

# 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 incumbentsdata <- read.csv("incumbents_subset.csv")
2 incumbentsdata
3 summary(incumbentsdata)
4 regression1 <- lm(voteshare ~ difflog, data=incumbentsdata)
5 regression1
```

```
lm(formula = voteshare ~ difflog, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.26832	-0.05345	-0.00377	0.04780	0.32749

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.579031	0.002251	257.19 <2e-16 ***
difflog	0.041666	0.000968	43.04 <2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

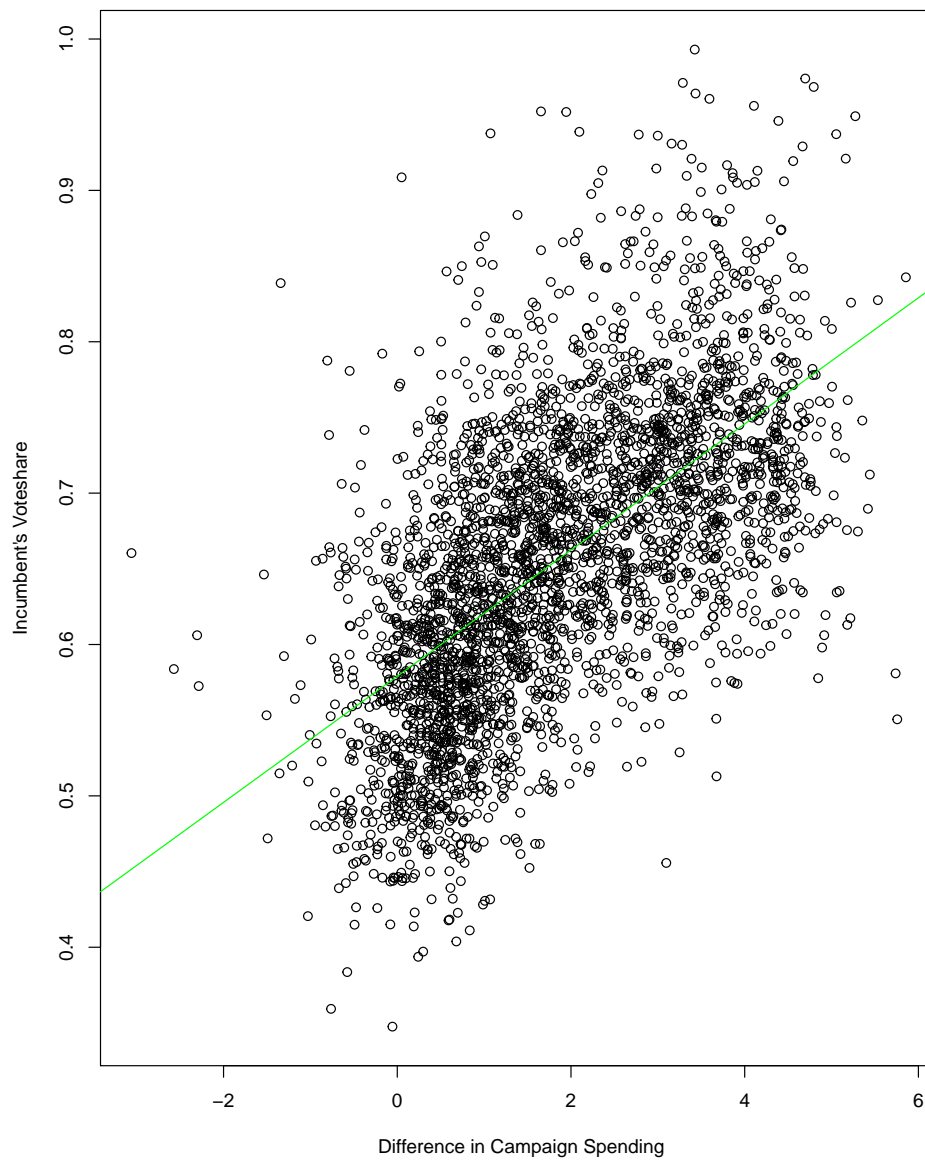
Residual standard error: 0.07867 on 3191 degrees of freedom

Multiple R-squared: 0.3673, Adjusted R-squared: 0.3671

F-statistic: 1853 on 1 and 3191 DF, p-value: < 2.2e-16

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

```
1 plot(incumbentsdata$difflog, incumbentsdata$voteshare, ylab = "Incumbent '  
   s Voteshare", xlab = "Difference in Campaign Spending")  
2 abline(regression1, col="green")
```



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

```
1 residualsregression1 <- residuals(regression1)
2 residualsregression1
```

4. Write the prediction equation.

$$\hat{Y} = 0.579031 + 0.041666X$$

## 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 regression2 <- lm(presvote ~ difflog, data=incumbentsdata)
2 regression2
3 summary(regression2)
```

