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`.
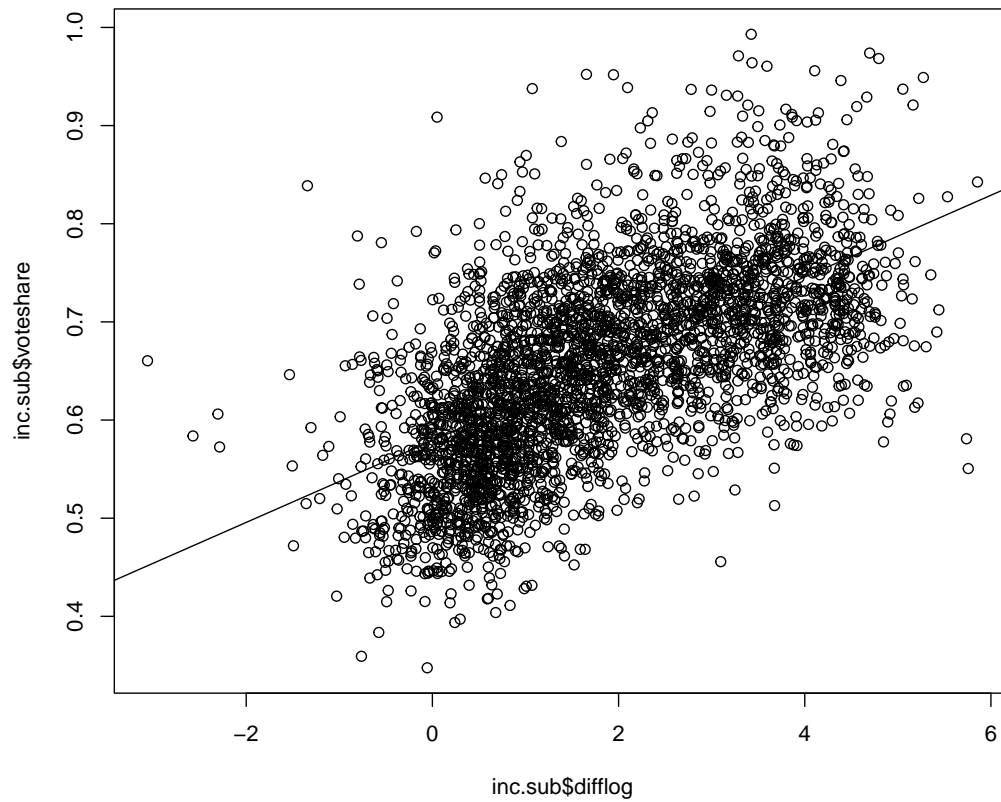
   The R script for regression is:

   ```
   1 model1 <- lm(voteshare~difflog, data= inc.sub)
   2 summary(model1)
   ```

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

   ```
   1 pdf("plot1.pdf", width=8)
   2 plot(inc.sub$difflog, inc.sub$voteshare)
   3 abline(model1)
   4 dev.off()
   ```

Figure 1: Scatterplot of relationship between $voteshare, difflog$.

The above scatter shows bgnuifdkmsas

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

```
1  res1 <- residuals(model1)
```

4. Write the prediction equation.

   voteshare = $B1$ + $B2$ * 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`.
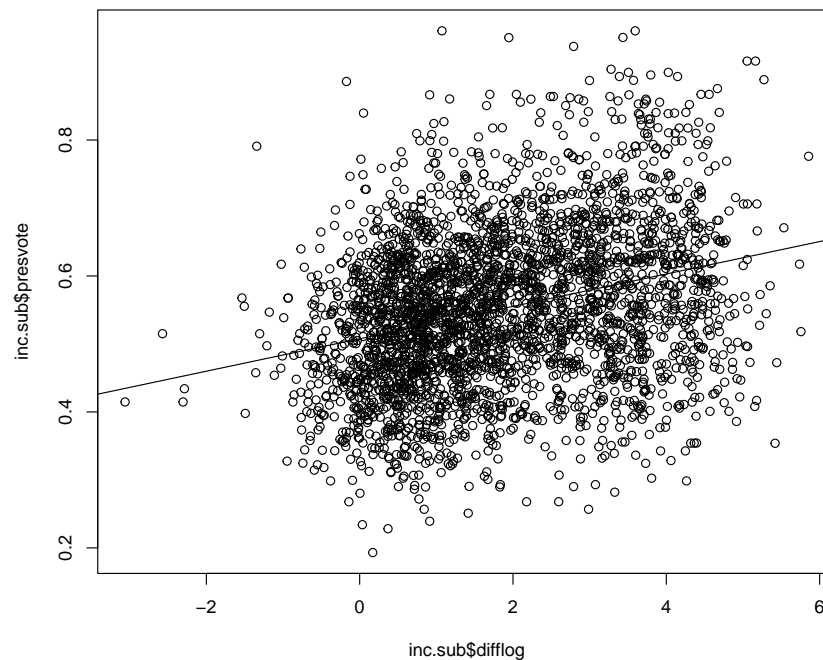
   The R script for regression is:

```
1 model2 <- lm(presvote~difflog, data= inc.sub)
2 summary(model2)
```

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

```
1 pdf("plot2.pdf", width=8)
2 plot(inc.sub$difflog,inc.sub$presvote)
3 abline(model2)
4 dev.off()
```

Figure 2: Scatterplot of relationship between Y, X_2.

