Problem Set 3

Linette Lim

Applied Stats/Quant Methods 1

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 20, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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, we make sure the global environment is clear and tidyverse is

loaded. Then, we enter the data.

```
data <- read.csv ("incumbents_subset.csv")
```

Then, we use the lm() function to fit a regression model where the outcome variable is voteshare and the explanatory variable is difflog.

```
lm(voteshare ~ difflog, data = data)
diffvote.lm <- lm(voteshare ~ difflog, data = data)
summary(diffvote.lm)
```

We get:

```
> summary(diffvote.lm)

Call:
lm(formula = voteshare ~ difflog, data = data)

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 ***
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
```

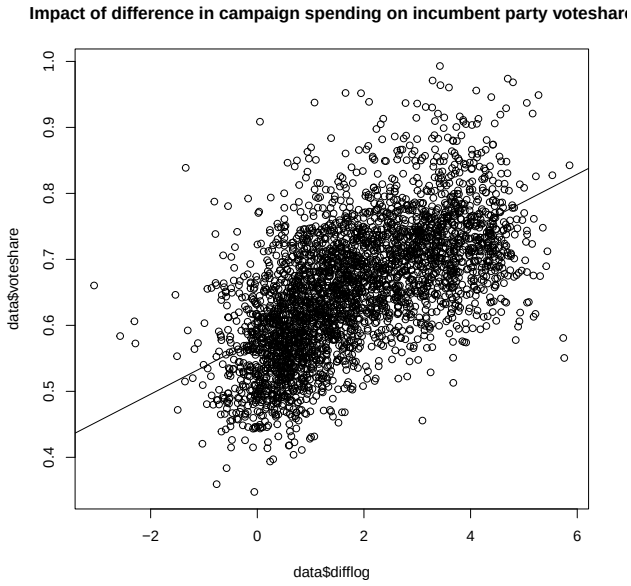
2. Make a scatterplot of the two variables and add the regression line. We use the plot()

function to make a scatterplot and abline () to add the regression line.

```
pdf("plot1_difflog_voteshare.pdf")
plot(data$difflog, data$voteshare, main
     = "Impact of difference in campaign spending on incumbent party
   voteshare")
abline(diffvote.lm)
dev.off()
```

We get:

Figure 1: Impact of Difference in Campaign Spend on Incumbent Party Voteshare.



3. Save the residuals of the model in a separate object. We use the resid() function to

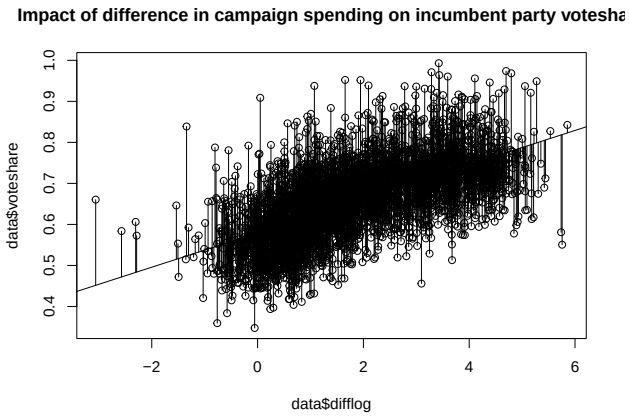
calculate the residuals and then save as separate object.

```
diffvote.res <- resid(diffvote.lm)
```

Let us plot the residuals using the predict() and segments() functions to check that the assumptions of the model have been satisfied.

```
preds.diffvote <- predict(diffvote.lm)
segments(data$difflog, data$voteshare, data$difflog, preds.diffvote)
```

Figure 2: Residuals and regression line of Plot 1.



