

# Problem Set 3

## Applied Stats/Quant Methods 1

Due: November 20, 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 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`.

```
1 mod1 <- lm(difflog ~ voteshare , data = dat)
2
3 summary(mod1)
```

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

```
1 splot1 <-  
2   dat %>%  
3     ggplot(aes(x=difflog ,y=voteshare))+  
4     geom_point()+  
5     geom_smooth(method='lm', formula= y~x)  
6  
7 plot(splot1)
```

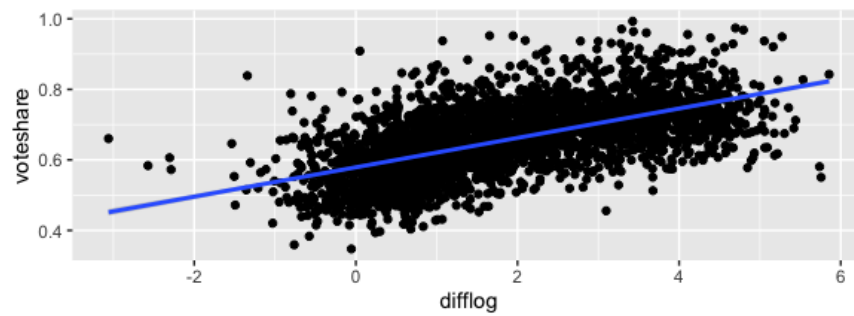


Figure 1:

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

```
1 resid1 <- mod1$residuals
```

4. Write the prediction equation.

```
1 Yi = difflog + voteshare
```

## 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 mod2 <- lm(difflog ~ presvote , data = dat)
2
3 summary(mod2)
```

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

```
1 splot2 <-
2   dat %>%
3   ggplot(aes(x=difflog ,y=presvote))+
4   geom_point()+
5   geom_smooth(method='lm' , formula= y~x)
```

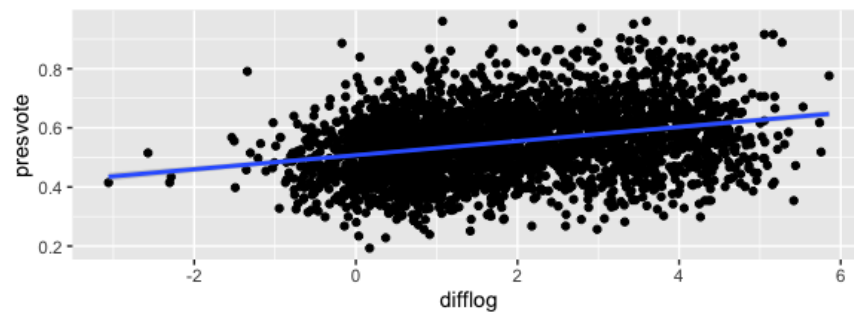


Figure 2:

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

```
1 resid2 <- mod2$residuals
```

4. Write the prediction equation.

```
1 Yi = Voteshare + 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**.

```
1 mod3 <- lm( presvote ~ voteshare , data = dat )  
2  
3 summary(mod3)
```

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

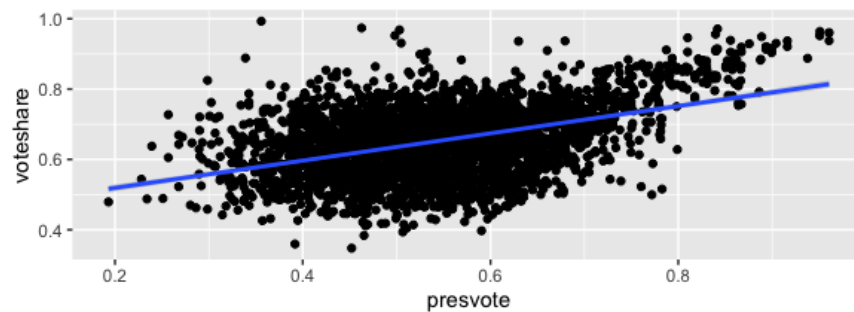


Figure 3:

```
1 splot3 <-  
2   dat %>%  
3   ggplot(aes(x=presvote , y=voteshare )) +  
4   geom_point() +  
5   geom_smooth(method='lm' , formula= y~x)
```

3. Write the prediction equation.

```
1 resid3 <- mod3$residuals
```

## 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 mod4 <- lm(resid1 ~ resid2)
2 summary(mod4)
3 )
```

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

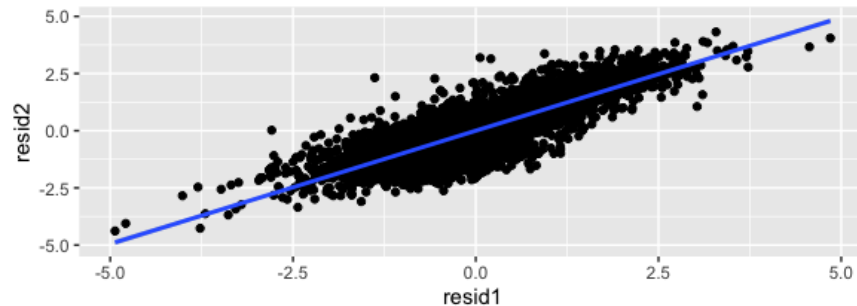


Figure 4:

```
1 splot4 <-
2   dat %>%
3   ggplot(aes(x=resid1, y= resid2))+
4   geom_point()+
5   geom_smooth(method='lm', formula= y~x)
```

3. Write the prediction equation.

```
1 Yi = mod1$residuals + mod2$residuals
```

## 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 mod5 <- lm( difflog + presvote ~ voteshare , data = dat )  
2 summary(mod5)
```

2. Write the prediction equation.

```
1 Y = presvote / difflog + voteshare
```

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 P value is the same as in 4