

Problem Set 3

Applied Stats/Quant Methods 1

Due: November 11, 2024

Name: Eimhin O'Neill
Student Number: 20332107

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 Monday November 11, 2024. 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

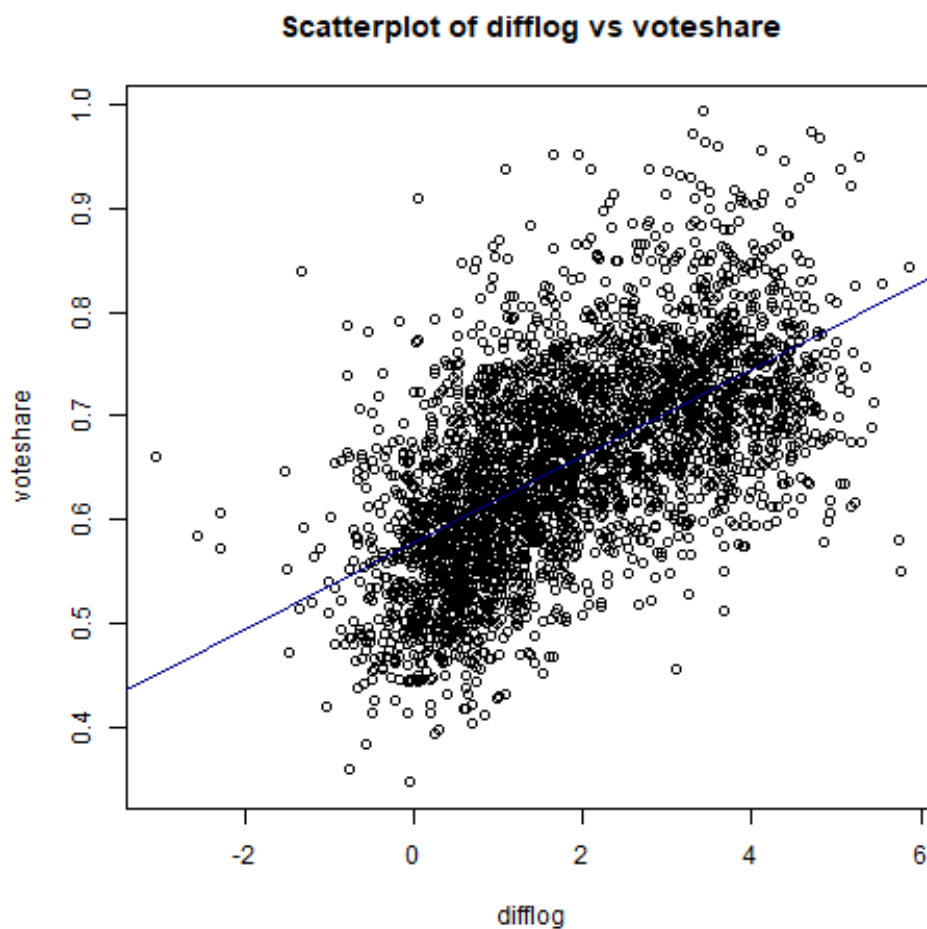
We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

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

```
1 reg_q1_vs_difflog <- lm(voteshare ~ difflog, data = inc.sub)
```

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

```
1 png(file="vs_difflog_plot.png")
2 plot(inc.sub$difflog, inc.sub$voteshare,
3       main="Scatterplot of difflog vs voteshare",
4       xlab="difflog",
5       ylab="voteshare")
6 abline(reg_q1_vs_difflog, col="darkblue")
7 dev.off()
```



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

```
1 resid_q1 <- residuals(reg_q1_vs_difflog)
2 print(resid_q1)
```

4. Write the prediction equation.

```
1 coef(reg_q1_vs_difflog)
2
3 # prediction equation
4 # voteshare = 0.5790 + 0.0417(difflog)
5
6 # so if the incumbent outspends the challenger
7 # their vote share tends to increase
8
9 # But the effect size is modest
10 # a one unit increase in difflog
11 # only raises the vote share by about 4.17%
```

