

Problem Set 3

Applied Stats/Quant Methods 1

Due: November 19, 2022

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 GitHub.
- This problem set is due before 23:59 on Sunday November 19, 2023. No late assignments will be accepted.

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

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`.

```
Answer # (1)
2
3 #run the regression
4 lm_vote_diff=lm(inc.sub$voteshare~inc.sub$difflog)
5 summary(lm_vote_diff)
6 #According to the regression ,we can see the based on the t value of
  Intercept and difflog ,
7 #the doesn't equal to 0 significantly. And according to the Muultiple R-
  squared=0.3673
8 #we can say difflog only can exaplain around 37% change in voteshare.
```

```

Call:
lm(formula = inc.sub$voteshare ~ inc.sub$difflog)

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 ***
inc.sub$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

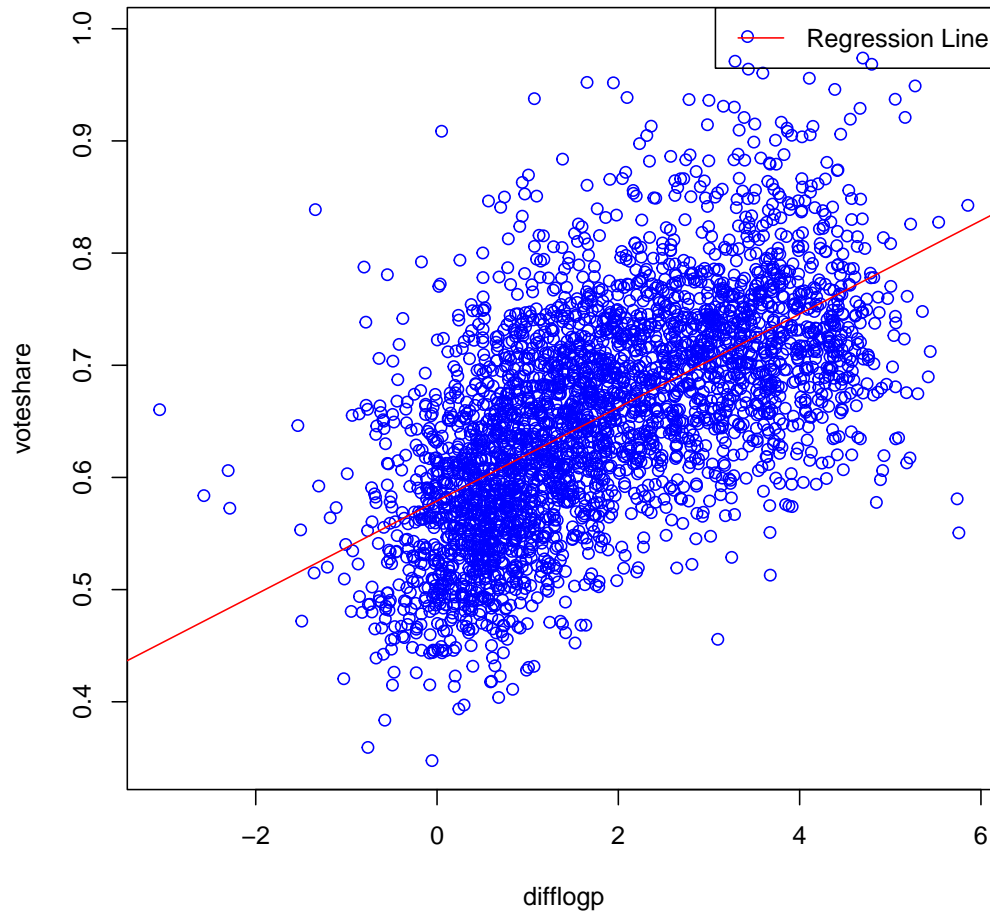
```

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

```

Answer # (2)
1 #name the picture
2 pdf("lm_vote_diff.pdf")
3
4
5 # Create a scatterplot
6 plot(inc.sub$difflog, inc.sub$voteshare, ch = 16, col = "blue", main = "
   Scatterplot with Regression Line", xlab = "difflogp", ylab = "
   voteshare")
7
8 # Add regression line to the plot
9 abline(lm_vote_diff, col = "red")
10
11 # Add a legend
12 legend("topright", legend = "Regression Line", col = "red", lty = 1)
13
14 #save the picture

```

Scatterplot with Regression Line

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

Answer # (3)

```
2 #Save the residuals as a separate object
3 residuals_vote_diff <- residuals(lm_vote_diff)
4 summary(residuals_vote_diff)
```

4. Write the prediction equation.

Answer # (4)

```
2 #the equationb: voteshare = 0.58+0.04*difflog ,
3 #where 1)B0=0.58, which means when difflog equals to 0, the
4 #.      mean of voteshare equals to 0.58
5 #      2)B1=0.04, which means when difflog increase by one unite ,
6 #.      the mean of voteshare would increase by 0.04
```

Question 2

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`.

