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`.

```
# read in incumbents data subset from online .csv
incumbents <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2021/
# run regression model with voteshare regressed on difflog
regression_model_problem2 <- lm(voteshare ~ difflog, data=incumbents)
# get summary of model with coefficient estimates
summary(regression_model_problem1)

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

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
```

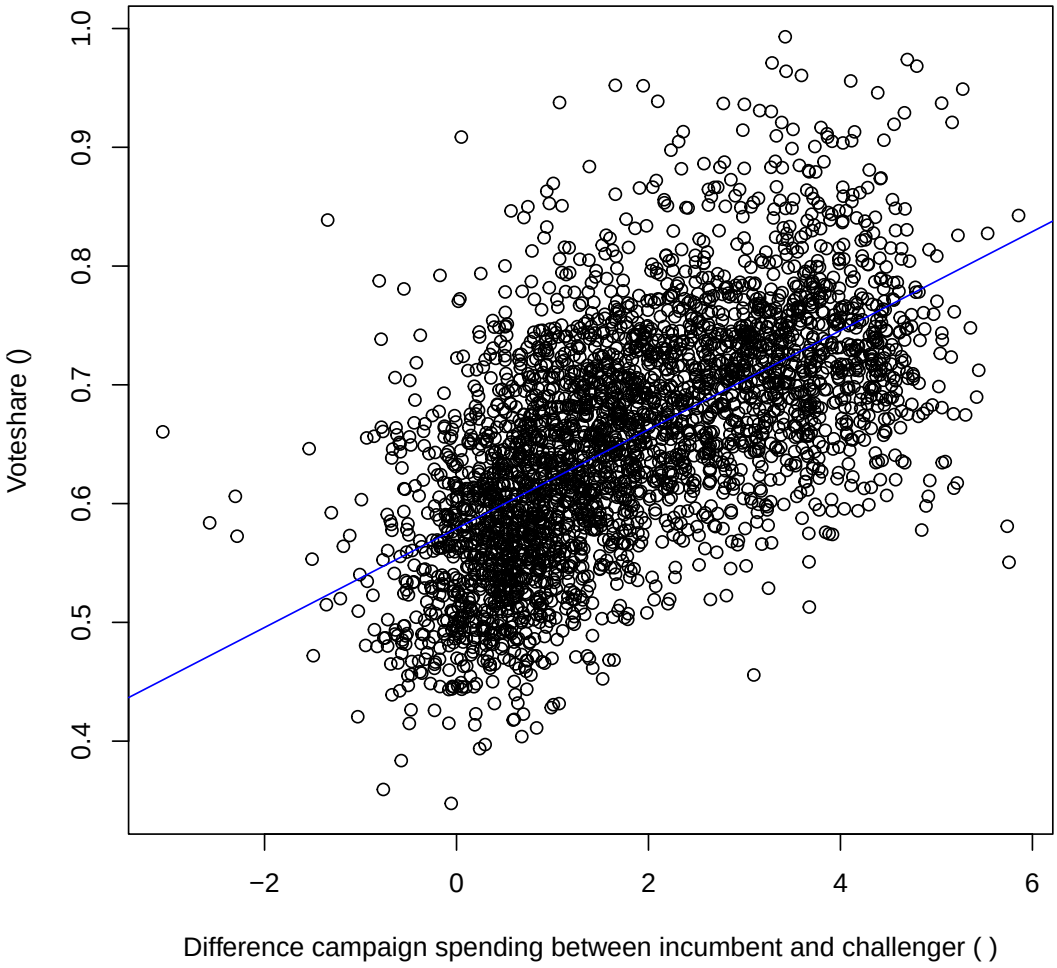
2. Make a scatterplot of the two variables and add the regression line.

```
begin{figure}[width=0.9\textwidth]
\includegraphics[width=8cm]{Figure1.pdf}
\end{figure}
plot(incumbents$difflog, incumbents$voteshare,
xlab = "Difference campaign spending between incumbent and challenger ( )", ylab =

abline(regression_model_problem1, col = "blue")
```

