Problem Set 3

Applied Stats/Quant Methods 1

Due: November 19, 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 19, 2023. No late assignments will be accepted.

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
# Run the regression
reg_model1 <- lm(voteshare ~ difflog, data = inc_sub)

# Print the regression summary
summary(reg_model1)
```

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

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.

```r
# Create the scatterplot
plot(inc_sub$difflog, inc.sub$voteshare,
     main = "Scatterplot with Regression Line", xlab = "difflog", ylab = "voteshare")

# Fit the regression line
reg_line1 <- lm(inc_sub$voteshare ~ inc_sub$difflog)
abline(reg_line1, col = "red")
```

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

```r
# Save the residuals in a separate object
residualsq1 <- reg_model1$residuals

# Print the residuals
print(residualsq1)
```

```r
#The prediction equation for the given model is:

#voteshare = 0.579031 + 0.041666 * difflog
```

4. Write the prediction equation.

```r
# Run the regression
reg_model2 <- lm(presvote ~ difflog, data = inc_sub)

# Print the regression summary
summary(reg_model2)
```

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

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
```

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`.

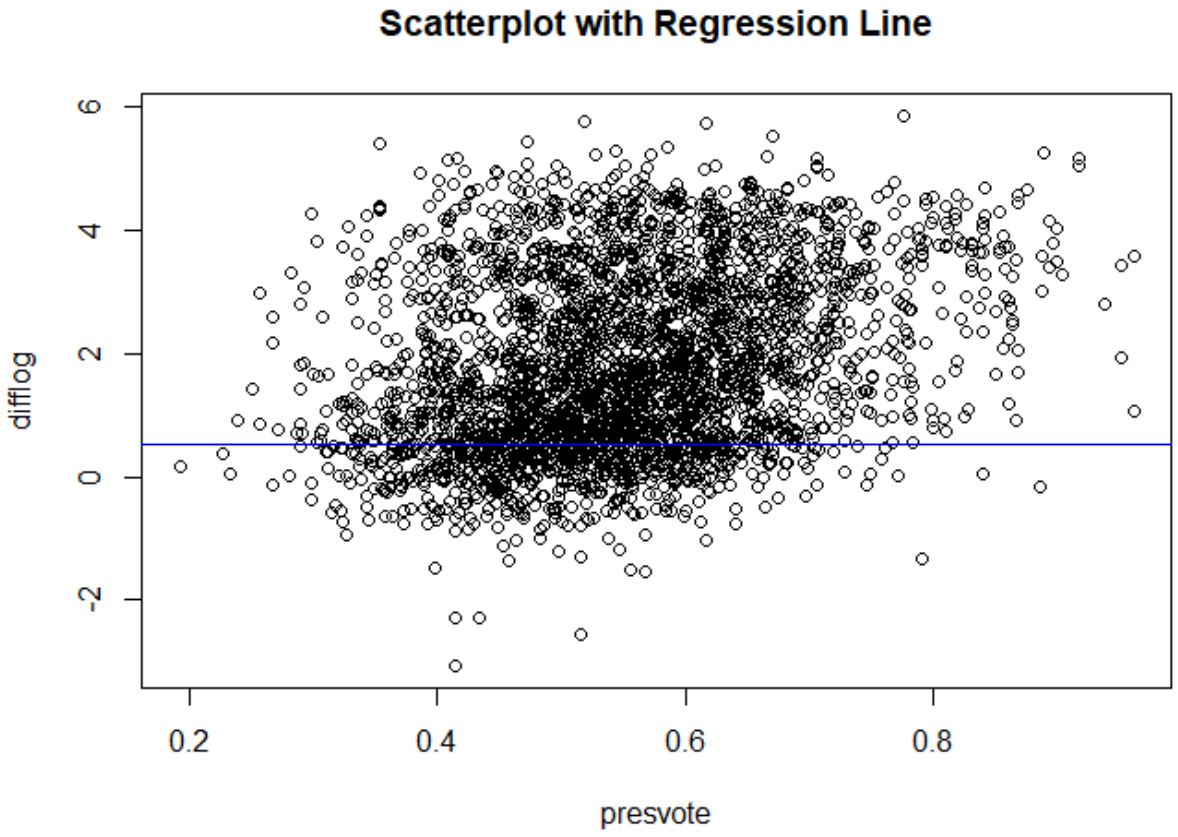
2. Make a scatterplot of the two variables and add the regression line.

```r
# Create the scatterplot
plot(inc_sub$presvote, inc.sub$difflog, main = "Scatterplot with Regression Line",
     xlab = "presvote", ylab = "difflog")

# Fit the regression line
reg_line2 <- lm(inc_sub$presvote ~ inc_sub$difflog)
abline(reg_line2, col = "blue")
```