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

```
1  res2 <- residuals(model2)
```

4. Write the prediction equation.

presvote = *B1* +*B2* * 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.
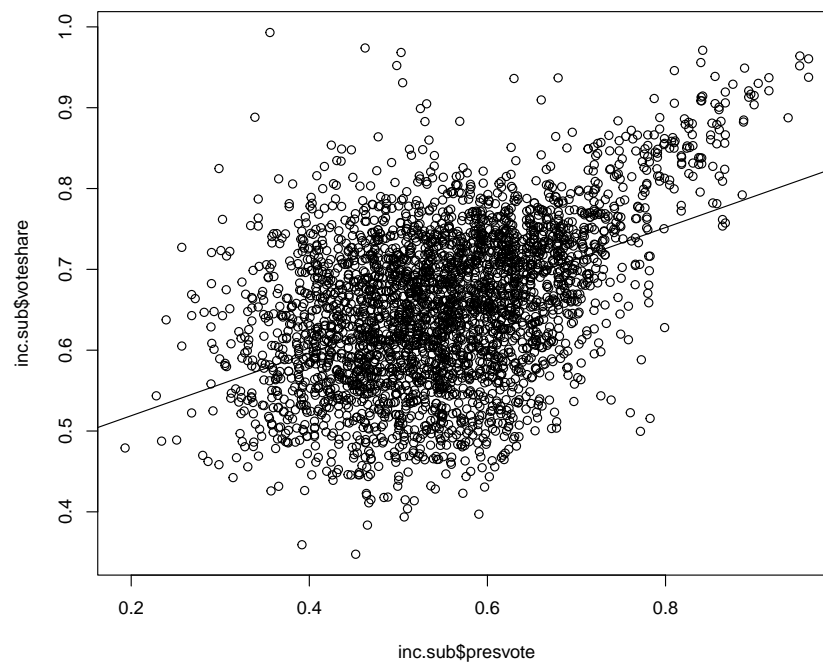
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`. **The R script for regression is:**

```
1 model3 <- lm(voteshare~presvote, data= inc.sub)
2 summary(model3)
```

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

```
1 pdf("plot3.pdf", width=8)
2 plot(inc.sub$presvote, inc.sub$voteshare)
3 abline(model3)
4 dev.off()
```

Figure 3: Scatterplot of relationship between *voteshare, presvote.*

3. Write the prediction equation.

   voteshare = *B1* + *B2* * presvote

5

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.
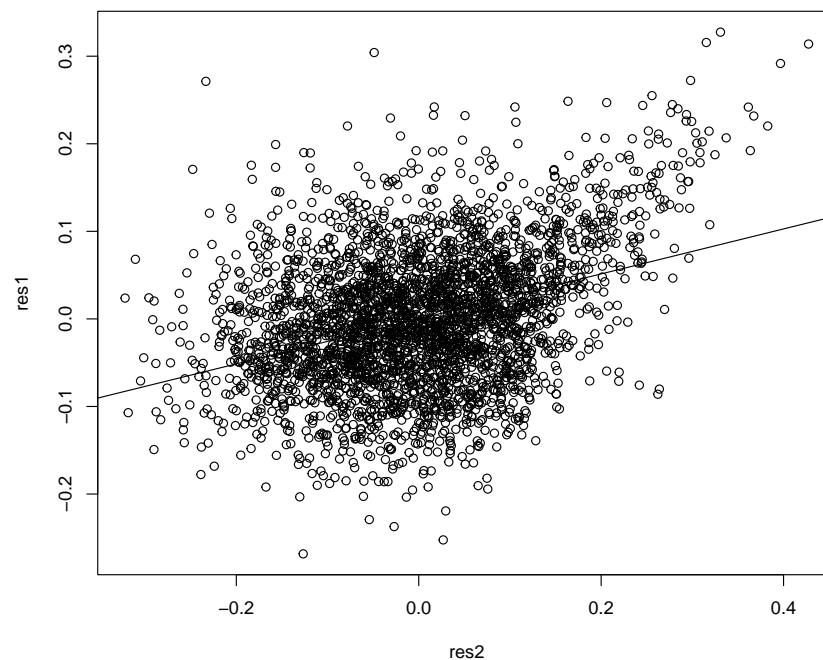
    The R script for regression is:

    ```
    1  model4 <- lm(res1~res2)
    2  summary(model4)
    ```

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

    ```
    1  pdf("plot4.pdf", width=8)
    2  plot(res2, res1)
    3  abline(model4)
    4  dev.off()
    ```

Figure 4: Scatterplot of relationship between $res1, res2$.

3. Write the prediction equation.

    res1 = *B1* + *B2* * res2

6

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`.

   The R script for regression is:

   ```
   1  model5 <- lm(voteshare~difflog+presvote, data= inc.sub)
   2  summary(model5)
   ```

2. Write the prediction equation.

   voteshare = *B1* + *B2* * difflog + *B3* * presvote

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