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

```
resids1 <- regression_model_problem1$residuals
```

4. Write the prediction equation.


Voteshare ()
0.4
0.5
0.6
0.7
0.8
0.9
1.0
−2
0
2
4
6
Difference campaign spending between incumbent and challenger ( )

```
#mean of outcome variable = the y-intercept or constant + slope of predictor1 multi
mean of outcome variable = t value of y intercept in table +
```

Question 2

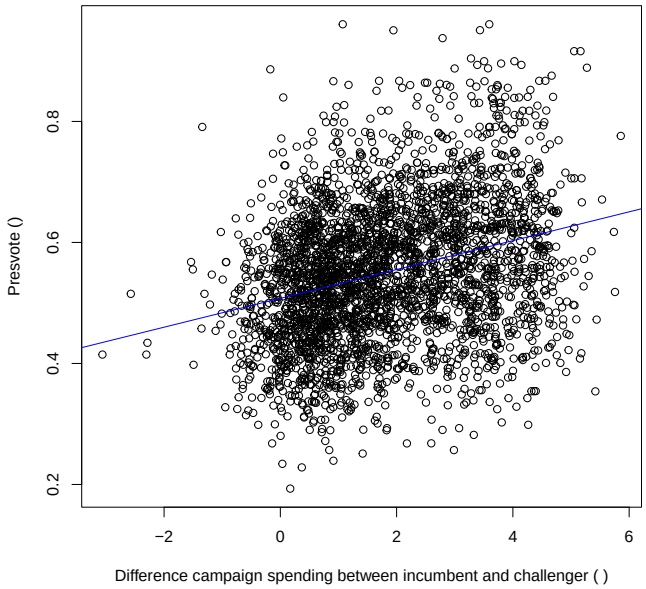
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

```
regression_model_problem2 <- lm(presvote ~ difflog, data=incumbents)

summary(regression_model_problem2)

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

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.

```
begin{figure}[width=0.9\textwidth]
\includegraphics[width=8cm]{Figure2.pdf}
\end{figure}
plot(incumbents$difflog, incumbents$presvote,
xlab = "Difference campaign spending between incumbent and challenger ( )", ylab =
```




```
# Add the regression line to the scatterplot
abline(regression_model_problem2, col = "blue")
```

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

```
resids2 <- regression_model_problem2$residuals
```

4. Write the prediction equation.

```
#mean of outcome variable = the y-intercept or constant + slope of predictor1 multiplied
mean of outcome variable = 257.19 +
```

Question 3

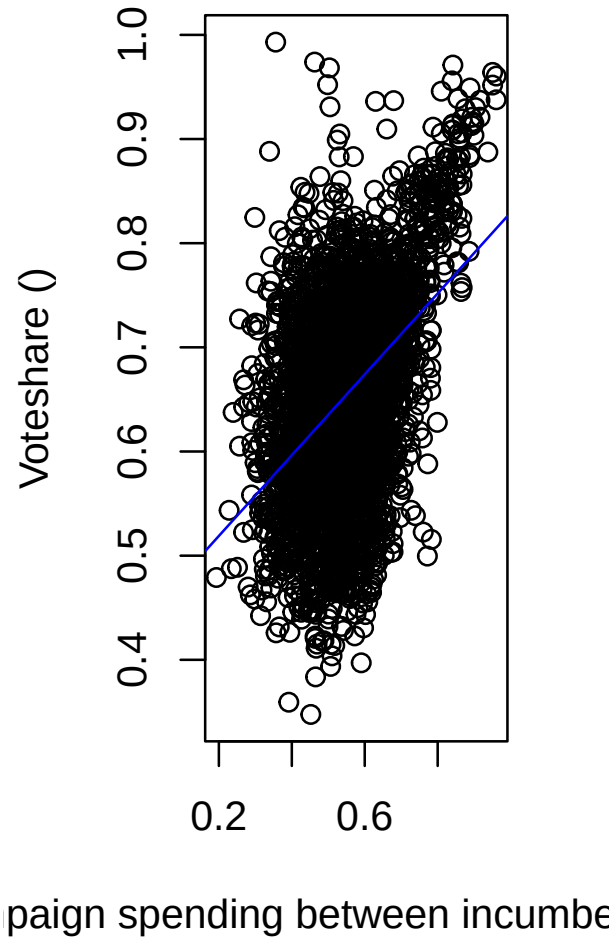
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