Question 2

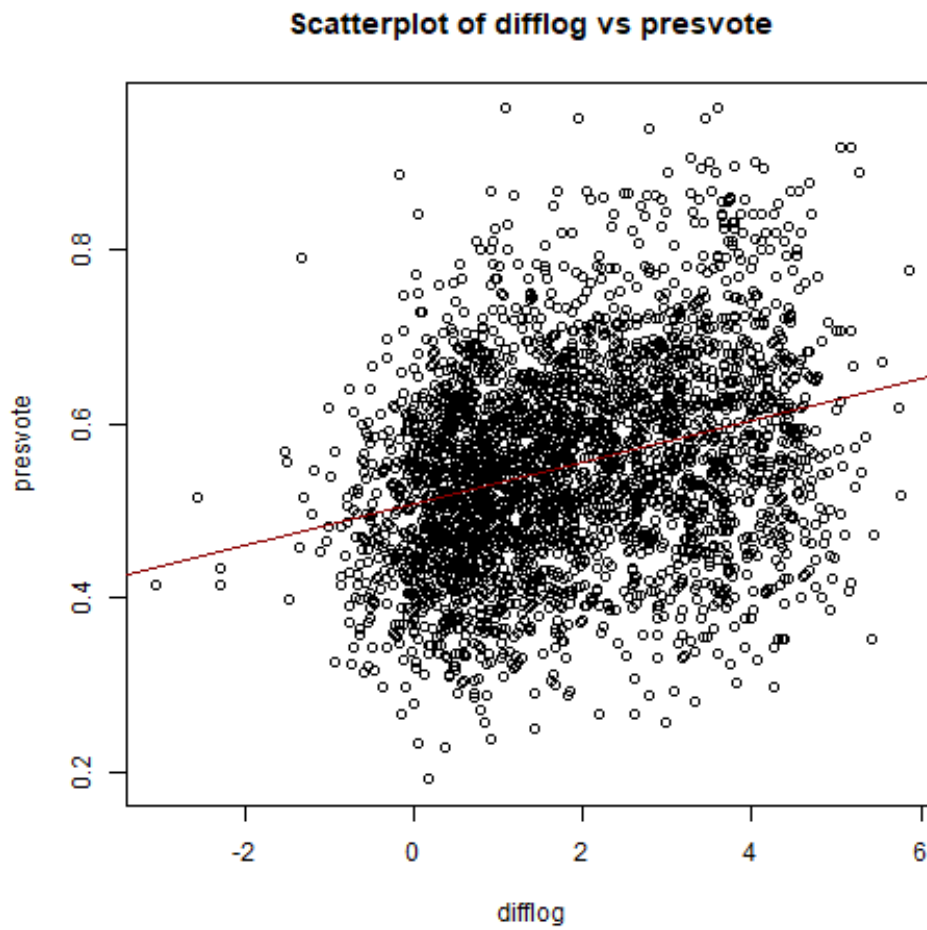
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

1. Run a regression where the outcome variable is **presvote** and the explanatory variable is **difflog**.

```
1 reg_q2_pv_difflog <- lm(presvote ~ difflog , data = inc.sub)
```

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

```
1 reg_q1_vs_difflog <- lm(votesshare ~ difflog , data = inc.sub)
```



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

```
1 resid_q2 <- residuals(reg_q2_pv_difflog)
2 print(resid_q2)
```

4. Write the prediction equation.

```
1 coef(reg_q2_pv_difflog)
2
3 # prediction equation
```

```
4 # presvote = 0.5076 + 0.0238(difflog)
5
6 # as incumbent spends more relative to the challenger
7 # the presidential candidate from incumbents party
8 # is slightly more likely
```