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

4. Write the prediction equation.




```
# Save the residuals in a separate object
residualsq2 <- reg_model2$residuals

# Print the residuals
print(residualsq2)
```

```
#The prediction equation for the given model is:

#presvote = 0.507583 + 0.023837 * difflog
```

```r
# Run the regression
reg_model3 <- lm(voteshare ~ presvote, data = inc.sub)

# Print the regression summary
summary(reg_model3)
```

```
Call:
lm(formula = voteshare ~ presvote, data = inc.sub)

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
```

Question 3

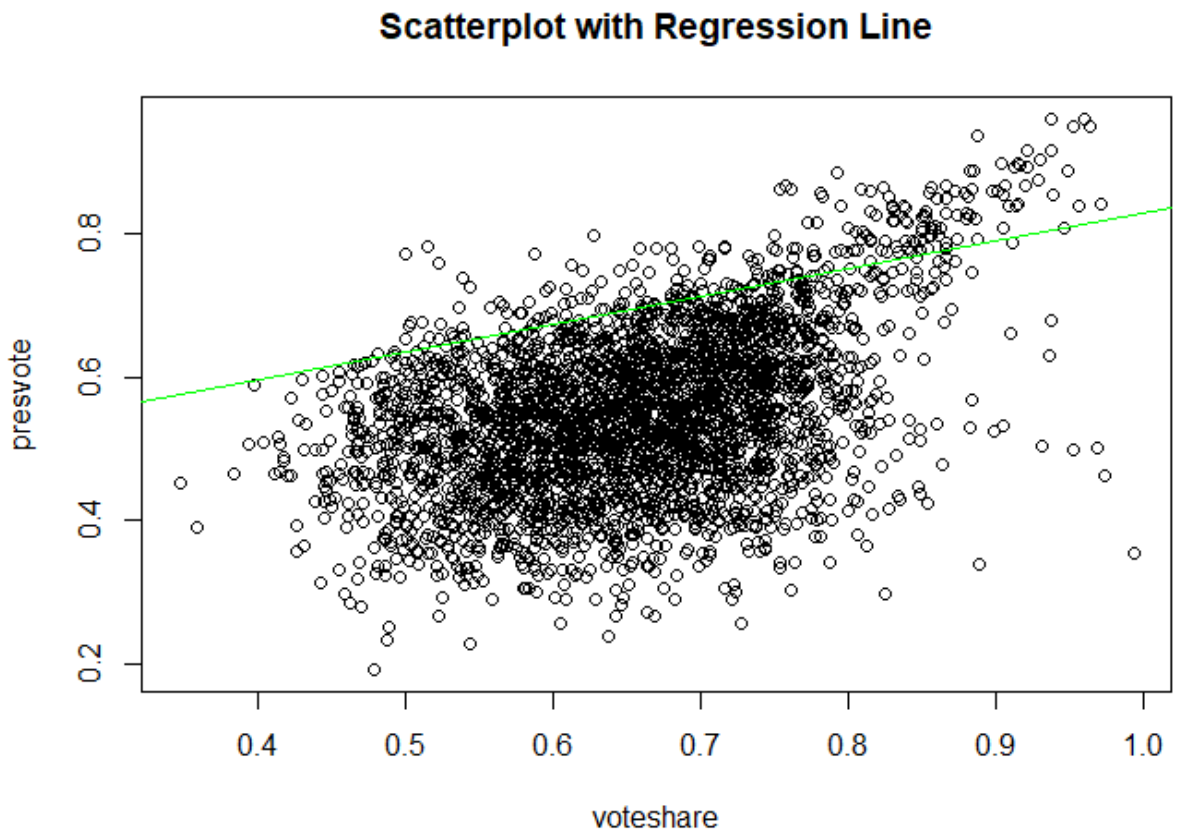
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

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

```r
plot(inc_sub$voteshare, inc.sub$presvote, main = "Scatterplot with Regression Line",
     xlab = "voteshare", ylab = "presvote")

# Fit the regression line
reg_line3 <- lm(inc_sub$voteshare ~ inc_sub$presvote)
abline(reg_line3, col = "green")
```



```
#The prediction equation for the given model is:

#voteshare = 0.441330 + 0.388018 * presvote
```

3. Write the prediction equation.

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.

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

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`.

2. Write the prediction equation.

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