Answer

```
# Question 2 -----
# (1)
#run the regression
lm_pres_diff=lm(inc.sub$presvote~inc.sub$difflog)
summary(lm_pres_diff)
#According to the regression ,we can see the based on the t value of
  Intercept and difflog ,
#they don't equal to 0 significantly . And according to the Muultiple R-
  squared=0.08795
#we can say difflog only can exaplain around 9% change in presvote.
```

```
Call:
lm(formula = inc.sub$presvote ~ inc.sub$difflog)

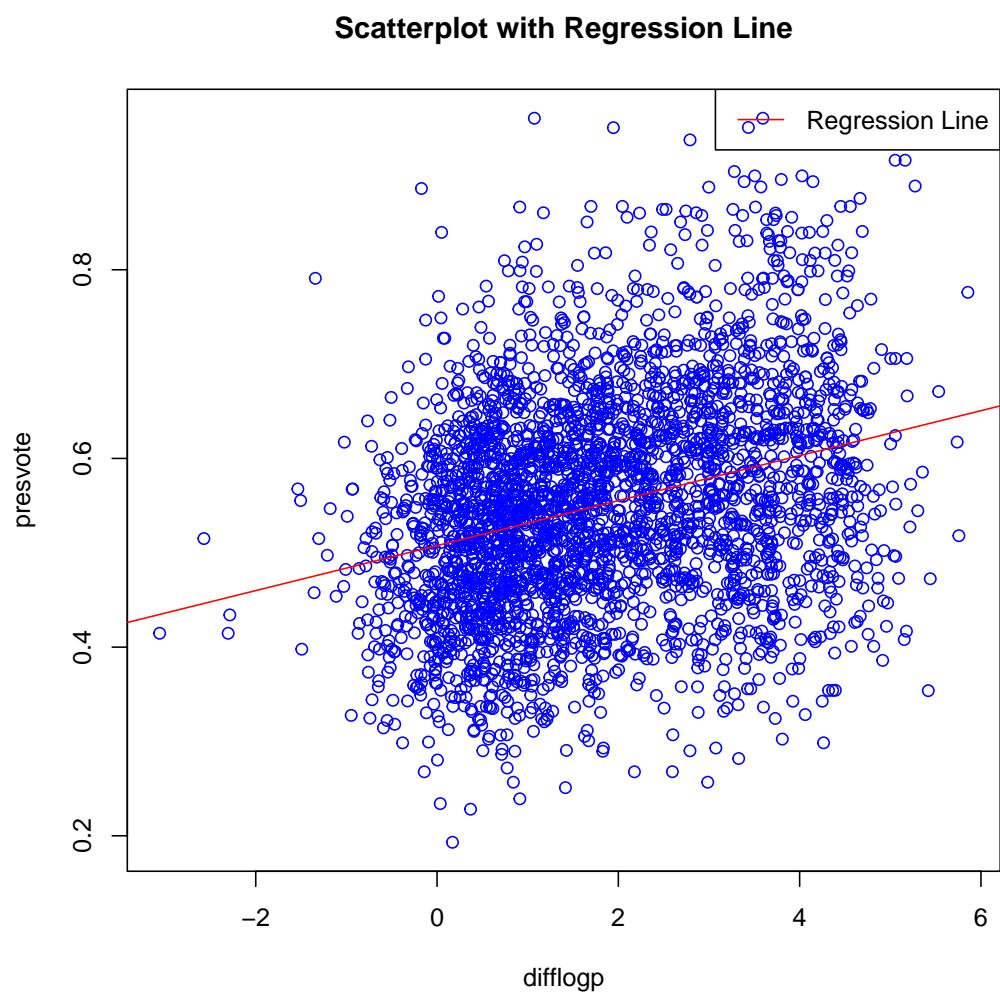
Residuals:
    Min       1Q   Median       3Q      Max
-0.32196 -0.07407 -0.00102  0.07151  0.42743

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.507583   0.003161  160.60  <2e-16 ***
inc.sub$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.