```
lm(formula = presvote ~ difflog, data = incumbentsdata)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.32196	-0.07407	-0.00102	0.07151	0.42743

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.507583	0.003161	160.60	<2e-16 ***
difflog	0.023837	0.001359	17.54	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

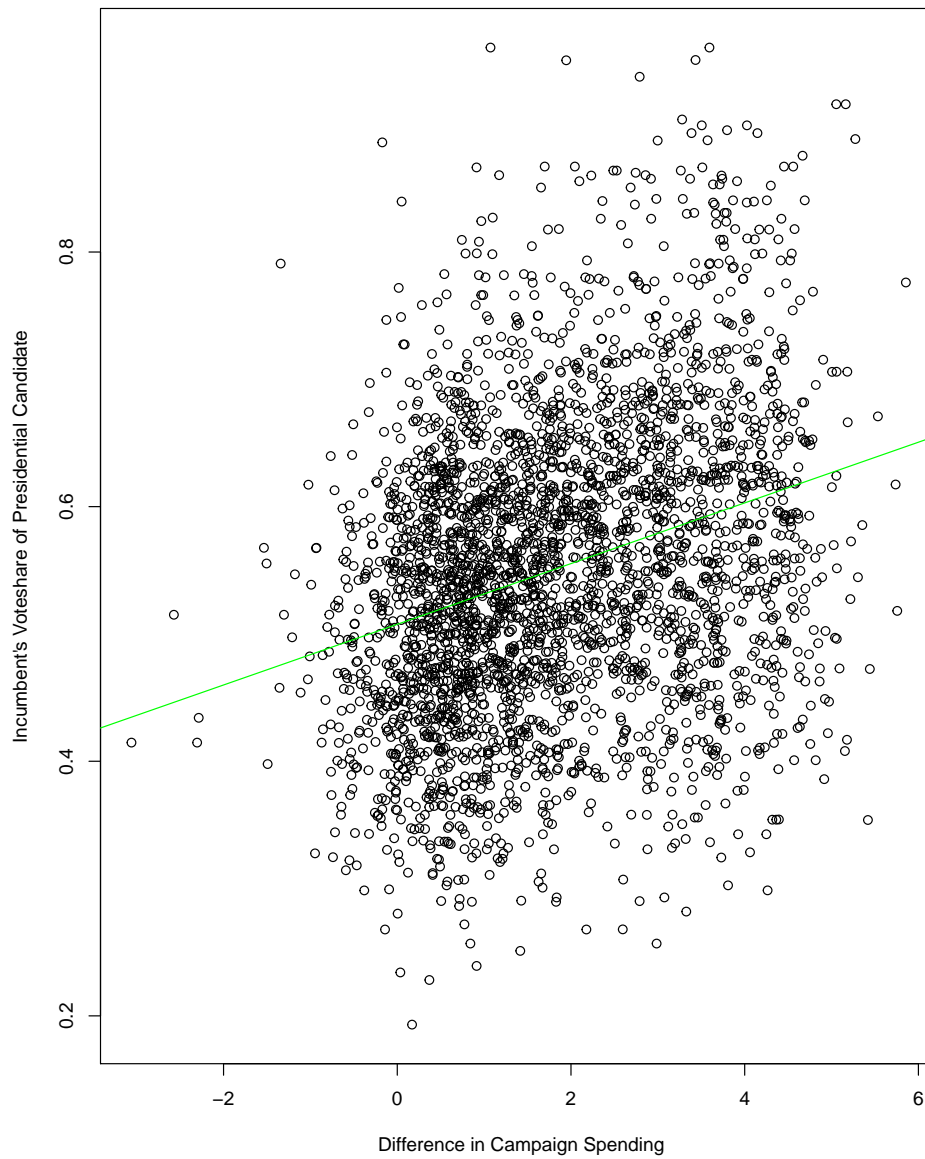
Residual standard error: 0.1104 on 3191 degrees of freedom

Multiple R-squared: 0.08795, Adjusted R-squared: 0.08767

F-statistic: 307.7 on 1 and 3191 DF, p-value: < 2.2e-16

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

```
1 str(incumbentsdata)
2 plot(incumbentsdata$difflog, incumbentsdata$presvote, ylab = "Incumbent's
      Voteshare of Presidential Candidate", xlab = "Difference in Campaign
      Spending")
3 abline(regression2, col="green")
```



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

```
1 residualsregression2 <- residuals(regression2)
2 residualsregression2
```

4. Write the prediction equation.

$$\hat{Y} = 0.507583 + 0.023837X$$

## 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 regression3 <- lm(voteshare ~ presvote, data=incumbentsdata)
2 regression3
3 summary(regression3)
```

