Problem Set 3

Applied Stats/Quant Methods 1

Due: November 12, 2021

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`.

```r
1
2       # 1. run a regression where the outcome variable is 'voteshare' and
        the explanatory variable is 'difflog'
3 # voteshare - outcome variable is the 'dependent variable on Y'
4
5 # difflog - explanatory variable is the 'independent variable (predictor
        variable) on X'
6 # unknown parameter
7 # the basic prediction equation expresses a linear relationship between
        an independent variable
8 # or (x, predictor variable) and a dependent variable (y, a criterian
        variable)
9
10 incumbents <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI
        _Fall2021/main/datasets/incumbents_subset.csv")
11 summary(incumbents) # this is a simple regression or summary(..)
12 str(incumbents)
13 attach(incumbents)
14 names(incumbents)
15
16
17 # difflog depends on voteshare ..
18 # q1 <- lm(difflog ~ voteshare, data = incumbents)
19 # summary(q1)
20 q1 <- lm(voteshare ~ difflog, data = incumbents)
21 summary(q1)
22
23 plot(difflog ~ voteshare, data = incumbents)
24 abline(lm(difflog ~ voteshare, data = incumbents), col = "blue")
25
26
```

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

```r
1
```

```r
2      ### 2. Make a scatterplot of the two variables and add the regression
        line ####
3
4  # the ggplot plots the regression line
5  #q1_plot <- ggplot(incumbents, aes(difflog, voteshare)) +
6    # geom_point(alpha = 0.5) + # add a scatterplot
7  #  geom_smooth(method = "lm") # add a linear regression line
8  # run the plot
9  #q1_plot
10
11 class(voteshare)
12 class(difflog)
13 plot(difflog, voteshare, main = "Scatterplot")
14 cor(difflog, voteshare) # x,y
15
16 # testing the incumbents_subset dataset
17 # scatter.smooth(x=incumbents$difflog, y=incumbents$voteshare, main="
      Scatter")
18
19 # the ggplot plots the regression
20 q1_plot <- ggplot(incumbents, aes(difflog, voteshare)) +
21   geom_point(alpha = 0.5) + # add a scatterplot
22   geom_smooth(method = "lm") + # add a linear regression line
23   labs(x = "difflog", y = "voteshare")
24 q1_plot
25
26 # using the lm() function for plotting the output
27
28 summary(lm(data = incumbents, difflog~ voteshare))
29
30
```
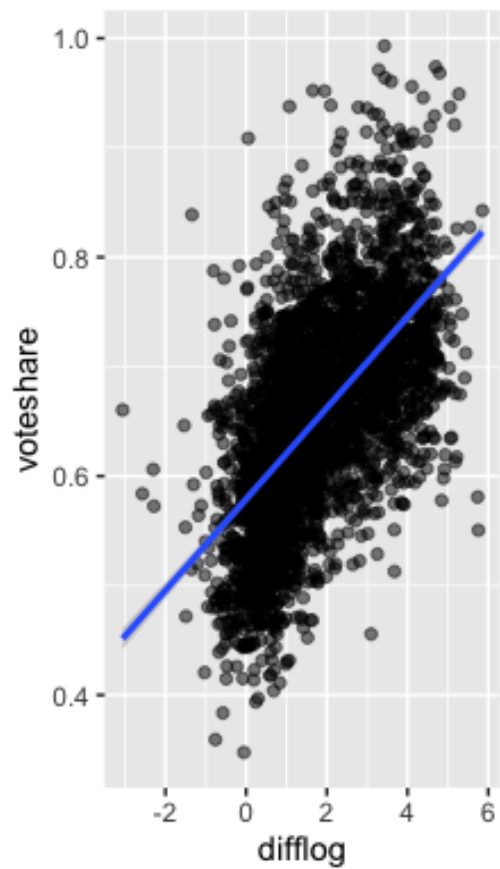
3. Save the residuals of the model in a separate object.

```
1
2    ### 3. Save the residuals of the model in a separate object
3
4 # plotting the residuals
5
```

```
6  q1_resid <- resid(q1) #a function for extracting the residuals from
7  #a model object
8  plot(incumbents$difflog, q1_resid)
9  abline(h = 0, col = "red")
10
11 q1resid <- q1$residuals
12 q1resid
13
14 q1_resid <- resid(q1) #a function for extracting the residuals from
15 #a model object
16 plot(incumbents$voteshare, q1_resid)
17 abline(h = 0, col = "red")
18
19
20
```