```
Answer # (2)
2 #name the picture
3 pdf("lm_pres_diff.pdf")
4
5 # Create a scatterplot
6 plot(inc.sub$difflog, inc.sub$presvote, ch = 16, col = "blue", main = "
    Scatterplot with Regression Line", xlab = "difflogp", ylab = "presvote
    ")
7
8 # Add regression line to the plot
9 abline(lm_pres_diff, col = "red")
10
11 # Add a legend
12 legend("topright", legend = "Regression Line", col = "red", lty = 1)
13
14 #save the picture
15 dev.off()
```



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

Answer `# (3)`

```
2 #Save the residuals as a separate object
3 residuals_pres_diff <- residuals(lm_pres_diff)
4 summary(residuals_pres_diff)
```

4. Write the prediction equation.

Answer # (4)

```
2 #the equationb: presvote = 0.58+0.04*difflog ,
3 #where 1)B0=0.58, which means when difflog equals to 0, the
4 #.      mean of presvote equals to 0.58
5 #      2)B1=0.04, which means when difflog increase by one unite ,
6 #.      the mean of presvote would increase by 0.04
```


Question 3

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 # (1)
2 #run the regression
3 lm_vote_pres=lm(inc.sub$voteshare~inc.sub$presvote)
4 summary(lm_vote_pres)
5 #According to the regression ,we can see the based on the t value of
  Intercept and difflog ,
6 #they don't equal to 0 significantly. And according to the Muultiple R-
  squared=0.2058
7 #we can say presvote only can exaplain around 20% change in voteshare.
```

```
Call:
lm(formula = inc.sub$voteshare ~ inc.sub$presvote)

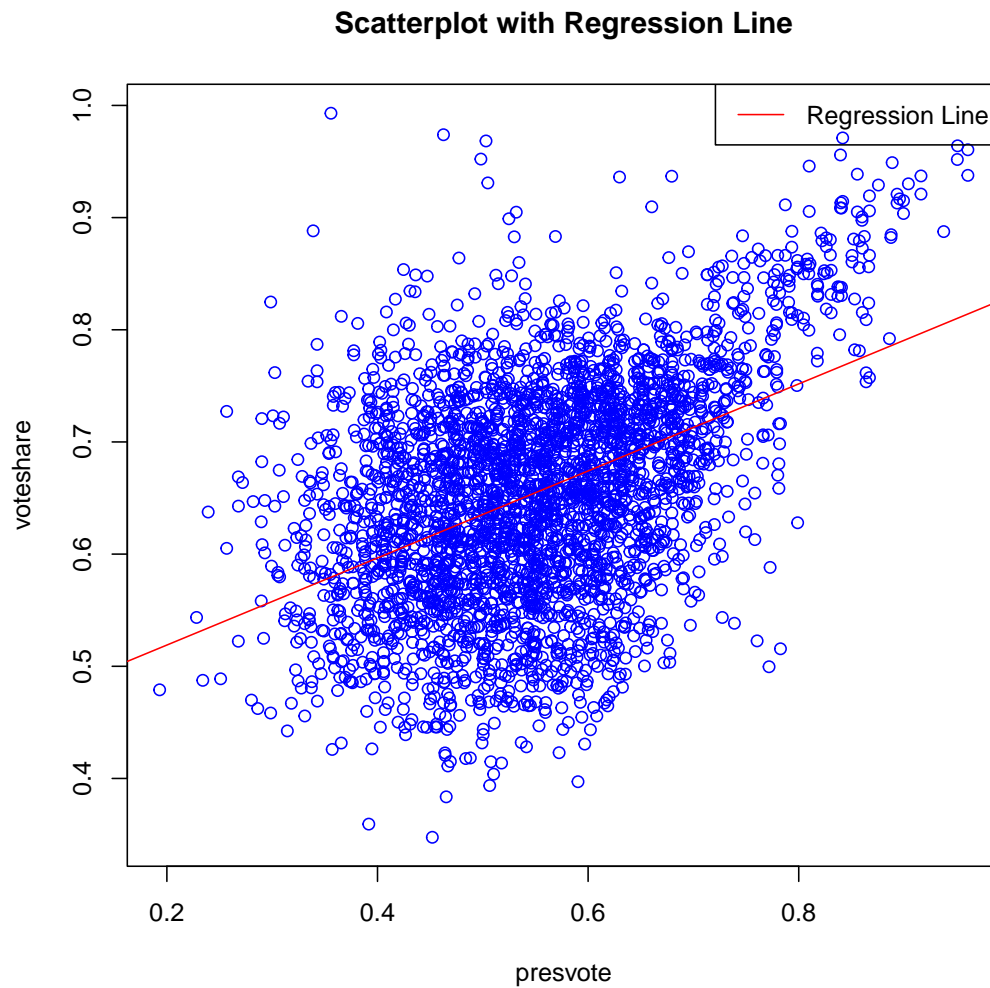
Residuals:
    Min       1Q   Median       3Q      Max
-0.27330 -0.05888  0.00394  0.06148  0.41365

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.441330   0.007599   58.08  <2e-16 ***
inc.sub$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 # (2)
2 #name the picture
3 pdf("lm_vote_pres.pdf")
4
5 # Create a scatterplot
6 plot(inc.sub$presvote, inc.sub$voteshare, ch = 16, col = "blue", main = "
   Scatterplot with Regression Line", xlab = "presvote", ylab = "
   voteshare")
7
8 # Add regression line to the plot
9 abline(lm_vote_pres, col = "red")
10
11 # Add a legend
12 legend("topright", legend = "Regression Line", col = "red", lty = 1)
13
14 #save the picture
15 dev.off()
```