```
lm(formula = voteshare ~ presvote, data = incumbentsdata)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.27330	-0.05888	0.00394	0.06148	0.41365

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.441330	0.007599	58.08	<2e-16 ***
presvote	0.388018	0.013493	28.76	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

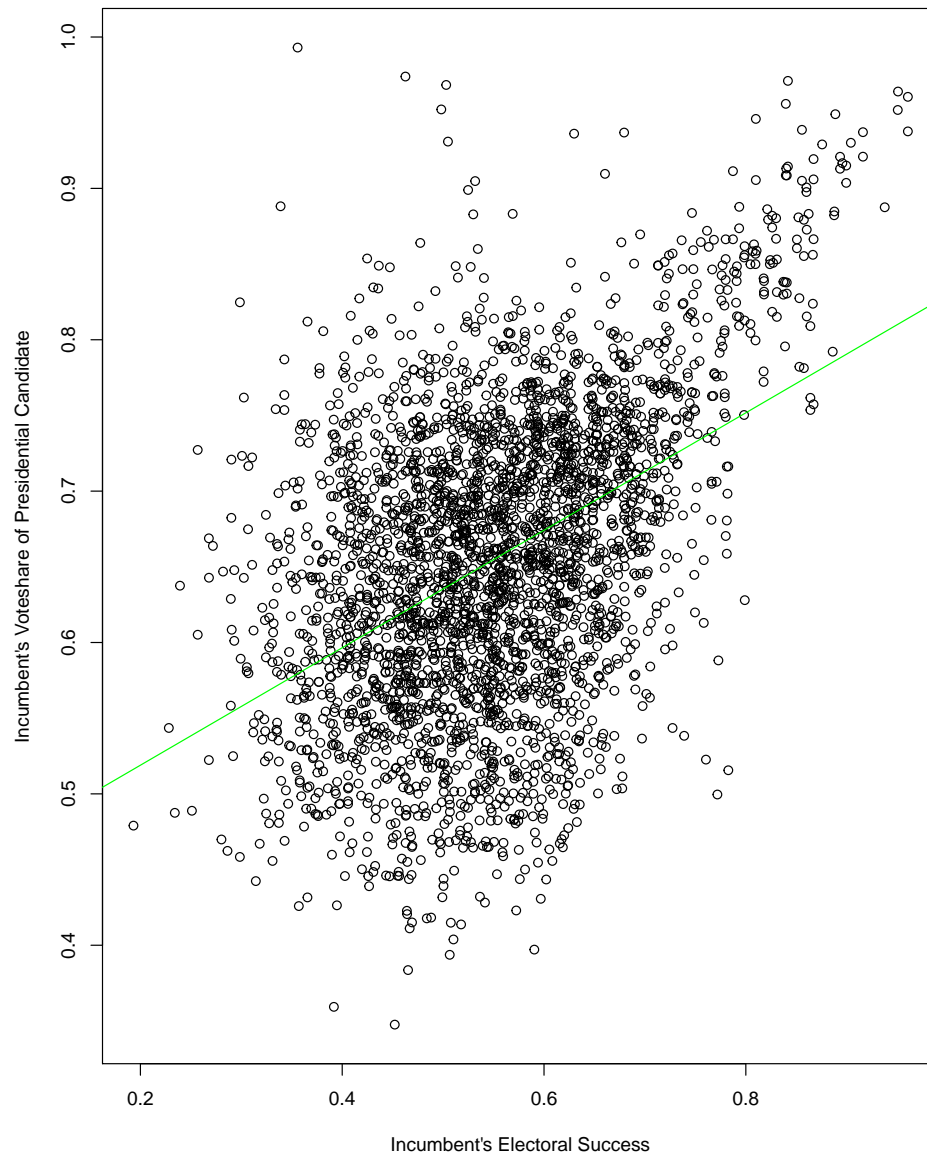
Residual standard error: 0.08815 on 3191 degrees of freedom

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.

```
1 str(incumbentsdata)
2 plot(incumbentsdata$presvote, incumbentsdata$voteshare, ylab = "Incumbent
    's Voteshare of Presidential Candidate", xlab = "Incumbent's Electoral
    Success")
3 abline(regression3, col="green")
```



3. Write the prediction equation.

$$\hat{Y} = 0.441330 + 0.388018X$$

## 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 regression4 <- lm(residualsregression1 ~ residualsregression2, data=
  incumbentsdata)
2 regression4
3 summary(regression4)
```

```
lm(formula = residualsregression1 ~ residualsregression2, data = incumbentsdata)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.25928	-0.04737	-0.00121	0.04618	0.33126

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-4.860e-18	1.299e-03	0.00
residualsregression2	2.569e-01	1.176e-02	21.84

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07338 on 3191 degrees of freedom

