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)
   ```

   The intercept is 0.579 while the slope of this regression line is 0.041. The R squared value of 0.37 means that around 37 percent of the variability in the voteshare (response variable) can be explained by difflog (explanatory variable). The p-value shows that the difference in campaign spending between incumbent and challenger is a significant predictor of incumbents' vote share.

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 $voteshare, difflog$.
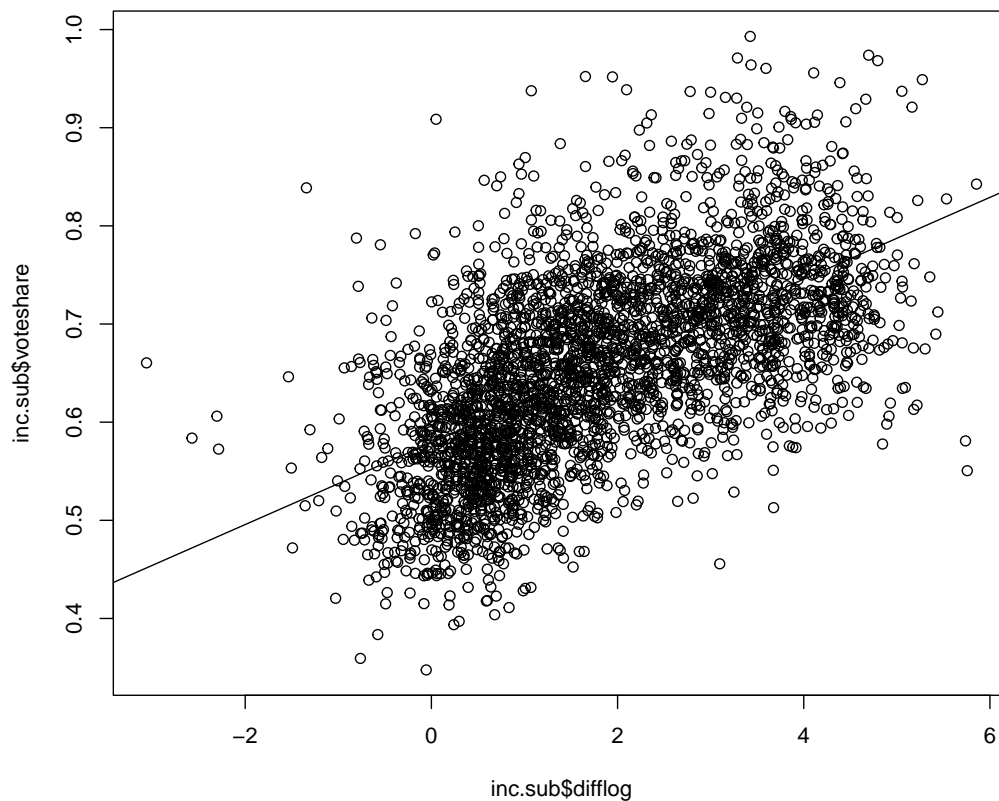
The above scatter plot shows a positive linear relationship between the variables.

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

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

The negative residual value shows that the model overestimates the incumbents' vote share and it is less than predicted by the model.

4. Write the prediction equation.

   voteshare = *B0* + *B1* * difflog

   The intercept is 0.579 while the slope of this regression line is 0.041 which means that with every percentage point increase in the difference of campaign spending between incumbent and challenger, we would expect the percentage increase in voteshare to be increased by 0.041 percent points.

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`.
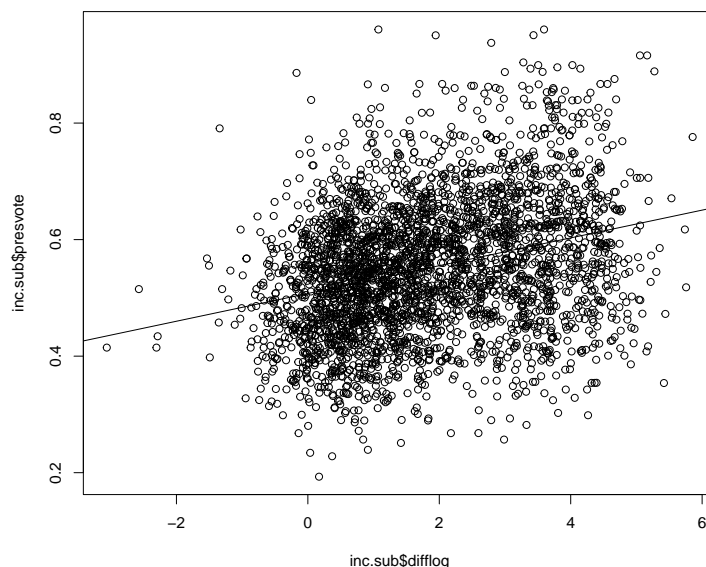
   The R script for regression is:

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

   The intercept is 0.507 while the slope of this regression line is 0.024. The R squared value of 0.087 means that around 9 percent of the variability in the presvote (response variable) can be explained by difflog (explanatory variable). The p-value shows that the difference in campaign spending between incumbent and challenger is a significant predictor of presvote.

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

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

Figure 2: Scatterplot of relationship between Y, X_2.

The above scatter plot shows a positive linear relationship between the variables.

4

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

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

The positive residual value shows that the model underestimates the presvote and it is more than predicted by the model.

4. Write the prediction equation.

presvote = *B0* + *B1* * difflog

The intercept is 0.507 while the slope of this regression line is 0.024 which means that with every percentage point increase in the difference of campaign spending between incumbent and challenger, we would expect the percentage increase in presvote to be increased by 0.024 percent points.

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)
```

The intercept is 0.441 while the slope of this regression line is 0.388. The R squared value of 0.20 means that around 20 percent of the variability in the voteshare (response variable) can be explained by presvote (explanatory variable). The p-value shows that the difference in campaign spending between incumbent and challenger is a significant predictor of incumbents' vote share.

