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 # Use the lm() function to run a regression analysis on the relationship
      between difflog and voteshare, with voteshare as the outcome variable
      (y) and difflog as the explanatory variable (x)
2 > difflog_voteshare_regression<- lm(incumbents_subset$voteshare~
      incumbents_subset$difflog, data = incumbents_subset)
3 > summary(difflog_voteshare_regression)
4
```

```
 5  Call:
 6  lm(formula = incumbents_subset$voteshare ~ incumbents_subset$difflog,
 7  data = incumbents_subset)
 8
 9  Residuals:
10  Min          1Q     Median          3Q         Max
11  −0.26832  −0.05345  −0.00377    0.04780    0.32749
12
13  Coefficients:
14  Estimate Std. Error t value Pr(>|t|)
15  (Intercept)                 0.579031    0.002251    257.19   <2e−16 ***
16  incumbents_subset$difflog  0.041666    0.000968     43.04   <2e−16 ***
17
18  Residual standard error: 0.07867 on 3191 degrees of freedom
19  Multiple R−squared:  0.3673,     Adjusted R−squared:  0.3671
20  F−statistic:   1853 on 1 and 3191 DF,   p−value: < 2.2e−16
21
22  # The p−value is very small, so reject the null hypothesis that there is
        no relationship between the difference in spending between the
        incumbent and the challenger and the incumbent's vote share. It is
        likely that the difference in spending is related to the incumbent's
        vote share. The intercept of the regression is 0.579031, meaning that
        when the difference in spending between the incumbent and the
        challenger is 0, the incumbent's vote share is 0.579031. The slope of
        the regression is 0.041666, meaning that the incumbents value of the
        voteshare increases by 0.041666 for every one unit that the difference
         in spending between the incumbent and the challenger increases.
```
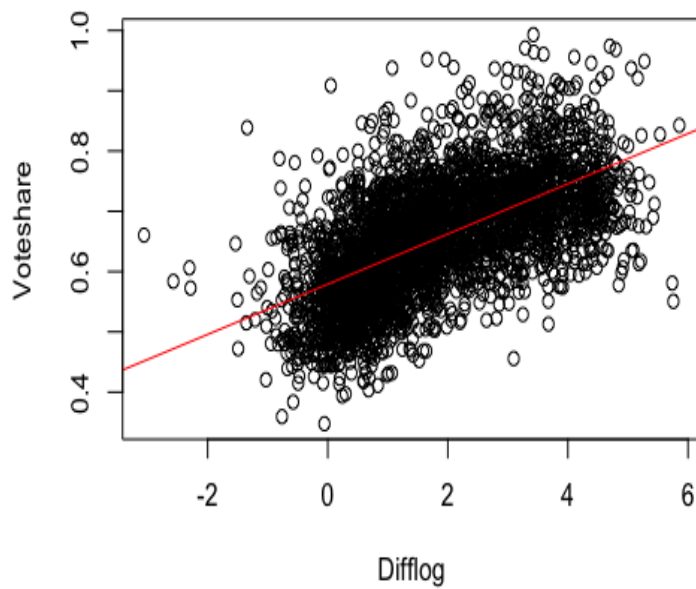
2. Make a scatterplot of the two variables and add the regression line.

```
1  # Make a scatterplot of the relationship between difflog and preshare
2  plot(incumbents_subset$difflog,incumbents_subset$voteshare, xlab = "
       Difflog", ylab = "Voteshare")
3  # Add the regression line
4  abline(difflog_voteshare_regression, col= "red")
```

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

```
1 # Create an object for the residuals of the regression
2 residuals_difflog_voteshare<- residuals(difflog_voteshare_regression)
3 residuals_difflog_voteshare
```

4. Write the prediction equation.

