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`.

   First I create my model using voteshare as the dependent variable and difflog as the indipendent.

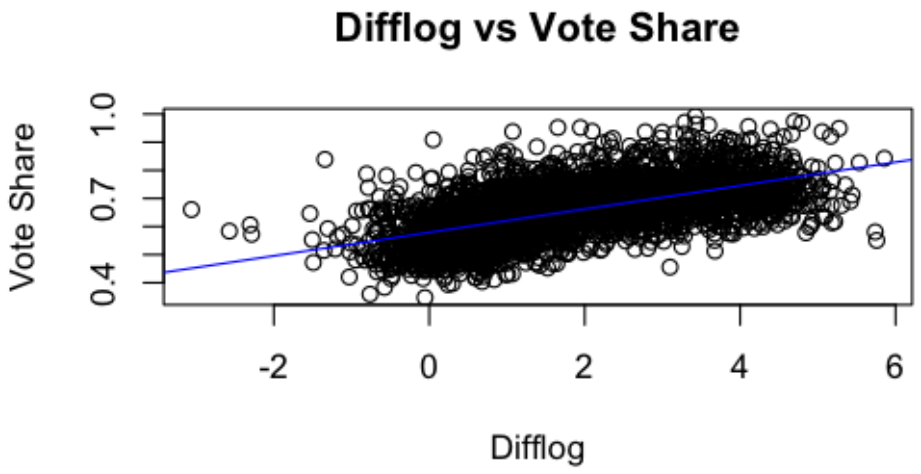
```
model <- lm(voteshare ~ difflog, data = inc.sub)
summary(model)

# Output
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 ***

```

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

   Here I make a plot with a regression line

```
plot(inc.sub$difflog, inc.sub$voteshare,
  main = 'Difflog vs Vote Share',
    xlab = 'Difflog', ylab = 'Vote Share')

abline(model, col = "blue")

```



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

```
resid <- residuals(model)

```

4. Write the prediction equation.

   Predicted voteshare = 0.579031 + 0.041666 * x

   0.57903 is the y intercept and 0.041666 is the expected increase of voteshare for each unit increase in difflog

Question 2

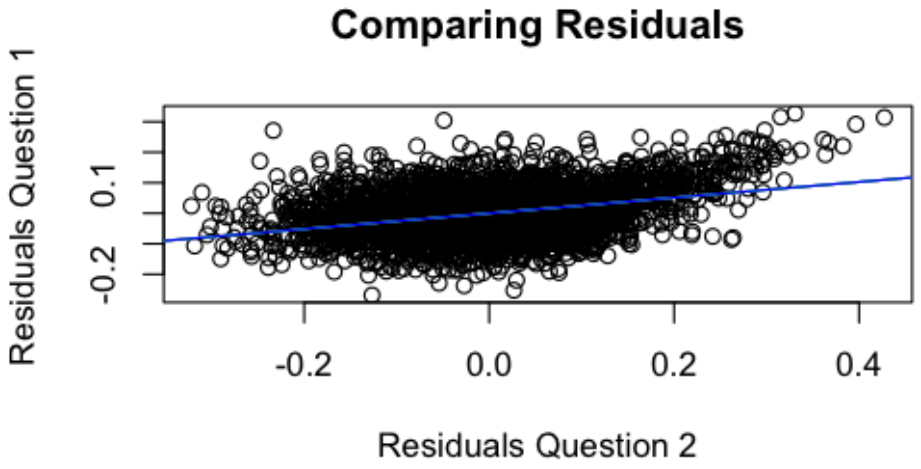
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
model <- lm(voteshare ~ difflog, data = inc.sub)
summary(model)

# Output
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 ***
```

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

```
plot(inc.sub$difflog, inc.sub$presvote,
     main = 'Difflog vs Presvote',
       xlab = "Difflog", ylab = "Presvote" )
abline(model_q2, col = "blue")

```



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

```
      resid_q2 <- residuals(model_q2)

```

4. Write the prediction equation.

   Predicted value of presvote = 0.507583 + 0.023837 * X

   0.507583 is the y intercept and 0.023837 is the expected increase of presvote for each unit increase in difflog

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`.

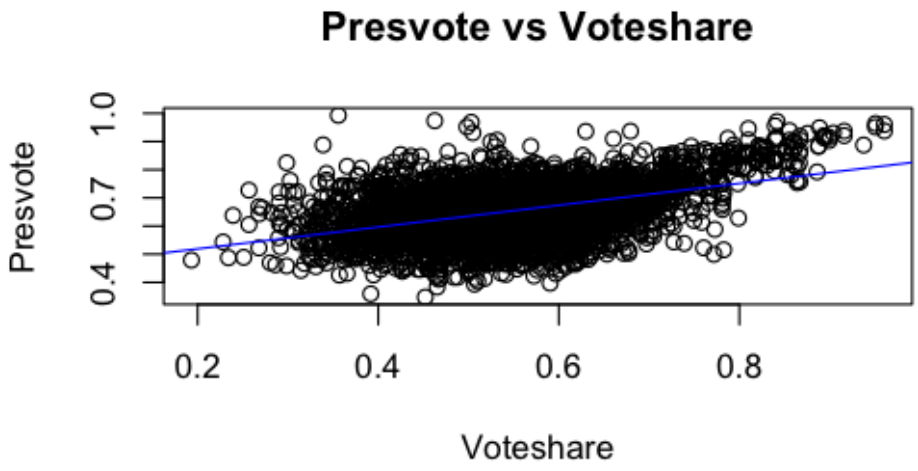
```
model_q3 <- lm(voteshare ~ presvote, data = inc.sub)
summary(model_q3)