Question 3

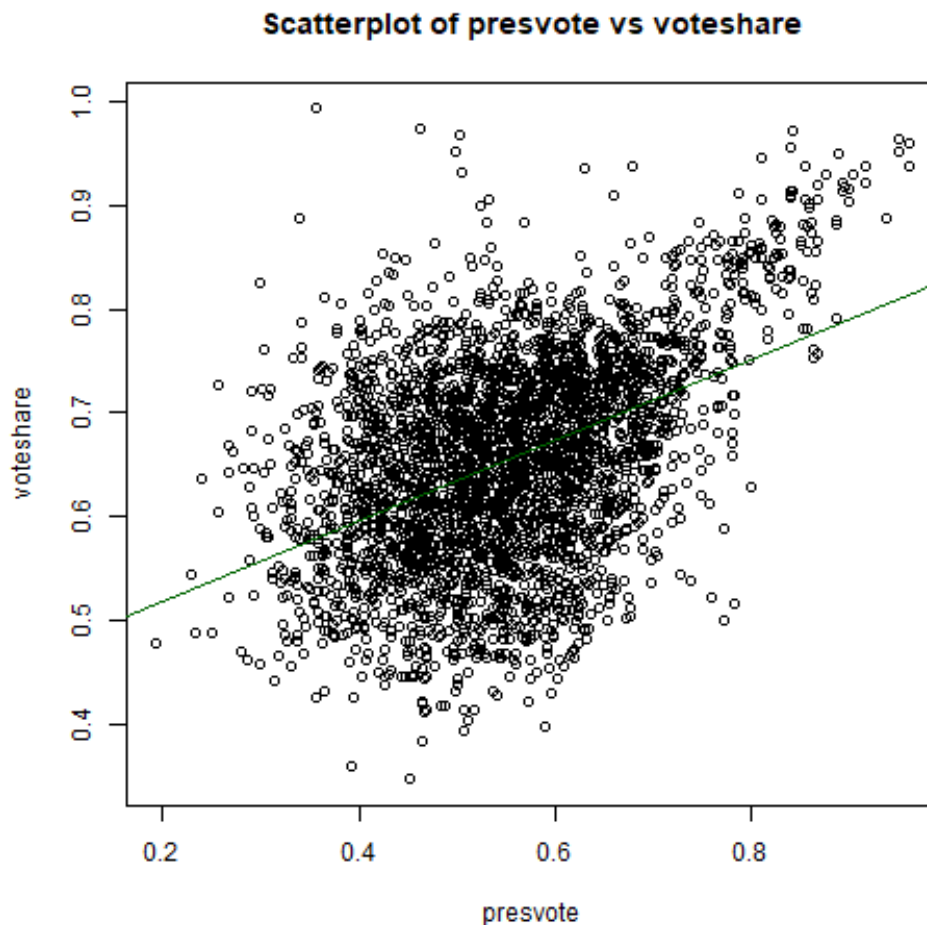
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
1 reg_q3_vs_pv <- lm(voteshare ~ presvote, data = inc.sub)
```

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

```
1 png(file="vs_pv_plot.png")
2 plot(inc.sub$presvote, inc.sub$voteshare,
3       main="Scatterplot of presvote vs voteshare",
4       xlab="presvote",
5       ylab="voteshare")
6 abline(reg_q3_vs_pv, col="darkgreen")
7 dev.off()
```



3. Write the prediction equation.

