Problem Set 3

Applied Stats/Quant Methods 1

Due: November 12, 2021

Name: Gareth Moen

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 in `.pdf` form.
- This problem set is due before class on Friday November 12, 2021. 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

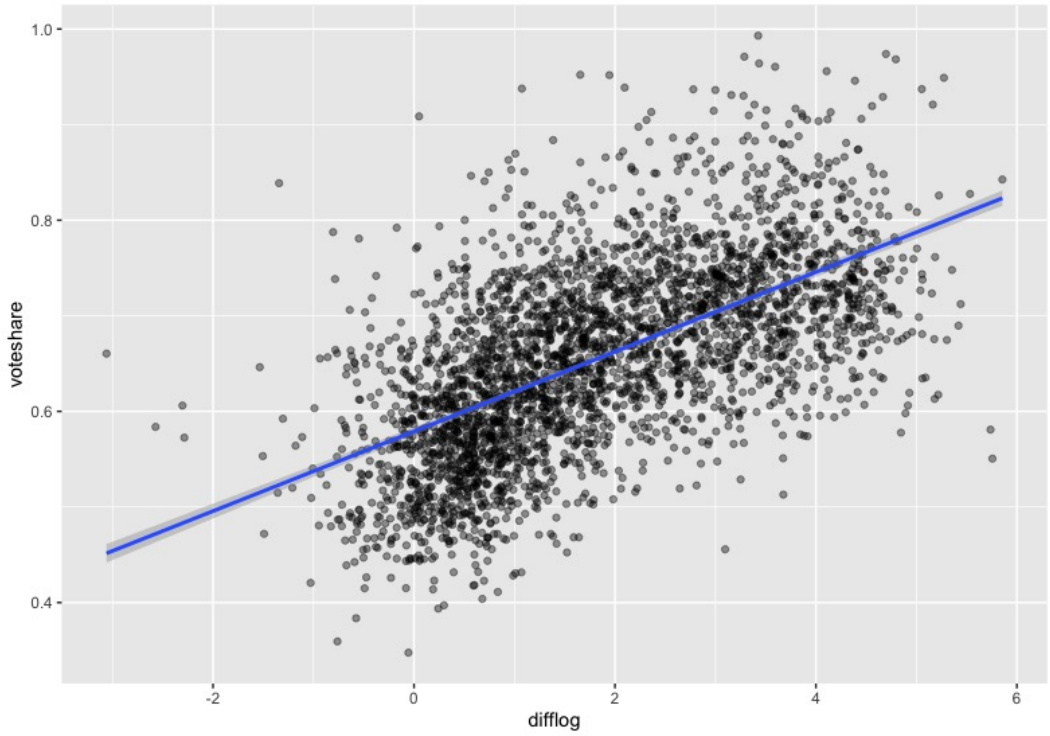
We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

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

```
#Runs a regression and assigns a name to it
Reg_01 <- lm(Incumbents$voteshare ~ Incumbents$difflog)
lm(Incumbents$voteshare ~ Incumbents$difflog)
#Regression coefficients
#y-intercept = 0.579
#slope = b = 0.042
```

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

```
ggplot(aes(x = difflog, y = voteshare), data = Incumbents) +
geom_point(alpha = 0.4) +
geom_smooth(method = "lm", formula = y ~ x)
```



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

```
summary(Reg_01) # summary of regression
#Saving the residuals from first regression as on object
resid_obj1 <- residuals(Reg_01)
```

4. Write the prediction equation.

```
#y-intercept = 0.579
#slope = b = 0.042

#Prediction equation formula is "\hat{y} = a + bx"
# \hat{y} = 0.579 + (0.042)x
```