# Output
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 ***
```

2. Make a scatterplot of the two variables and add the regression line. ** See next page**

```
1       plot(inc.sub$presvote, inc.sub$voteshare,
2           main = "Presvote vs Voteshare",
3           ylab = 'Presvote', xlab = "Voteshare")
4
5       abline(model_q3, col = "blue")
6
```

3. Write the prediction equation.
   Predicted Voteshare = 0.441330 + 0.388018*x

   0.441330 is the y ntercept and 0.388018 is the expected change of voteshare for each unit increase in presvote.



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.

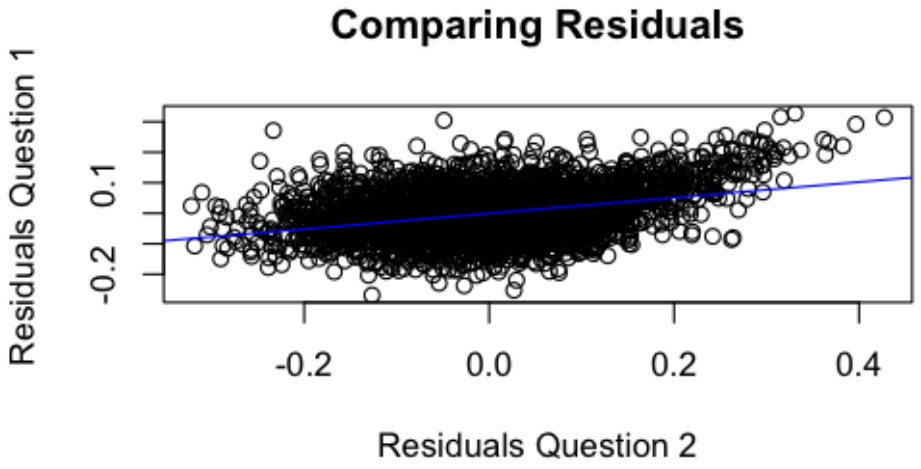
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
model_q4 <- lm(resid {$\sim$} resid_q2)
summary(model_q4)

# Output
Coefficients:
                        Estimate   Std. Error t      value Pr(>|t|)
(Intercept)   -4.860e-18      1.299e-03     0.00         1
resid_q2       2.569e-01       1.176e-02    21.84    <2e-16 ***

```

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

```
plot(resid_q2, resid,       xlab = "Residuals Question 2",
     ylab = "Residuals Question 1",
     main = 'Comparing Residuals')
abline(model_q4, col = "blue")

```

3. Write the prediction equation. I have changed the values out of scientific notation Predicted residual for question 1 model = 0.4860 + 0.2569* x



0.4860 is the Y intercept, 0.2569 is the expected change in residua question 1 for a unt increae in reidual question2.

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`.

```
model_q5 <- lm(voteshare ~ difflog + presvote, data = inc.sub)
summary(model_q5)

```

2. Write the prediction equation.

   Predicted voteshare = 0.4486442 + 0.0355431*x + 0.2568770*x

   0.4486442 is the expected value of voteshare when both indipendent variables are 0
   0.0355431 is the expected increae in voteshare when difflog increaes by one unit
   0.2568770 is the expected increase in voteshare when presvote increases by one unit

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

It appears that the residuals for both the question 5 model and question 4 model are the same.

```
# Question 5 Model Residuals
Residuals:
      Min                1Q        Median              3Q             Max
 -0.25928      -0.04737      -0.00121      0.04618      0.33126

#Question 4 Model Residuals
Residuals:
             Min                1Q        Median              3Q             Max
 -0.25928      -0.04737      -0.00121      0.04618      0.33126

```

Furthermore the Residual standard error: 0.07338 for both models are

However, if I run a test to compare the two, most residuals are not identical. I believe this may be due to rounding.

```
model_q4$residuals == model_q5$residuals

# Here is one line from the output for reference
55           56           57          58           59          60          61
       62            63
FALSE   FALSE    TRUE    TRUE      FALSE    TRUE     FALSE     FALSE      FALSE
```

The model of question 4 and 5 are essential using the same variables
Question 5 obviously compares voteshare $\sim$ difflog + presvote
Question 4 compares the rediduals of questions 1 $\sim$ residuals of question 2.

Lets bread that down.
Redisuals of question 1 result from the model comparing voteshare $\sim$ difflog.
Redisuals of question 1 result from the model comparing voteshare $\sim$ difflog.
As can be seen the variables are repeadeted but processed differently. This could be an explination for having the same residuals.

Another redundency is the slope of presvote in question 5 and the slope of question 4. Both are 0.2568770.