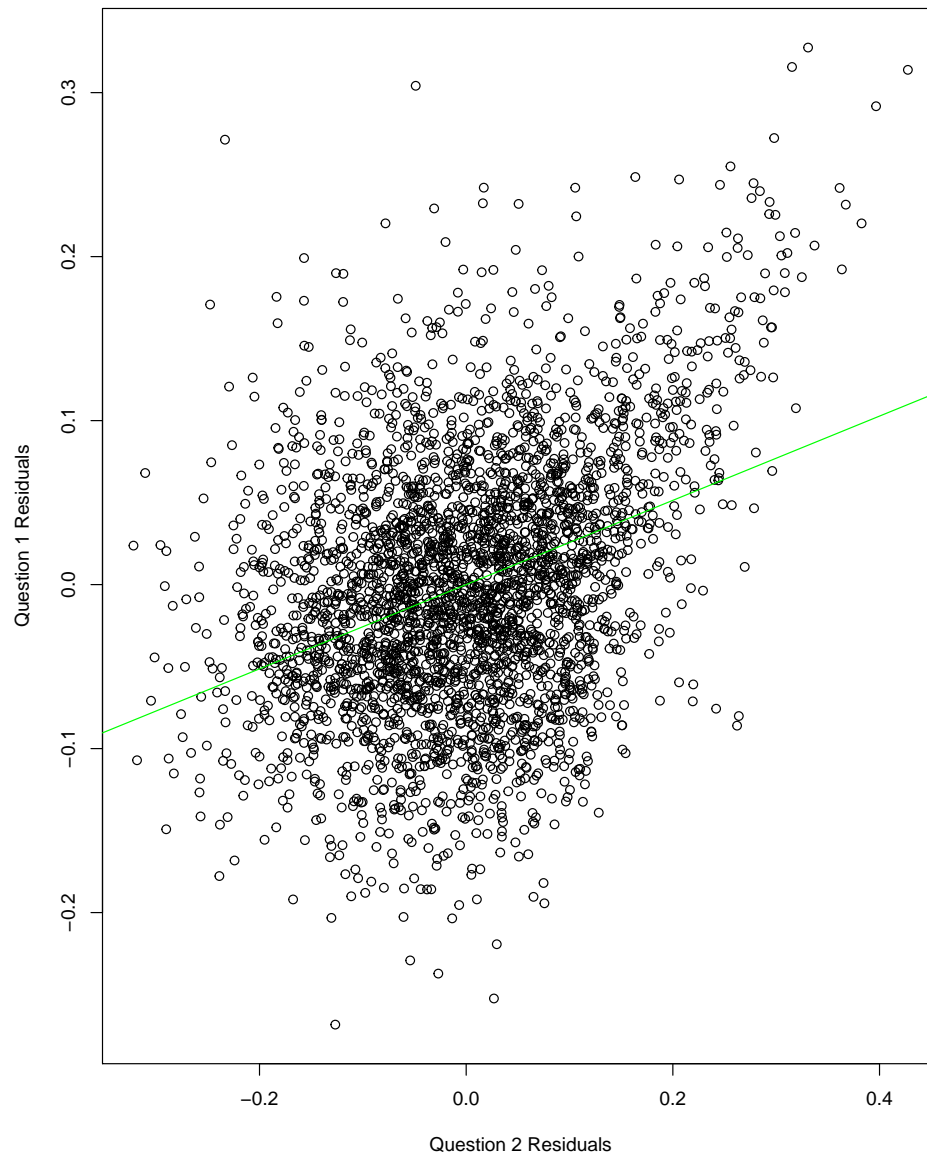
Multiple R-squared: 0.13, Adjusted R-squared: 0.1298

F-statistic: 477 on 1 and 3191 DF, p-value: < 2.2e-16

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

```
1 str(incumbentsdata)
2 plot(residualsregression2, residualsregression1, ylab = "Question 1
  Residuals", xlab = "Question 2 Residuals")
3 abline(regression4, col="green")
```





3. Write the prediction equation.

$$\hat{Y} = -4.860e-18 + 2.569e-01X$$

## 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 regression5 <- lm(voteshare ~ difflog+presvote , data=incumbentsdata)
2 regression5
3 summary(regression5)
```

```
lm(formula = voteshare ~ difflog + presvote, data = incumbentsdata)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.25928	-0.04737	-0.00121	0.04618	0.33126

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.4486442	0.0063297	70.88 <2e-16 ***
difflog	0.0355431	0.0009455	37.59 <2e-16 ***
presvote	0.2568770	0.0117637	21.84 <2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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.

$$\hat{Y} = 0.4486442 + 0.0355431X_1 + 0.2568770X_2$$

$Y$  = voteshare

$X_1$  = difflog and  $X_2$  = 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?

Reflected.