QTM 200: Applied Regression Analysis: Problem Set 3

Farris Sabir

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`.

```
## Part 1

#read incumbents subset data
incumbents <- read.csv("incumbents_subset.csv")
#run regression model with incumbent's vote share regressed on difference
    in campaign spending
incumbents_model <- lm(voteshare ~ difflog, data=incumbents)
#get summary of model
summary(incumbents_model)
```

```
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.

```r
## Part 2

#produce scatterplot
plot(incumbents$difflog, incumbents$voteshare, xlab = "Logarithmic
    Difference in Campaign Spending", ylab = "Incumbent Vote Share (%)")
#add regression line
abline(0.579031, 0.041666)
```

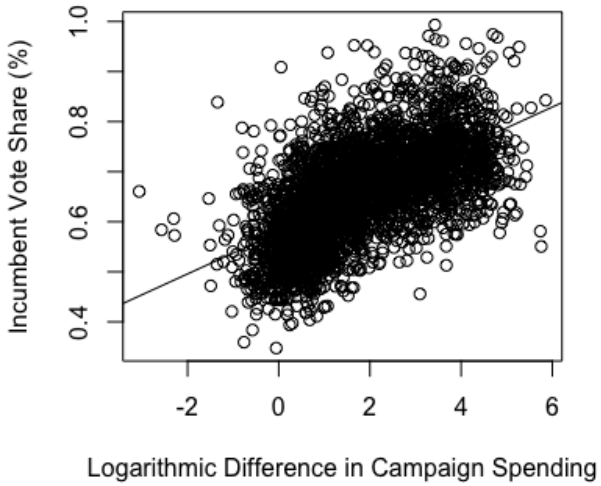
The scatterplot and regression line are shown in Figure 1.

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

```r
## Part 3

residual1 <- incumbents$voteshare - (0.5970321 + 0.041666*incumbents$
    difflog)
residual1 <- incumbents_model$residuals
```

4. Write the prediction equation. Prediction Equation: $\hat{Y} = 0.5970321 + 0.041666X$

Figure 1: Difference in Campaign Spending vs. Incumbent Vote Share



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`.

```
## Part 1

#run regression model with vote share of the presidential candidate of
    incumbent party regressed on difference in campaign spending
incumbents_model2 <- lm(presvote ~ difflog , data=incumbents)
#get summary of model
summary(incumbents_model2)
```

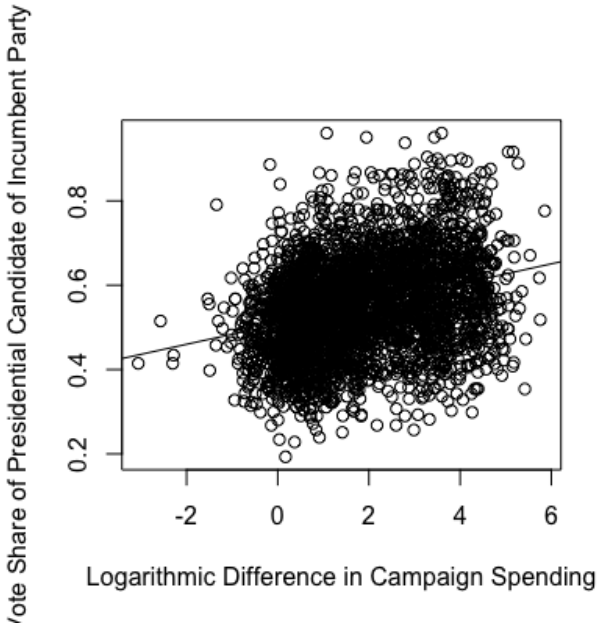
```
Residuals:
     Min       1Q   Median       3Q      Max
-0.32196 -0.07407 -0.00102  0.07151  0.42743

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.

```
## Part 2

#produce scatterplot
plot(incumbents$difflog, incumbents$presvote, xlab = "Logarithmic
    Difference in Campaign Spending", ylab = "Vote Share of Presidential
    Candidate of Incumbent Party (%)")
#add regression line
abline(0.507583, 0.023837)
```

Figure 2: Difference in Campaign Spending vs. Vote Share of Candidate of Incumbent Party



The scatterplot and regression line are shown in Figure 2.

