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`.

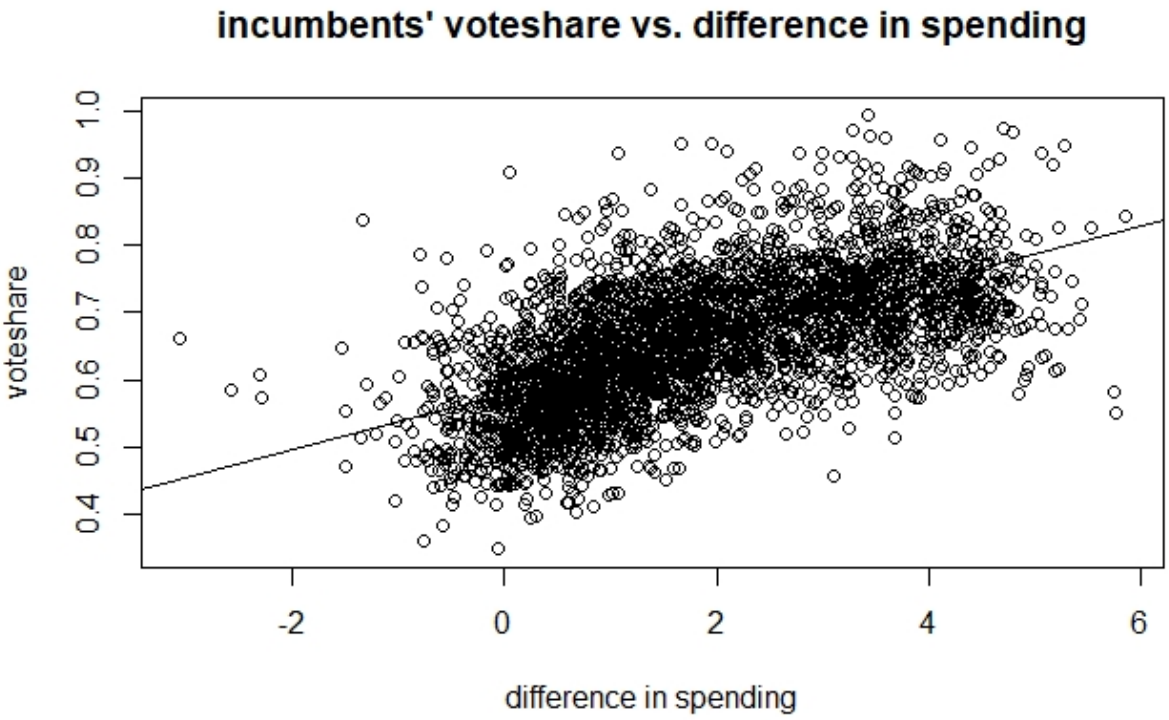
```
##############Question 1################
X1 <- incumb$difflog
Y1 <- incumb$voteshare
#1.regression
Reg1 <- lm(Y1~X1)
summary(Reg1)
#2.plot regression
plot(Y1~X1, main = "incumbents' voteshare vs. difference in spending",
    xlab = "difference in spending", ylab = "voteshare")
abline(Reg1)
```

```r
#4.prediction equation
#Coefficients:
  #Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.579031   0.002251  257.19   <2e-16 ***
  #X1          0.041666   0.000968   43.04   <2e-16 ***

#Residual standard error: 0.07867 on 3191 degrees of freedom
#Multiple R-squared:  0.3673, Adjusted R-squared:  0.3671
#F-statistic:  1853 on 1 and 3191 DF,  p-value: < 2.2e-16

##############Question 2################
```

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



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

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

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:

$$\bar{y} = 0.579 + 0.042 x_1 + \bar{\epsilon_1}$$

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`.

