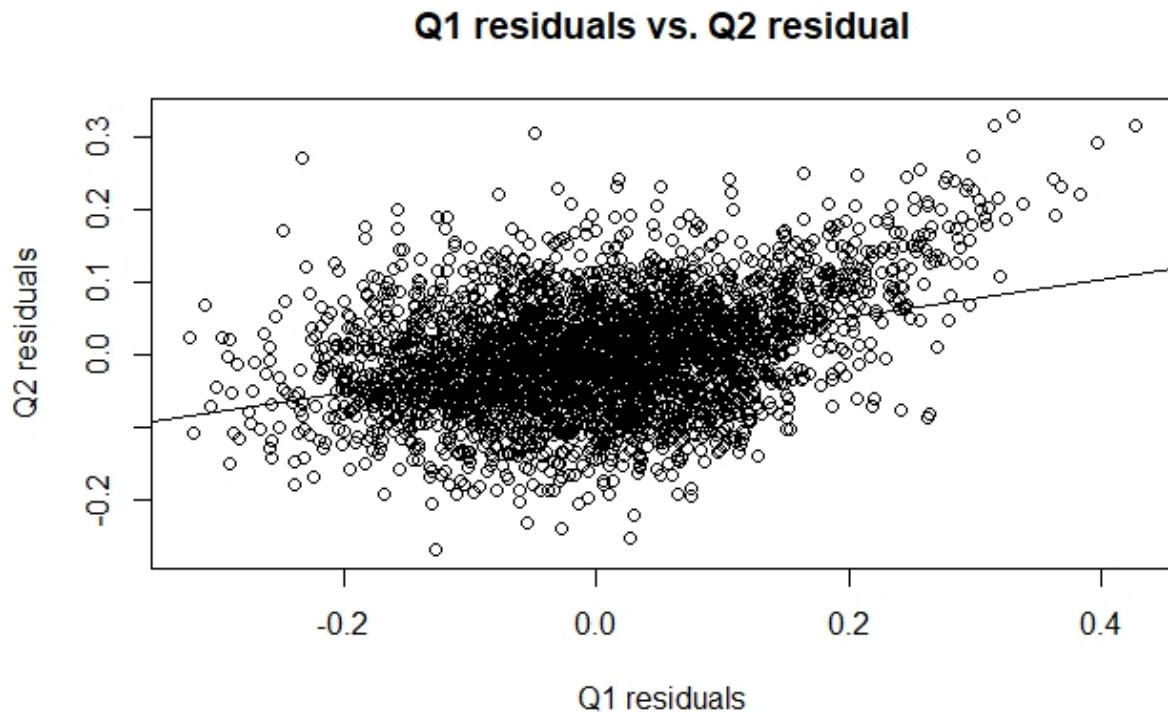
4. Write the prediction equation.

$$\bar{y} = \bar{\beta}_0 + \bar{\beta}_1 x_1 + \bar{\epsilon}_1$$

Where $\epsilon \sim N(0, \bar{\sigma}^2)$

We get from results in Question 1:

$$\bar{y} = -4.86e - 18 + 2.569e - 1x_1 + \bar{\epsilon}_1$$

Where $\epsilon \sim N(0, 0.0\bar{7}338^2)$

Q1 residuals vs. Q2 residual

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  ###############Question 5###############
2  X5a <- incumb$difflog
3  X5b <- incumb$presvote
4  Y5 <- incumb$voteshare
5  #1.regression
6  Reg5 <- lm(Y5~X5a+X5b)
7  #2.prediction equation
8  summary(Reg5)
9
10 #Coefficients:
11   #Estimate Std. Error  t value Pr(>|t|)
12 #(Intercept) 0.4486442  0.0063297   70.88   <2e-16 ***
13   #X5a         0.0355431  0.0009455   37.59   <2e-16 ***
14   #X5b         0.2568770  0.0117637   21.84   <2e-16 ***
15
16 #Residual standard error: 0.07339 on 3190 degrees of freedom
17 #Multiple R-squared:  0.4496, Adjusted R-squared:  0.4493
```

2. Write the prediction equation.

$$\bar{y} = \bar{\beta}_0 + \bar{\beta}_1 x_1 + \bar{\beta}_2 x_2 + \bar{\epsilon}_i$$

Where $\epsilon \sim N(0, \bar{\sigma}^2)$

We get from results in Question 1:

$$\bar{y} = 0.449 + 0.036x_1 + 0.257x_2 + \bar{\epsilon}_i$$

Where $\epsilon \sim N(0, 0.0\bar{7}339^2)$

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 Residual standard error is the same when rounded to third digits 0.0734