4. Write the prediction equation.

```
1
2 # (Intercept) 0.579031 and difflog  0.041666 (from the summary)
3 # x is the difference for challenger's spending
4  # y (hat) = 0.579031 + 0.041666x
5
6
7
```

(Intercept) 0.579031 and difflog 0.041666 is from summary(q1) x is the difference for challenger's spending

$\hat{y} = 0.579031 + 0.041666\text{x}$

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`.

```
1
2      ## 1. Run a regression where the outcome variable is 'prevote' and
       the explanatory variable
3  # is 'difflog'
4
5  # 'prevote' is the dependent variable (outcome variable) on Y axis
6  # 'difflog' is the independent variable (explanatory variable) on X axis
7
8      q2 <- lm(presvote ~ difflog, data = incumbents)
9  summary(q2)
10 plot(difflog ~ presvote, data = incumbents)
11 abline(lm(difflog ~ presvote, data = incumbents), col = "blue")
12
13
```
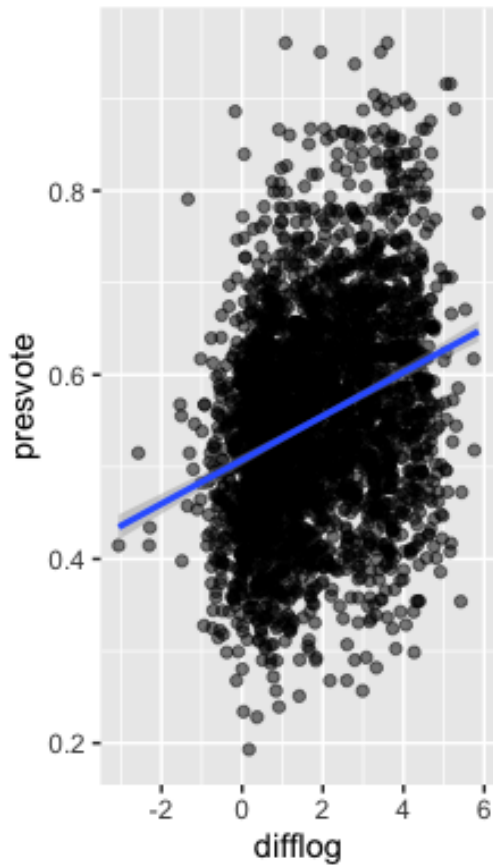
2. Make a scatterplot of the two variables and add the regression line.

```
1
```

```
2 ## 2. Make a scatterplot of the two variables and add the regression line
3
4 # the ggplot plots the regression
5 q2_plot <- ggplot(incumbents, aes(difflog, presvote)) +
6   geom_point(alpha = 0.5) + # add a scatterplot
7   geom_smooth(method = "lm") + # add a linear regression line
8   labs(x = "difflog", y = "presvote")
9 q2_plot
10
11
```


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

```
1
2     ## 3. Save the residuals of the model in a separate object.
3
4 q2_resid <- resid(q2) #a function for extracting the residuals from
5 #a model object
6 plot(incumbents$voteshare, q2_resid)
7 abline(h = 0, col = "red")
8
9 q2resid <- q2$residuals
10 q2resid
11
12
```

4. Write the prediction equation.

(Intercept) 0.507583 and difflog 0.023837 is from summary(q2)