3. Write the prediction equation.

```
1 # (3)
2 #the equationb: voteshare = 0.44+0.38*presvote ,
3 #where 1)B0=0.44, which means when presvote equals to 0, the
4 #.      mean of voteshare equals to 0.44
```

⁵ # 2) $B_1=0.38$, which means when presvote increase by one unite ,
⁶ #. the mean of voteshare would increase by 0.38

Question 4

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 # (1)
2 #run the regression
3 lm_residuals=lm(residuals_vote_diff~residuals_pres_diff)
4 summary(lm_residuals)
5 #According to the regression ,we can see the based on the t value of
   residual_pres_diff
6 #they don't equal to 0 significantly. However, 95% CI of Intercept
   include 0, so
7 # it is plausible that residuals_vote_cliff equals to 0 when residual_
   pres_diff euquals to 0
8 #And according to the Muultiple R-squared=0.13, we can say residual_pres_
   diff
9 #only can exaplain around 13% change in residuals_vote_cliff.
```

```
Call:
lm(formula = residuals_vote_diff ~ residuals_pres_diff)

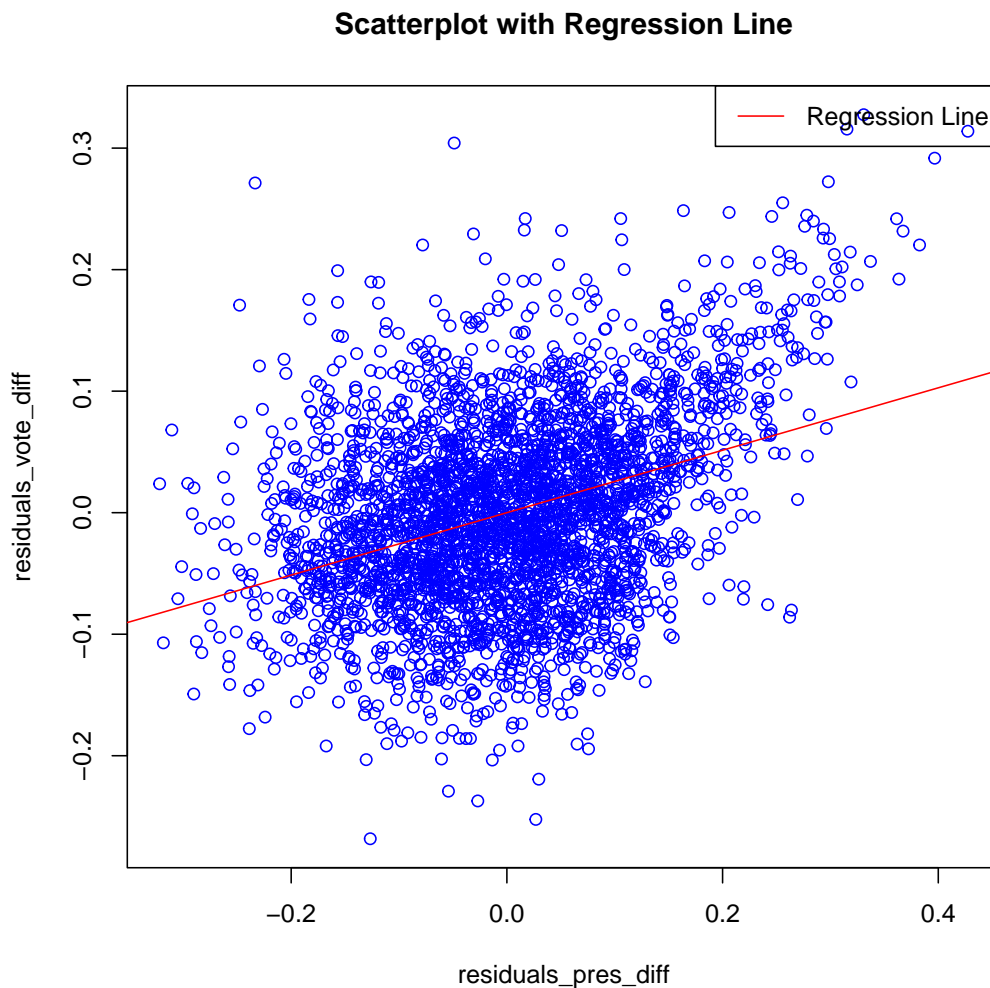
Residuals:
    Min       1Q   Median       3Q      Max
-0.25928 -0.04737 -0.00121  0.04618  0.33126