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

```
#Coefficients:
  #Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.507583   0.003161  160.60   <2e-16 ***
  #X2          0.023837   0.001359   17.54   <2e-16 ***

#Residual standard error: 0.1104 on 3191 degrees of freedom
#Multiple R-squared:  0.08795,  Adjusted R-squared:  0.08767
#F-statistic: 307.7 on 1 and 3191 DF,  p-value: < 2.2e-16

##############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.

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

4. Write the prediction equation.

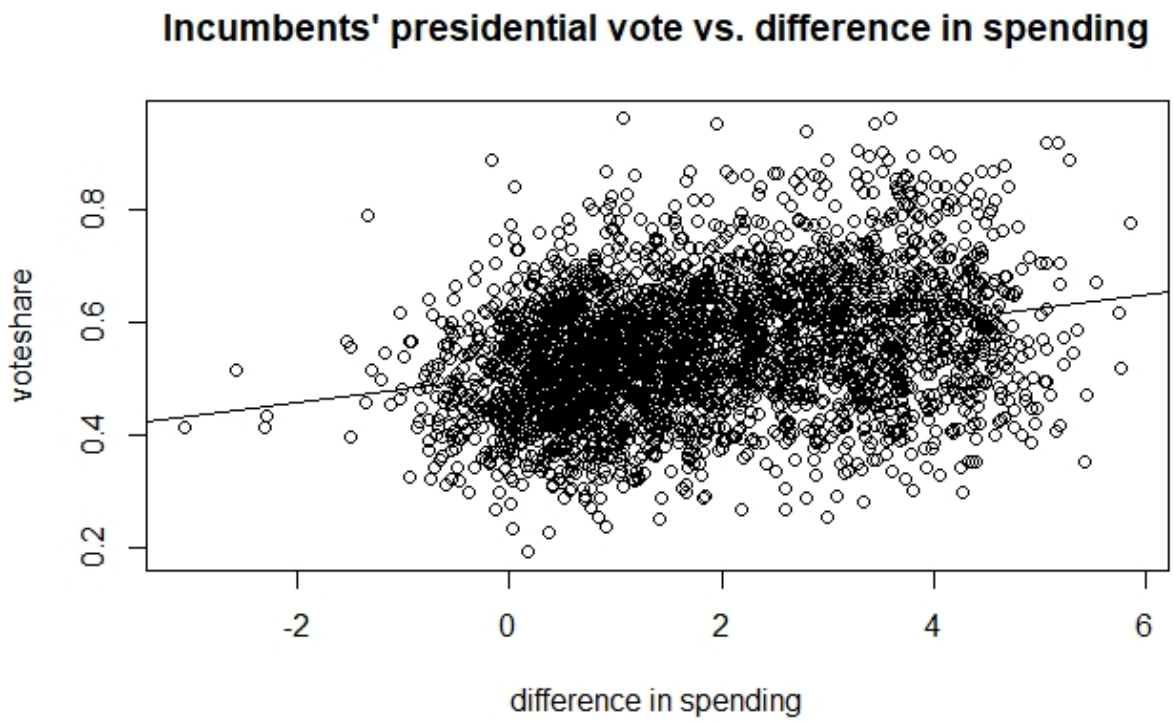
$$\bar{y} = \bar{\beta_0} + \bar{\beta_1} x_1 + \bar{\epsilon_1}$$

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

We get from results in Question 1:

$$\bar{y} = 0.507 + 0.023 x_1 + \bar{\epsilon_1}$$

Where $\epsilon \sim N(0, 0.1\bar{1}04^2)$



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.

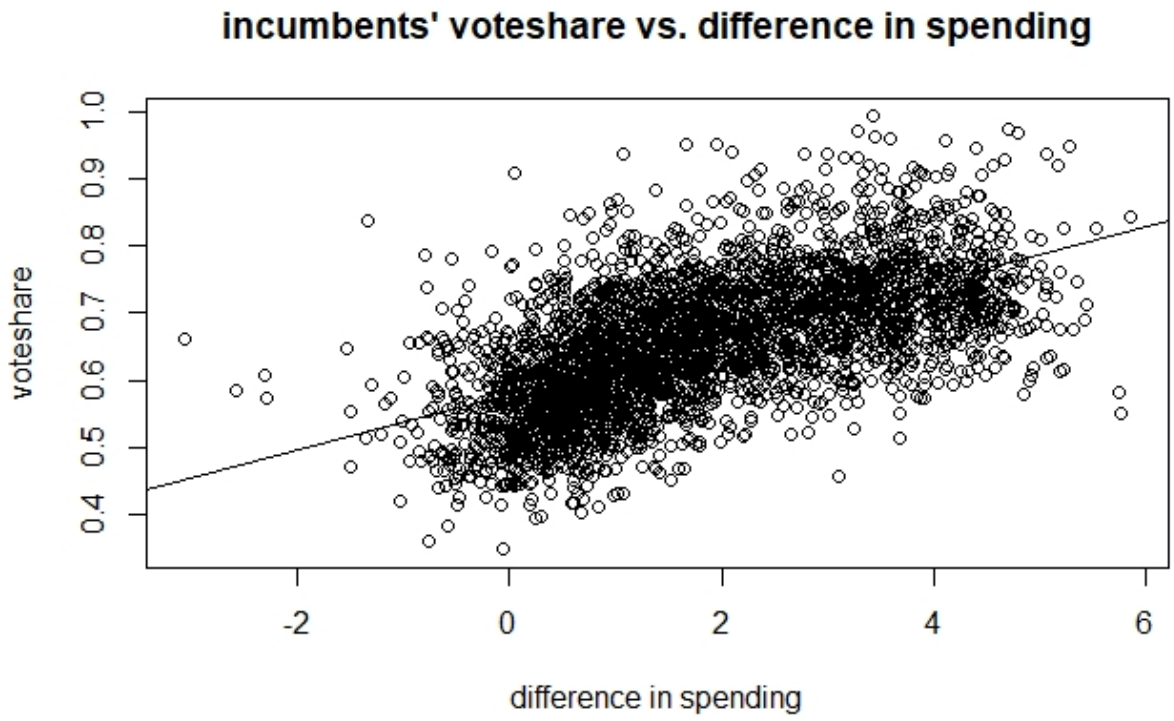
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `voteshare`.

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

```
#Coefficients:
  #Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.441330   0.007599   58.08   <2e-16 ***
  #X3          0.388018   0.013493   28.76   <2e-16 ***

#Residual standard error: 0.08815 on 3191 degrees of freedom
#Multiple R-squared:  0.2058, Adjusted R-squared:  0.2056
#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.