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

```
## Part 3

residual2 <- incumbents$presvote - 0.407583 - 0.023837*incumbents$difflog
residual2 <- incumbents_model2$residuals
```

4. Write the prediction equation. 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`.

```
## Part 1

#run regression model with vote share of the presidential candidate of
    incumbent party regressed on incumbent's electoral success
incumbents_model3 <- lm(voteshare ~ presvote, data=incumbents)
#get summary of model
summary(incumbents_model3)
```

```
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
```

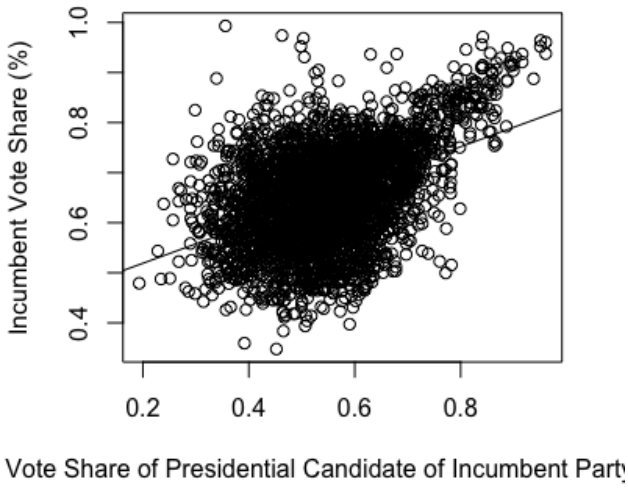
2. Make a scatterplot of the two variables and add the regression line.

```
## Part 2

#produce scatterplot
plot(incumbents$presvote, incumbents$voteshare, ylab = "Incumbent Vote
    Share (%)", xlab = "Vote Share of Presidential Candidate of Incumbent
    Party (%)")
#add regression line
abline(0.441330, 0.388018)
```

The scatterplot and regression line are shown in Figure 3.

3. Write the prediction equation. Prediction Equation: $\hat{Y} = 0.441330 + 0.388018X$

Figure 3: Vote Share of Candidate of Incumbent Party vs. Incumbent's Electoral Success



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.

```
## Part 1

#run regression model with variation in voteshare not explained by
    difference in spending and variation in incumbent electoral success
    not explained by difference in spending
incumbents_model4 <- lm(residual1 ~ residual2)
#get summary of model
summary(incumbents_model4)
```

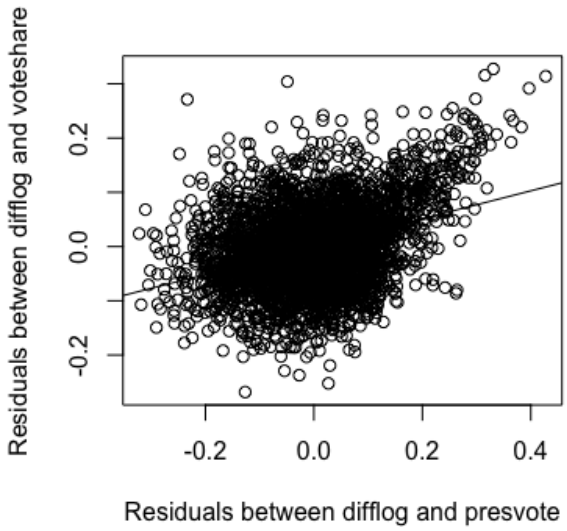
```
Residuals:
     Min       1Q   Median       3Q      Max
-0.25928 -0.04737 -0.00121  0.04618  0.33126

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.860e-18  1.299e-03    0.00        1
```

```
residual2    2.569e-01  1.176e-02   21.84   <2e-16 ***
---
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.

```
## Part 2

#produce scatterplot
plot(residual2, residual1, xlab = "Residuals between difflog and presvote
    ", ylab = "Residuals between difflog and voteshare")
#add regression line
abline(-4.860e-18, 2.569e-01)
```

Figure 4: Question 2 Residuals vs. Question 1 Residuals



The scatterplot and regression line are shown in Figure 4.

3. Write the prediction equation. Prediction Equation: $\hat{Y} = -4.860 * 10^{-}18 + 0.2569X$

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`.

```
## Part 1

#run regression model with incumbent's vote share against difference in
    campaign spending and vote share of presidential candidate of
    incumbent party
incumbents_model5 <- lm(voteshare ~ difflog + presvote, data=incumbents)
#get summary of model
summary(incumbents_model5)
```

```
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. Prediction Equation: $\hat{Y} = 0.4486442 + 0.0355431X_1 + 0.2568770X_2$

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

   (Reflection hidden)