Question 2

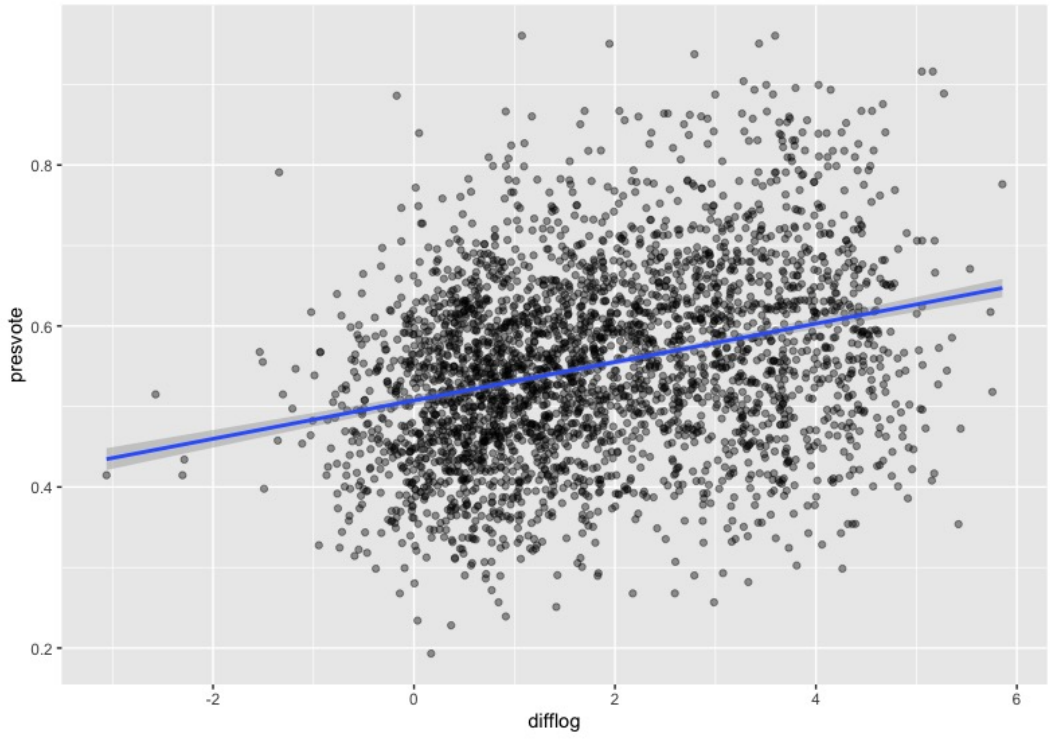
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
#Runs a regression and assigns a name to it
Reg_02 <- lm(Incumbents$presvote ~ Incumbents$difflog)
lm(Incumbents$presvote ~ Incumbents$difflog)
#Regression coefficients
#y-intercept = 0.508
#slope = b = 0.024
```

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

```
#a ggplot view of presvote and difflog with a regression line
ggplot(aes(x = difflog, y = presvote), data = Incumbents) +
geom_point(alpha = 0.4) +
geom_smooth(method = "lm", formula = y ~ x)
```



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

```
# summary of regression
summary(Reg_02)

#Saving the residuals from first regression as on object
resid_obj2 <- residuals(Reg_02)
```

4. Write the prediction equation.

```
#y-intercept = 0.508
#slope = b = 0.024

#Prediction equation formula is "\hat{y} = a + bx"
# \hat{y} = 0.508 + (0.024)x
```

Question 3

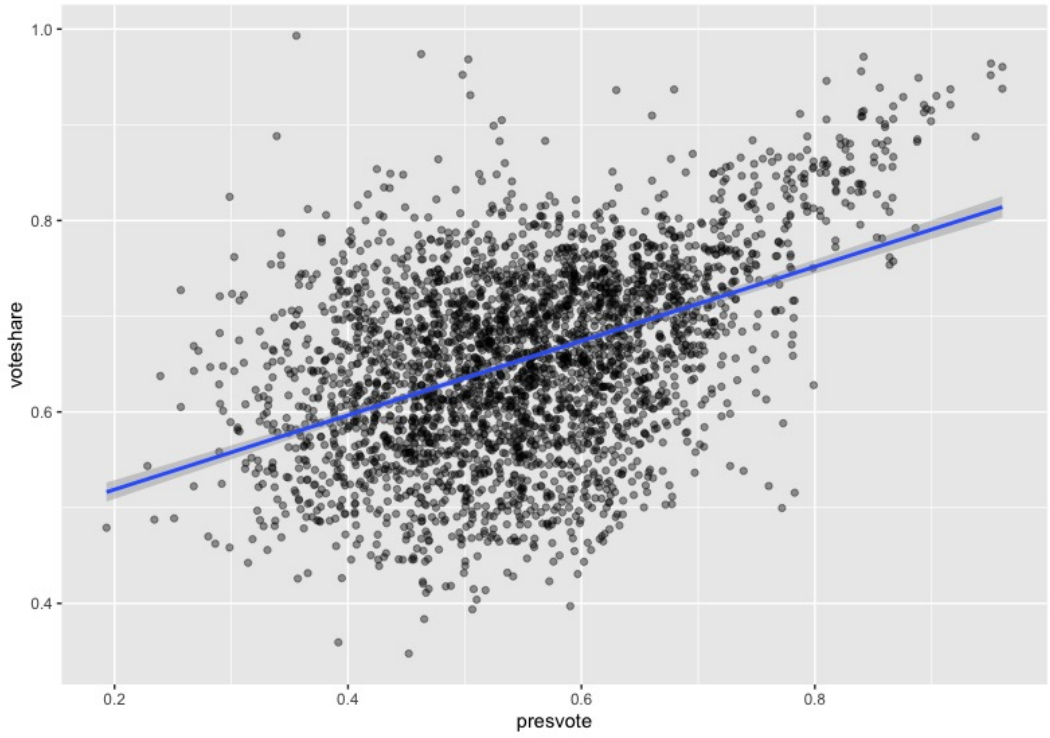
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
#Runs a regression and assigns a name to it
Reg_03 <- lm(Incumbents$voteshare ~ Incumbents$presvote)
lm(Incumbents$voteshare ~ Incumbents$presvote)
#Regression coefficients
# y-intercept is 0.4413
#slope = b = 0.388
```