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

```r
#3.residual analysis
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.

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

```
#Coefficients:
  #Estimate Std. Error t value Pr(>|t|)
#(Intercept) -4.860e-18  1.299e-03    0.00        1
#X4           2.569e-01  1.176e-02   21.84   <2e-16 ***

#Residual standard error: 0.07338 on 3191 degrees of freedom
#Multiple R-squared:   0.13,   Adjusted R-squared:  0.1298
#F-statistic:   477 on 1 and 3191 DF,  p-value: < 2.2e-16

##############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.

```
#3.residual analysis
residual4 <- residuals(Reg4)
```

4. Write the prediction equation.

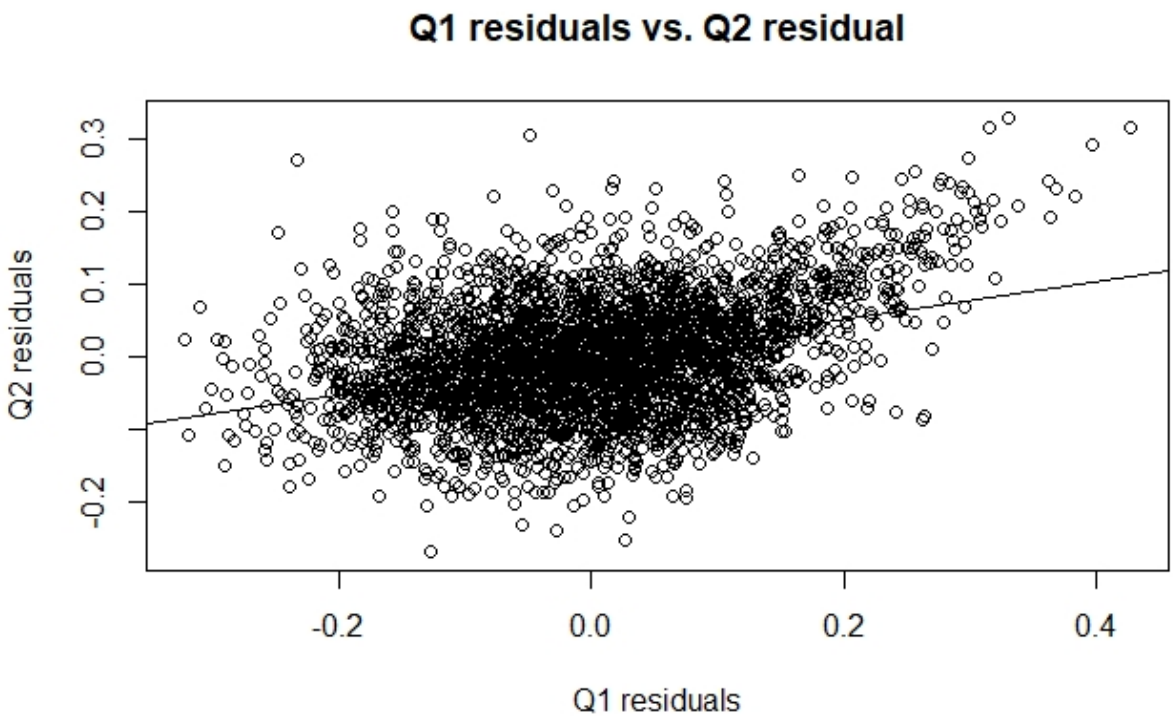
$$\bar{y} = \bar{\beta_0} + \bar{\beta_1} x_1 + \bar{\epsilon_1}$$

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

We get from results in Question 1:

$$\bar{y} = -4.86e-18 + 2.569e-1 x_1 + \bar{\epsilon_1}$$

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



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`.

```
##############Question 5###############
X5a <- incumb$difflog
X5b <- incumb$presvote
Y5 <- incumb$voteshare
#1.regression
Reg5 <- lm(Y5~X5a+X5b)
#2.prediction equation
summary(Reg5)

#Coefficients:
  #Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.4486442  0.0063297   70.88   <2e-16 ***
  #X5a         0.0355431  0.0009455   37.59   <2e-16 ***
  #X5b         0.2568770  0.0117637   21.84   <2e-16 ***

#Residual standard error: 0.07339 on 3190 degrees of freedom
#Multiple R-squared:  0.4496, Adjusted R-squared:  0.4493
```

```
#F-statistic:  1303 on 2 and 3190 DF,  p-value: < 2.2e-16
```

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:

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

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