```
regression_model_problem3 <- lm(voteshare ~ presvote, data=incumbents)


# get summary of model with coefficient estimates
summary(regression_model_problem3)
Call:
lm(formula = voteshare ~ presvote, data = incumbents)

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.

```
plot(incumbents$presvote, incumbents$voteshare,
xlab = "Difference campaign spending between incumbent and challenger ( )", ylab =

# Add the regression line to the scatterplot
abline(regression_model_problem3, col = "blue")
```



beginfigure

Write the prediction equation.

```
#mean of outcome variable = the y-intercept or constant + slope of predictor1 multiplied
mean of outcome variable = 58.08 +
```

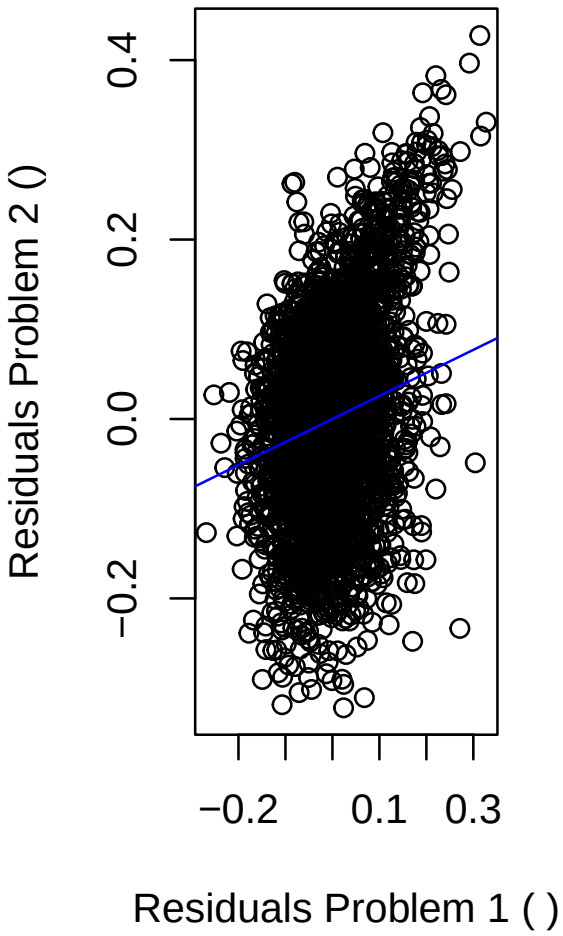
Question 4

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.

```
regression_model_problem4 <- lm(resids1 ~ resids2 , data=incumbents)

summary(regression_model_problem4)

Call:
lm(formula = resids1 ~ resids2, data = incumbents)

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
resids2      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.



beginfigure

Write the prediction equation.

```
#mean of outcome variable = the y-intercept or constant + slope of predictor1 multiplied
mean of outcome variable = 00.00 + 21.84 x
```

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`.

```
regression_model_problem5 <- lm( voteshare ~ difflog + presvote, data=incumbents)
\\
View(regression_model_problem5)
regression_model_problem5[["coefficients"]]
(Intercept)     difflog    presvote
0.44864422  0.03554309  0.25687701

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

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

(Intercept) ***
difflog     ***
presvote    ***
---
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.

```
#mean of outcome variable = the y-intercept or constant + slope of predictor1 multi
```

```
mean of outcome variable = 0.4486442 + 0.03554309 x  + 0.256877 x
```

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 mean of outcome variable - the slope numeric amount is the same in each equation due
```