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

```
#a ggplot view of voteshare and presvote with a regression line
ggplot(aes(x = presvote, y = voteshare), data = Incumbents) +
geom_point(alpha = 0.4) +
geom_smooth(method = "lm", formula = y ~ x)
```



3. Write the prediction equation.

```
#y-interecept = 0.441
#slope = b = 0.388

#Prediction equation formula is "\hat{y} = a + bx"
# \hat{y} = 0.441 + (0.388)x
```

Question 4

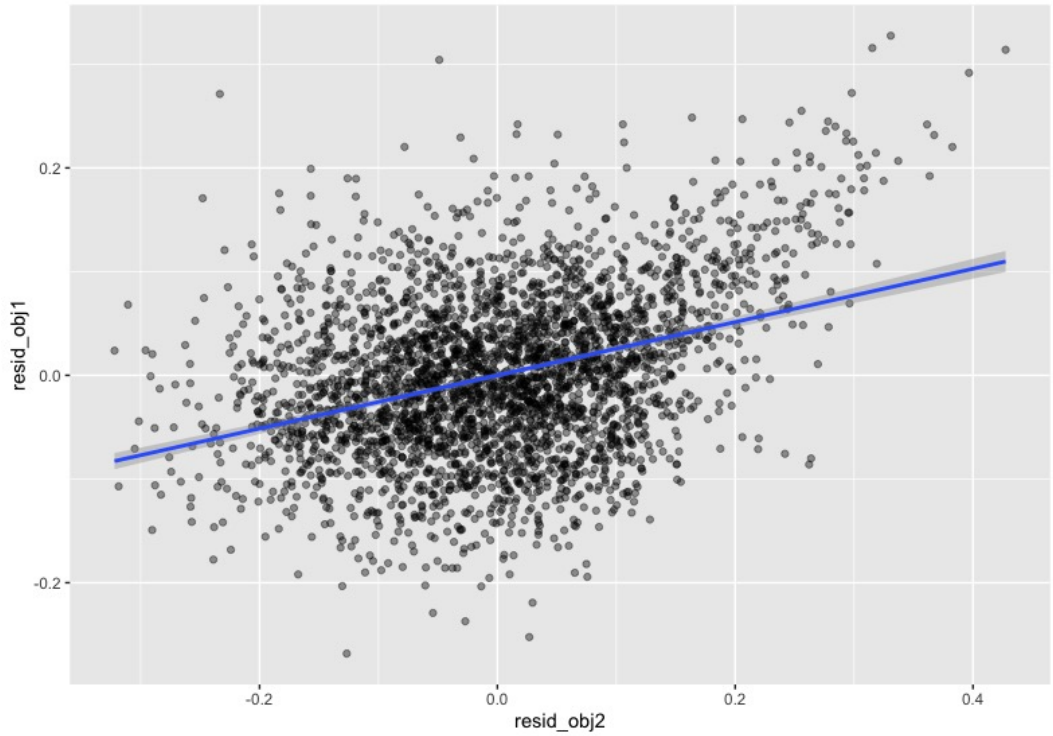
The residuals from part (a) tell us how much of the variation in `voteshare` is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in `presvote` is *not* explained by the difference in spending between incumbent 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.

```
#Runs the regression, assigns a name to it and prints the values
(Reg_04 <- lm(resid_obj1 ~ resid_obj2))
#Regression coefficients
# y-intercept = -1.674e-18
#slope = b = 2.569e-01
```

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

```
#a ggplot view of the residuals with a regression line, not working yet
ggplot(aes(x = resid_obj2, y = resid_obj1), data = Incumbents) +
geom_point(alpha = 0.4) +
geom_smooth(method = "lm", formula = y ~ x)
```



3. Write the prediction equation.

```
#a = y-interecept = -1.674e-18
#slope = b = 2.569e-01

#Prediction equation formula is
\hat{y} = -1.674e-18 + (2.569e-01)x
#When x = 0, y is almost = 0
```

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`.

```
(Reg_05 <- lm(Incumbents$voteshare ~ Incumbents$difflog + Incumbents$presvote))
#Regression coefficients
# y-intercept = 0.449
#slope for Incumbents$difflog = 0.036
#slope for Incumbents$presvote = 0.257
```

2. Write the prediction equation.

```
#a = y-interecept = 0.449

#Prediction equation formula is "\hat{y} = a + b1x1 + b2x2"
# \hat{y} = 0.449 + (0.036)x1 + (0.257)x2
```

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 slope of 0.257 is the same in this equation as in question 4
```