```
1 # Y = beta0 + beta1 * x
2 # beta0 = intercept
3 # beta1 = slope
4 # SOLUTION: 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`.

```
# Use the lm() function to run a regression analysis on the relationship
    between difflog and presvote, with presvote as the outcome variable (y
    ) and difflog as the explanatory variable (x)
> difflog_presvote_regression<- lm(incumbents_subset$presvote~incumbents_
    subset$difflog, data = incumbents_subset)
> summary(difflog_presvote_regression)
Call:
lm(formula = incumbents_subset$presvote ~ incumbents_subset$difflog,
data = incumbents_subset)

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 ***
incumbents_subset$difflog 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

# The p-value is very small, so reject the null hypothesis that there is
    no relationship between the difference in spending between the
    incumbent and the challenger and the vote share of the incumbent's
    party. It is likely that the difference in spending is related to the
    vote share of the incumbent's party. The intercept of the regression
    is 0.507583, meaning that when the difference in spending between the
    incumbent and the challenger is 0, the vote share of the incumbent's
    party is 0.507583. The slope of the regression is 0.023837, meaning
    that the vote share of the incumbent's party increases by 0.023837 for
     every one unit that the difference in spending between the incumbent
    and the challenger increases.
```
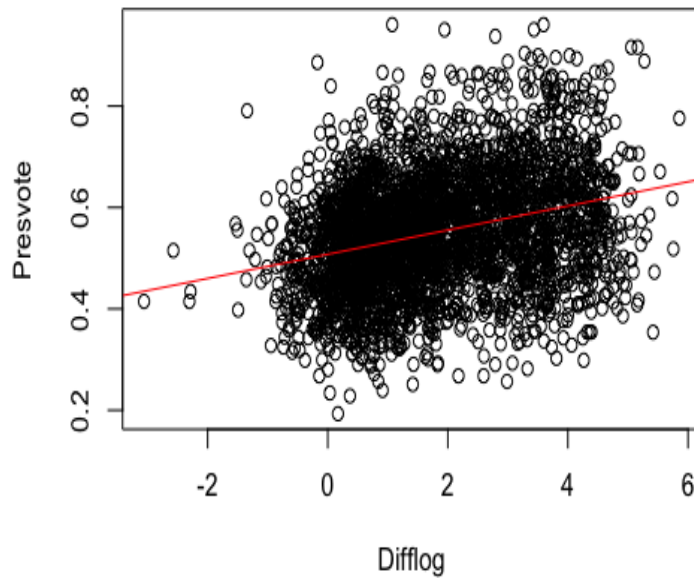
2. Make a scatterplot of the two variables and add the regression line.

```
# Make a scatterplot of the relationship between difflog and preshare
plot(incumbents_subset$difflog,incumbents_subset$voteshare, xlab = "
    Difflog", ylab = "Voteshare")
```

```
3 # Add the regression line
4 abline(difflog_voteshare_regression, col="red")
```


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

```
1 # Create an object for the residuals of the regression
2 residuals_difflog_presvote<- residuals(difflog_presvote_regression)
3 residuals_difflog_presvote
```

4. Write the prediction equation.

```
1 # Y = beta0 + beta1 * x
2 # beta0 = intercept
3 # beta1 = slope
4 # SOLUTION: Y = 0.507583 + .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`.

```
Use the lm() function to run a regression analysis on the relationship
    between voteshare and presvote, with voteshare as the outcome variable
    (y) and presvote as the explanatory variable (x)
presvote_voteshare_regression<- lm(incumbents_subset$voteshare~incumbents
    _subset$presvote, data = incumbents_subset)
presvote_voteshare_regression
summary(presvote_voteshare_regression)

Call:
lm(formula = incumbents_subset$voteshare ~ incumbents_subset$presvote,
data = incumbents_subset)

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 ***
incumbents_subset$presvote 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

The p-value is very small, so reject the null hypothesis that there is
    no relationship between the the vote share of the incumbent's party
    and the incumbent's electoral success. It is likely that the the vote
    share of the incumbent's party is related to the the incumbent's
    electoral success. The intercept of the regression is 0.441330,
    meaning that when the the incumbent's electoral success is 0, the vote
     share of the incumbent's party is 0.441330. The slope of the
    regression is 0.388018, meaning that the vote share of the incumbent's
     party increases by 0.388018 for every one unit that the the the
    incumbent's electoral success increases.
```

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

6

```
1 # Make a scatterplot of the relationship between difflog and presvote
2 plot(incumbents_subset$presvote, incumbents_subset$voteshare, xlab = "
      Presvote", ylab = "Voteshare")
3 # Add the regression line
4 abline(presvote_voteshare_regression, col= "red")
```
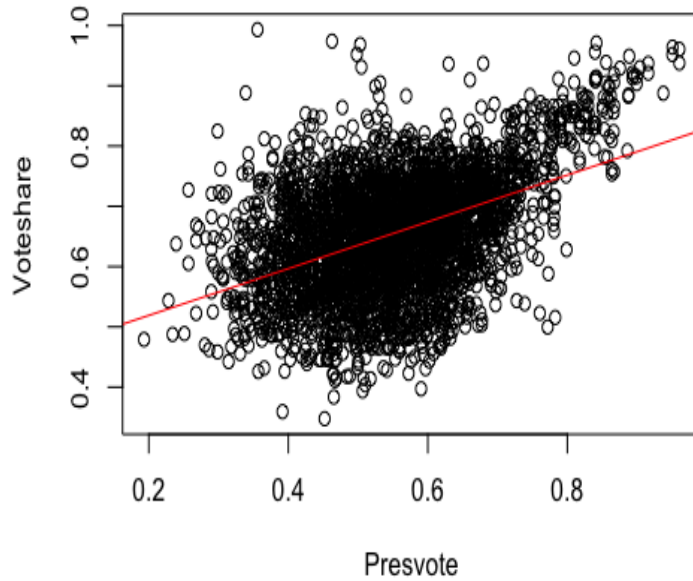

3. Write the prediction equation.