```

1 coef(reg_q3_vs_pv)
2
3 # prediction equation
4 # voteshare = 0.4413 + 0.388      presvote
5
6 # When the presidential candidate from the incumbent performs better
7 # incumbentt tends to also see a rise in voteshare
8
9 # the 1 % increase in presvote generates a 0.388% increase in the
   incumbent vs
10 # this suggests alignment with a popular pres candidate can boost an inc
   support

```


Question 4

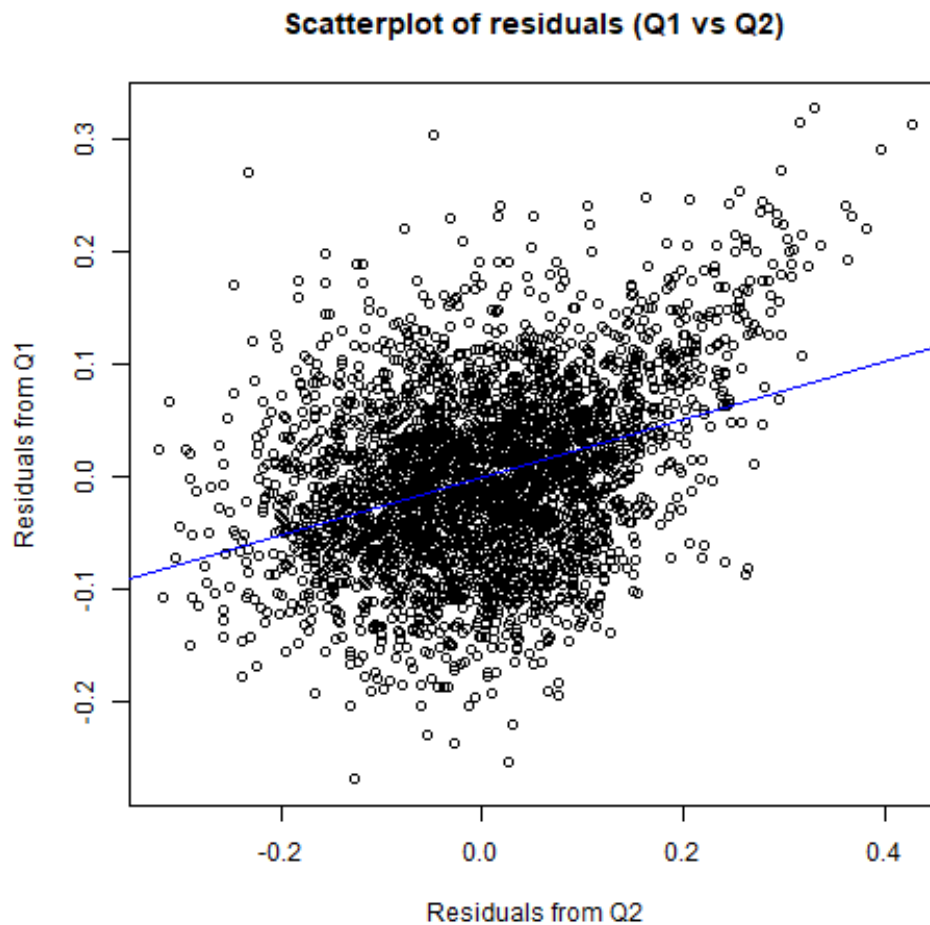
The residuals from part (a) tell us how much of the variation in **voteshare** is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in **presvote** is *not* explained by the difference in spending between incumbent 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 reg_q4_resid12 <- lm(resid_q1 ~ resid_q2)
```

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

```
1 png( file=" resid_plot.png" )
2 plot( resid_q2, resid_q1,
3       main=" Scatterplot of residuals (Q1 vs Q2)",
4       xlab=" Residuals from Q2",
5       ylab=" Residuals from Q1" )
6 abline( reg_q4_resid12, col=" blue" )
7 dev.off()
```



3. Write the prediction equation.

```
1 coef(reg_q4_resid12)
2
3 # prediction equation
4 # resid_q1 = 5 .93 10 * 18 + 0.2569 resid_q2
5
6 # positive slope indicates the unexplained variation in pv is positively
7 # associated with the unexplained variation in vs
8
9 # a 1 unit increase in the residuals of pv g to an enerates
10 # an increase of approx 0.2569 in the residuals of vs
11 # once again, positive association
```

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 reg_q5_vs_dl_pv <- lm(voteshare ~ difflog + presvote, data = inc.sub)
```

2. Write the prediction equation.

```
1 coef(reg_q5_vs_dl_pv)
2
3 # prediction equation
4 # voteshare = 0.4486 + 0.0355 difflog + 0.2569 presvote
5
6 # difflog
7 # as incumbent spends more compared to the challenger
8 # their voteshare slightly increases
9
10 # presvote
11 # larger coeff for presvote indicates that the incumbent vs
12 # more strongly affected by the pres candidates performance
13
14 # potentially showing a rising tide carries all boats phenomenon
15 # where the big race for president could determine lower level races
16 # more locally for example
```

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 slope for `presvote` (0.2569) here is identical to the slope from Question 4's regression of Q.1's residuals on Q.2's residuals (0.2569). This happens because both isolate the relationship between `presvote` and `voteshare`, and control for `difflog`. In this question, it's isolated within the context of a multiple regression with both predictors, but for Q4, it's isolated by analysing the residuals, or more simply, the variation not explained by `difflog`.

Overall, this confirms presvote's independent effect on voteshare, irrespective of difflog.