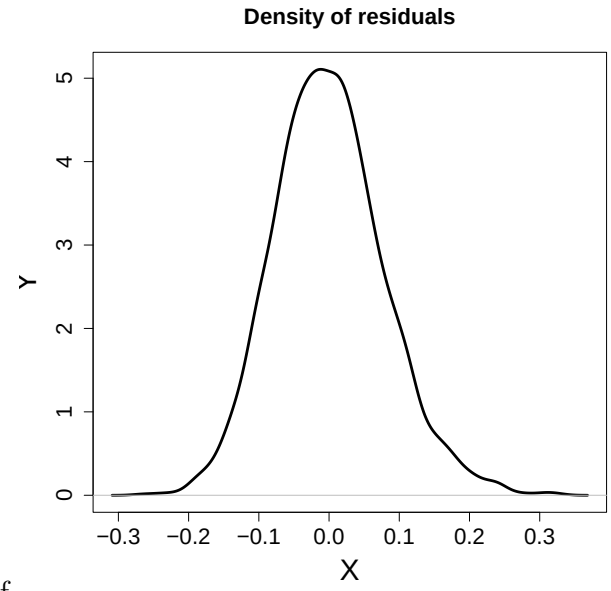
and regression line.pdf

We want to check that the residuals are zero in expectation (Figure 3).

```
pdf("plot3_density of residuals.pdf")
plot(density(data$voteshare-preds.diffvote),
     main = "Density of residuals",
     ylab = "Y", xlab = "X",
     cex.axis=1.5, cex.lab=2,
     cex.main=1.5, lwd=3)
dev.off()
```

We would also expect the residuals to be randomly scattered without showing any systematic patterns when plotted against the predictor variable (Figure 4):

Figure 3: Density of residuals.

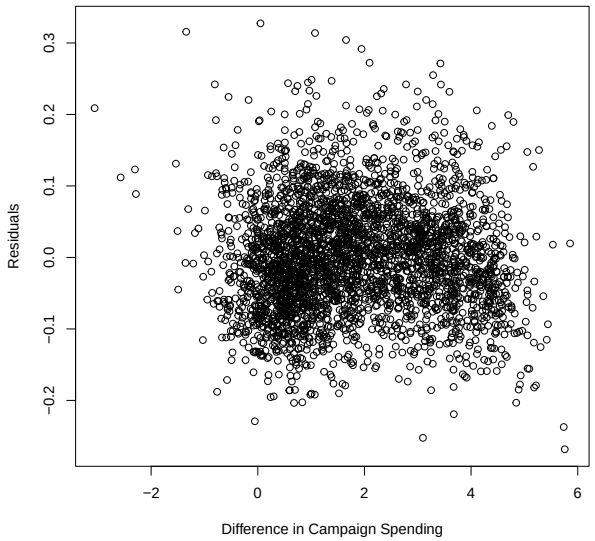


of residuals.pdf

```
pdf("plot4_residual against predictor.pdf")
plot(data$difflog, diffvote.res,
     ylab = "Residuals", xlab= "Difference in Campaign Spending")
dev.off()
```

4. Write the prediction equation. Using Y = b0 + b1X1 and extracting the coefficients from diffvote.lm in Q1.1, we have: Incumbent party voteshare = 0.58 + 0.04*Difference in Campaign Spending, or Y = 0.58 + 0.04X1

Figure 4: Residual against predictor plot.



against predictor.pdf

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`.
   We use lm() to fit a regression model where the outcome variable is presvote and the explanatory variable is difflog.

```
lm(presvote ~ difflog , data = data)
diffpres.lm <- lm(presvote ~ difflog , data = data)
summary(diffpres.lm)
```

We get:

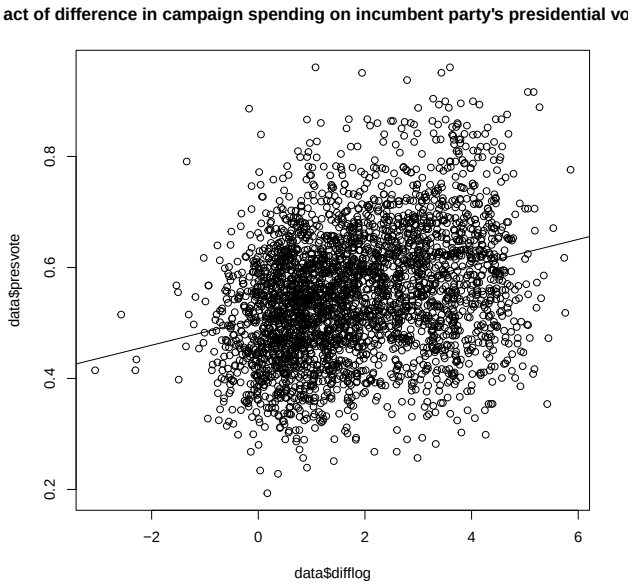
```
> summary(diffpres.lm)

Call:
lm(formula = presvote ~ difflog, data = data)
```

```
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 ***
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.
   We use the plot() to make a scatterplot and abline () to add the regression line. (Figure 5)

```
pdf("plot5_difflog_presvote.pdf")
plot(data$difflog, data$presvote, main
     = "Impact of difference in campaign spending on incumbent party's
   presidential voteshare")
abline(diffpres.lm)
dev.off()
```

Figure 5: Impact of Difference in Campaign Spend on Incumbent Party's Presidential Voteshare.



3. Save the residuals of the model in a separate object.
   We use the resid() function to calculate the residuals and then save as separate object.

```
diffpres.res <- resid(diffpres.lm)
```

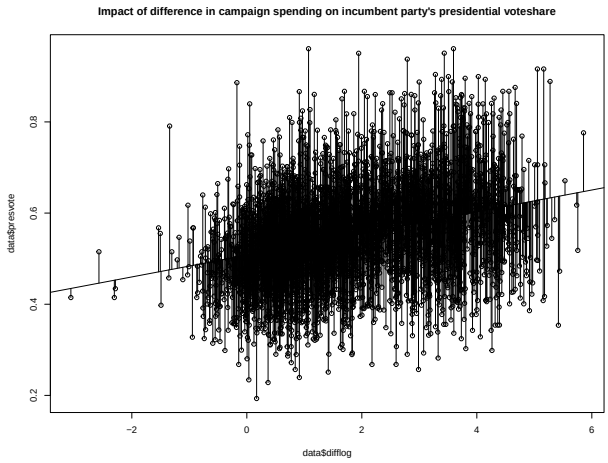
Let us plot the residuals using the predict() and segments() functions to check that the assumptions of the model have been satisfied (Figure 6).