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.942e-18  1.299e-03   0.00    1
residuals_pres_diff  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.

```
1 # (2)
2 pdf("lm_residuals.pdf")
3
4 # Create a scatterplot
5 plot(residuals_pres_diff, residuals_vote_diff, ch = 16, col = "blue", main
      = "Scatterplot with Regression Line", xlab = "residuals_pres_diff",
      ylab = "residuals_vote_diff")
6
7 # Add regression line to the plot
8 abline(lm_residuals, col = "red")
9
10 # Add a legend
11 legend("topright", legend = "Regression Line", col = "red", lty = 1)
12
13 #save the picture
14 dev.off()
```



3. Write the prediction equation.

```
1 # (3)
2 #the equationb: residuals_vote_diff = (-1.942e-18)+(2.569e-01)*residuals_pres_diff,
3 #where 1)B0=(-1.942e-18), which means when residuals_pres_diff equals to 0, the
4 #.      mean of residuals_vote_diff equals to (-1.942e-18)
5 #      2)B1=(2.569e-01), which means when residual_vote_diff increase by
6 #.      one unite,
7 #.      the mean of residuals_vote_diff would increase by (2.569e-01)
```

Question 5

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 # (1)
2 #run the regression
3 lm_vote_diff_pres=lm(inc.sub$voteshare~inc.sub$difflog+inc.sub$presvote)
4 summary(lm_vote_diff_pres)
5 #According to the regression,we can see the based on the t value of
   Intercept difflog
6 #and presvote,they don't equal to 0 significantly. And according to the
   Muultiple
7 #R-squared=0.4496,we can say difflog and presvote only can explain
   around 45% change in presvote.
```

```
Call:
lm(formula = inc.sub$voteshare ~ inc.sub$difflog + inc.sub$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)    0.4486442  0.0063297   70.88  <2e-16 ***
inc.sub$difflog 0.0355431  0.0009455   37.59  <2e-16 ***
inc.sub$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.

```
1 # (2)
2 #the equationb: voteshare = 0.45+0.04*difflog+0.26*presvote ,
3 #where 1)B0=0.45, which means when difflog and presvote equals to 0, the
4 #.      mean of residuals _vote_diff equals to 0.45
5 #      2)B1=0.04, which means when presvote keep the same and difflog
6 #.      increase
7 #.      by one unite , the mean of rvoteshare would increase by 0.04
8 #.      3)B2=0.26, which means when difflog keep the same and presvote
9 #.      increase
10 #.      by one unite , the mean of voteshare would increase by 0.26
```

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

```
1 # (3)
2 #the residuals and residual standard error in the output of Question 5 is
3 #identical to that of Question 4. But the resason is unknown.
```