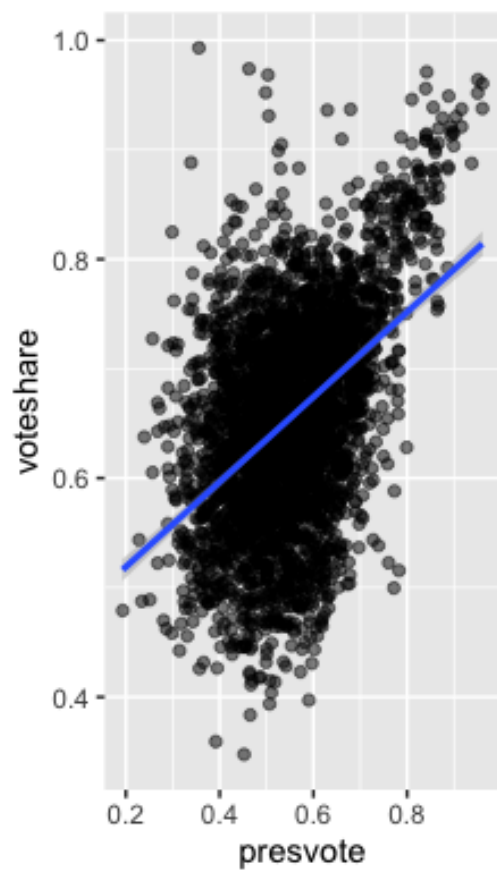$\hat{y} = 0.507583 + 0.023837\text{x}$

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`.

```r
# 1. Run a regression where the outcome variable is 'voteshare' and
   the explanatory varaible'
# is 'presvote'

q3 <- lm(voteshare ~ presvote, data = incumbents)
summary(q3)
plot(presvote ~ voteshare, data = incumbents)
abline(lm(presvote ~ voteshare, data = incumbents), col = "blue")
```

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

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

# the ggplot plots the regression
q3_plot <- ggplot(incumbents, aes(presvote, voteshare)) +
  geom_point(alpha = 0.5) + # add a scatterplot
  geom_smooth(method = "lm") # add a linear regression line
q3_plot
```

3. Write the prediction equation.

(Intercept) 0.441330 and presvote 0.388018 is from the summary(q3)

$\hat{y} = 0.507583 + 0.023837x$

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
2     ## 1. Run a regression where the outcome variable is the residuals
        from Question 1 and
3 # the explanatory variable is the residuals from Question 2.
4
5
6 # q1resid depends on q2resid
7 q4resid <- lm(q1resid ~ q2resid)
8 summary(q4resid)
9
10
```

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

```
1
2      ## 2. Make a scatterplot of the two residuals and add the regression
         line
3
4  q4_plot <- ggplot(aes(q2resid, q1resid), data = NULL) + # add the two
         residuals
5    geom_point() + # add a scatterplot
6    geom_smooth(method = "lm") # add the linear regression line
7  q4_plot
8
9
```
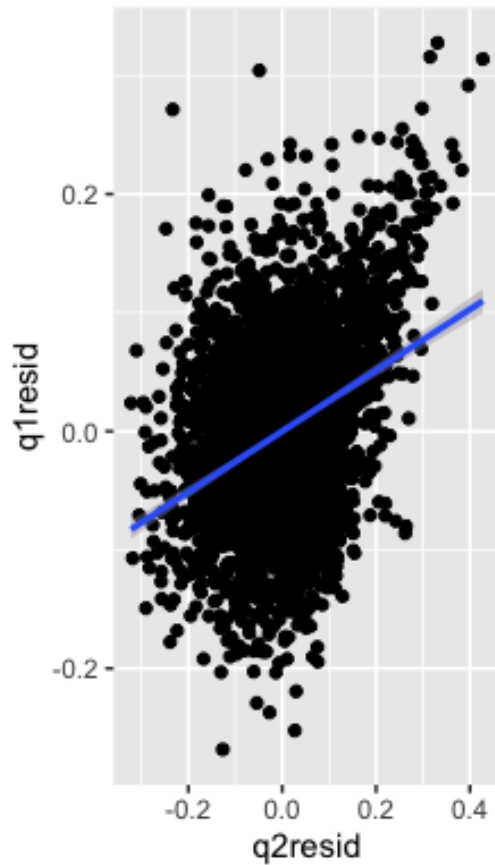

3. Write the prediction equation.

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
2
3       ## 1. Run a regression where the outcome variable is the incumbent'
        s 'voteshare' and
4  # the explanatory variables are 'difflog' and 'presvote'
5
6
7  q5resid <- lm(voteshare ~ difflog + presvote, incumbents)
8  # summary(lm(voteshare ~ difflog + presvote, incumbents))
9  summary(q5resid)
10
11
```

2. Write the prediction equation.

```
1
2       ## 2. Write the prediction equation
3
4  # (Intercept)  0.4486442
5  # difflog      0.0355431
```

```
6 # presvote    0.2568770
7
8
```

$$\hat{y} = 0.4486442 + (0.0355431)\text{x1} + (0.2568770)\text{x2}$$

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