```
preds.diffpres <- predict(diffpres.lm)
segments(data$difflog, data$presvote, data$difflog, preds.diffpres)
```

We want to check that the residuals are zero in expectation (Figure 7).

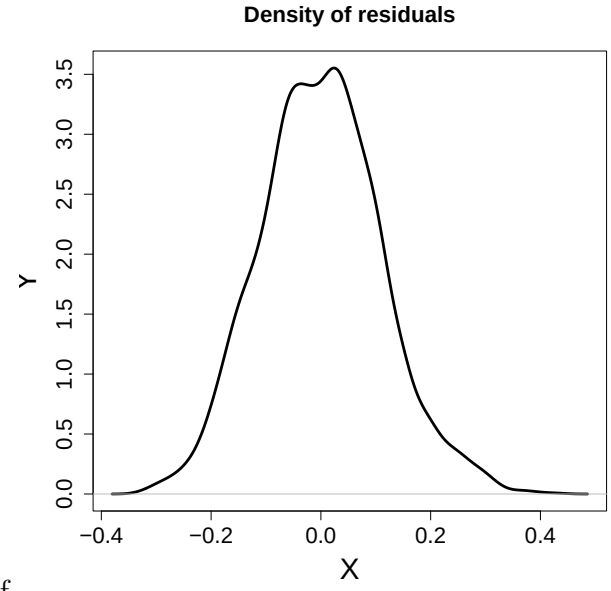
```
pdf("plot7_density of residuals.pdf")
plot(density(data$presvote-preds.diffpres),
     main = "Density of residuals",
     ylab = "Y", xlab = "X",
     cex.axis=1.5, cex.lab=2,
     cex.main=1.5, lwd=3)
dev.off()
```

Figure 6: Residuals and regression line of Plot 5.



and regression line.pdf

Figure 7: Density of residuals.

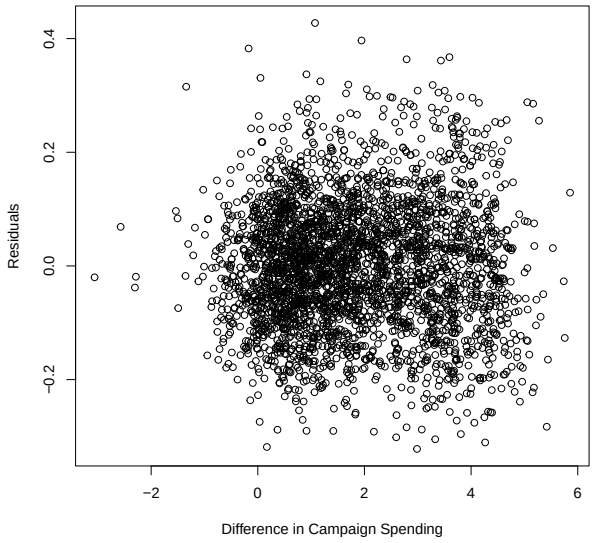


of residuals.pdf

We would also expect the residuals to be randomly scattered without showing any systematic patterns when plotted against the predictor variable (Figure 8):

```
pdf("plot8_residual against predictor.pdf")
plot(data$difflog, diffpres.res,
     ylab = "Residuals", xlab= "Difference in Campaign Spending")
dev.off()
```

Figure 8: Residual against predictor plot.



against predictor.pdf

4. Write the prediction equation. Using Y = b0 + b1X1 and extracting the coefficients from diffpres.lm in Q2.1, we have: Incumbent party's presidential voteshare = 0.51 + 0.02*Difference in Campaign Spending, or Y = 0.51 + 0.02X1

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`.
   We use lm() to fit a regression model where the outcome variable is voteshare and the explanatory variable is presvote.

```r
lm(voteshare ~ presvote, data = data)
presvoteshare.lm <- lm(voteshare ~ presvote, data = data)
summary(presvoteshare.lm)
```

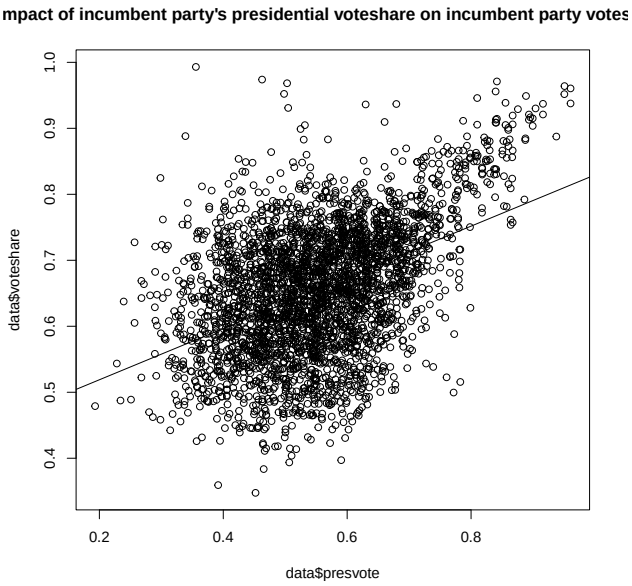
We get:

```
> summary(presvoteshare.lm)

Call:
lm(formula = voteshare ~ presvote, data = data)

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 ***
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.

```
pdf("plot9_presvote_voteshare.pdf")
plot(data$presvote, data$voteshare, main
     = "Impact of incumbent party's presidential voteshare on incumbent
   party voteshare")
abline(presvoteshare.lm)
dev.off()
```

Figure 9: Impact of presvote on voteshare.