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*.
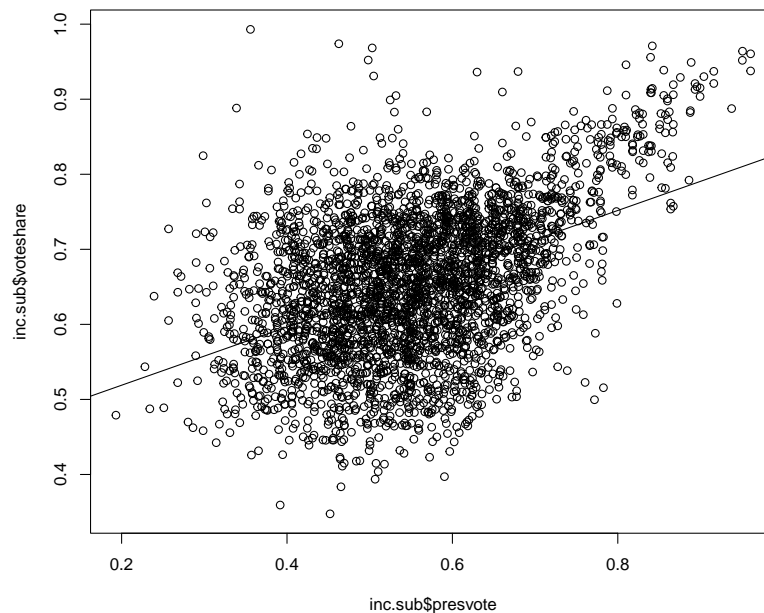
The above scatter plot shows a positive linear relationship between the variables.

3. Write the prediction equation.

voteshare = *B0* + *B1* * presvote

The intercept is 0.441 while the slope of this regression line is 0.388 which means that with every percentage point increase in the presvote, we would expect the percentage increase in voteshare to be increased by 0.388 percent points.

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:**

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

   Both the intercept and slope are having very smaller values. The multiple R squared value of 0.13 means that around 13 percent of the variability in the 1st residual (response variable) can be explained by 2nd residual (explanatory variable). The p-value shows that residual 2 is a significant predictor of residual 1.

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

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

Figure 4: Scatterplot of relationship between $res1, res2$.
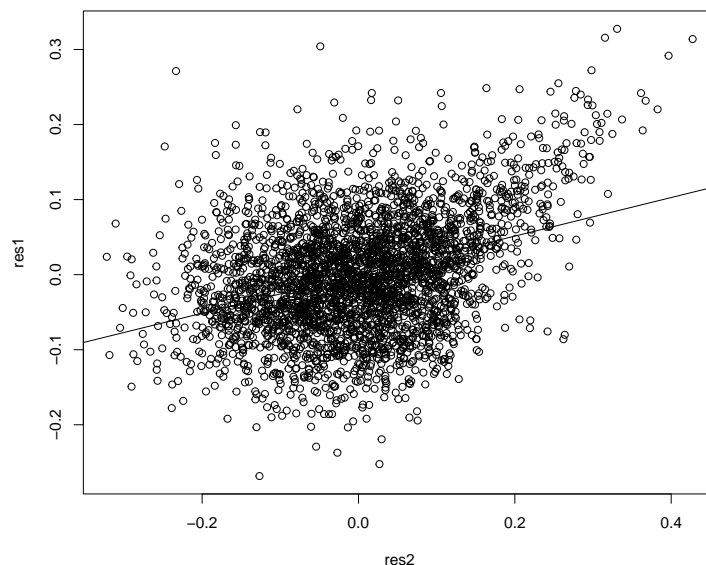
The above scatter plot shows a positive linear relationship between the variables.

3. Write the prediction equation.

   res1 = *B0* + *B1* * res2

   The intercept and slope of this regression line is very smaller which means that with every percentage point increase in the residual2, we would expect a tiny percentage increase in residual1.

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)
   ```

   The intercept is 0.448 while the slopes are 0.0355 and 0.256 respectively. The R squared value of 0.45 means that around 45 percent of the variability in the voteshare (response variable) can be explained by presvote and difference in campaign spending between incumbents and challenger (explanatory variables). The p-value shows that the difference in campaign spending between incumbent and challenger and presvote are significant predictors of incumbents' vote share.

2. Write the prediction equation.

   voteshare = *B0* + *B1* * difflog + *B2* * presvote

   The intercept is 0.448 while the slope of 0.0355 and 0.256 means that with every percentage point increase in the difference in campaign spending between the incumbent and challenger, we would expect the percentage increase in voteshare to be increased by 0.0355 percent points while with every percentage point increase in the presvote, we would expect the percentage increase in voteshare to be increased by 0.256 percentage points.

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

   In both Question 4 and Question 5, although the p-value indicates that there is a significant relationship between the explanatory and response variables
   , the R squared value and multiple R squared values are low which indicates that there are other variables that may have a significant impact on the incumbents' vote share.