```
1 # Y = beta0 + beta1 * x
2 # beta0 = intercept
3 # beta1 = slope
4 # SOLUTION: 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.

```
# Use the lm() function to run a regression analysis on the relationship
    between the residuals in question 1 and the residuals in question 2,
    with the residuals in question 1 as the outcome variable (y) and the
    residuals in question 2 as the explanatory variable (x)

residual_regression<- lm(residuals_difflog_voteshare~residuals_difflog_
    presvote)
summary(residual_regression)

Call:
lm(formula = residuals_difflog_voteshare ~ residuals_difflog_presvote)

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
residuals_difflog_presvote    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
```

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

```
# Make a scatterplot of the relationship between residuals_difflog_
    voteshare and residuals_difflog_presvote
plot(residuals_difflog_presvote, residuals_difflog_voteshare, xlab = "
    Question 2 Residuals", ylab = "Question 1 Residuals")
# Add the regression line
abline(residual_regression, col= "red")
```
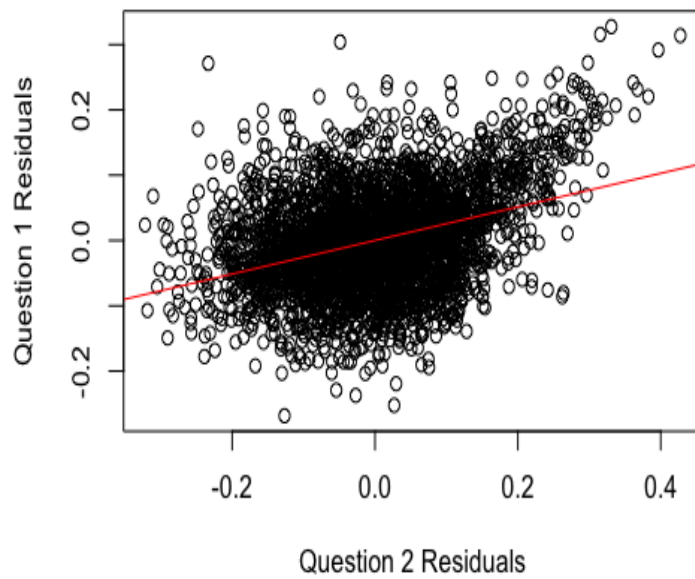
Question 1 Residuals (y-axis)
Question 2 Residuals (x-axis)

3. Write the prediction equation.

```
1 # Y = beta0 + beta1 * x
2 # beta0 = intercept
3 # beta1 = slope
4 # SOLUTION: 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`.

```
> voteshare_difflog_presvote <−lm(incumbents_subset$voteshare˜ incumbents
    _subset$difflog + incumbents_subset$presvote , data = incumbents_subset
    )
> summary(voteshare_difflog_presvote)

Call:
lm(formula = incumbents_subset$voteshare ˜ incumbents_subset$difflog +
incumbents_subset$presvote , data = incumbents_subset)

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 ***
incumbents_subset$difflog   0.0355431   0.0009455    37.59    <2e−16 ***
incumbents_subset$presvote 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.

```
# # Y = beta0 + beta1 * x1 + beta2 * x2
# beta0 = intercept
# beta1 = slope of the relationship between difflog and voteshare
# x1 = difflog
# beta2 = slope of the relationship between presvotes and voteshare
# x2 = voteshare
# SOLUTION: Y = 0.4486442 + 0.0355431 * x1 + 0.2568770 * x2
```

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 analysis in question 4 gives the effect of the incumbent's
    electoral success on the vote share that is not explained by the
    effect of the difference in spending between the incumbent and the
    challenger. By running a multivariat regression analysis that also
```

10

takes into account the difference in spending between the incumbent on the outcome of the incumbent's vote share, the effect of the incumbent's electoral success on the incumbent's vote share is calculated without the effect explained by the difference in spending between the incumbent and the challenger. As a result, the slope of the regression explained by the incumbent's electoral success on the vote share is 0.2568770 for both question 4 and question 5.