3. Write the prediction equation. Using Y = b0 + b1X1 and extracting the coefficients from presvoteshare.lm in Q3.1, we have: Incumbent party voteshare = 0.44 + 0.39*incumbent party's presidential voteshare, or Y = 0.44 + 0.39X

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.

```
lm(diffvote.res ~ diffpres.res, data = data)
residuals.lm <- lm(diffvote.res ~ diffpres.res, data = data)
summary(residuals.lm)
```

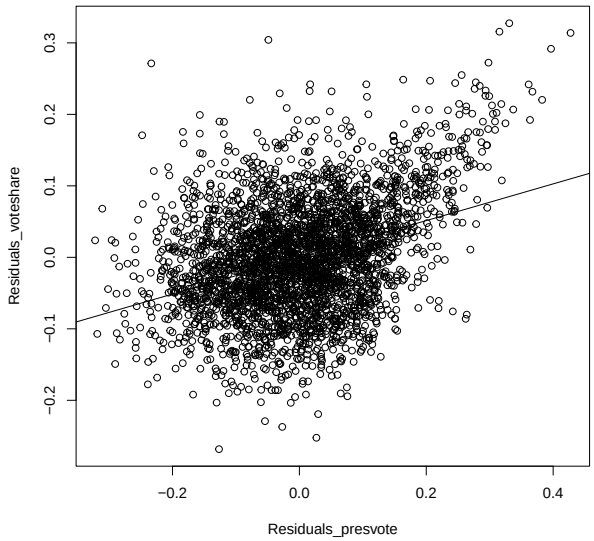
We get:

```
> summary(residuals.lm)

Call:
lm(formula = diffvote.res ~ diffpres.res, data = data)

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)  -4.860e-18  1.299e-03    0.00        1
diffpres.res  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. We use plot() and abline () to make a scatterplot and add the regression line.

```
pdf("plotres_diffpres_diffvote.pdf")
plot(diffpres.res, diffvote.res,
     ylab = "Residuals_voteshare", xlab= "Residuals_presvote")
abline(residuals.lm)
dev.off()
```

Figure 10: Residuals plot.



3. Write the prediction equation. Retrieving the coefficients from Q4.1, we get: Residuals of incumbent party voteshare = -4.860e-18 + 2.569e-01*residuals of incumbent party's presidential voteshare, or Y = -4.860e-18 + 2.569e-01X.
   After rounding the coefficients:

```
> round(-4.860e-18, digits = 4)
[1] 0
> round(2.569e-01, digits = 4)
[1] 0.2569
```

   We get: Y = 0 + 0.2569X1 = 0.2569X1

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`.

```
lm(voteshare ~ difflog + presvote, data = data)
reg2 <- lm(voteshare ~ difflog + presvote, data = data)
summary(reg2)
```

We get:

```
> summary(reg2)

Call:
lm(formula = voteshare ~ difflog + presvote, data = data)

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 ***
difflog     0.0355431  0.0009455   37.59   <2e-16 ***
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.
   The prediction equation is Y = b0 + b1X1 + b2X2. In other words, it is: Incumbent party voteshare = 0.45 + 0.04*difference in campaign spending + 0.26*incumbent party's presidential voteshare, or Y = 0.45 + 0.04X1 + 0.26X2

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

Looking at the regression output for Q5, we see that the coefficient estimate for presvote (0.2568770). This means that for a 1 unit increase in incumbent party's presidential voteshare, we have a 0.26 unit increase in incumbent pary voteshare. Comparing it with the regression output for Q4, we see that the coefficient estimate for diffpres.res (0.26) is identical to the coefficent estimate for presvote in Q5. Recall that in Q4, we regressed the residuals from voteshare difflog against the residuals from presvote difflog. By doing so, we are controlling for difflog (the difference in campaign spend between incumbent and challenger). In other words, this comparison affirms the impact of presvote on voteshare as 0.26.