Problem Set 3

Applied Stats/Quant Methods 1
Jianxiong Wu—23354731

Due: November 11, 2024

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 11, 2024. 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`.

```
1  #1.1
2  q1_regression <- lm(voteshare ~ difflog, data = inc.sub)
3  summary(q1_regression)
```

```
Results:
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
```

This regression shows that the difference in campaign spending between incumbent and challenger has a significant positive effect on the incumbent's vote share. On average, each unit increase in difflog increases the incumbent's vote share by about 0.042.

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

```
1 #1.2
2 png(file = "q1_regression.png")
3 plot(inc.sub$difflog, inc.sub$voteshare,
4     xlab = "difference in campaign spending between incumbent and
       challenger",
5     ylab = "the incumbent's vote share",
6     main = "scatterplot of voteshare & difflog")
7 abline(q1_regression, col = "red")
8 dev.off()
```

scatterplot of voteshare & difflog

difference in campaign spending between incumbent and challenger
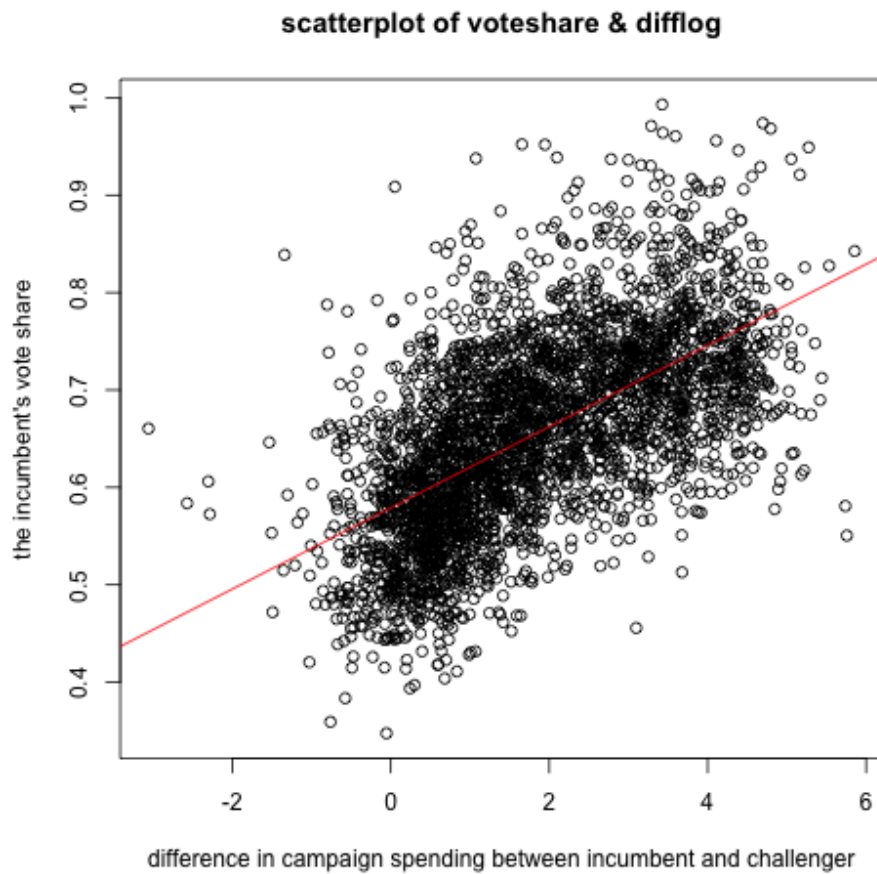
3. Save the residuals of the model in a separate object.

```
1 #1.3
2 q1_residuals <- q1_regression$residuals
3 q1_residuals
```

4. Write the prediction equation.

```
Results:
voteshare = intercept + coefficient * difflog
voteshare = 0.579031 + 0.041666 * difflog
```

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`.

```
#2.1
q2_regression <- lm(presvote ~ difflog, data = inc.sub)
summary(q2_regression)
```

```
Results:
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


This regression shows that the difference in campaign spending
between incumbent and challenger has a significant positive effect
on the vote share of the presidential candidate. On average, each
unit increase in difflog increases the vote share of the presidential
candidate by about 0.024.
```

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

```
#2.2
png(file = "q2_regression.png")
```

4

```
3  plot(inc.sub$difflog, inc.sub$presvote,
4      xlab = "difference between incumbent and challenger's spending",
5      ylab = "the vote share of the presidential candidate of the
      incumbent's party",
6      main = "scatterplot of presvote & difflog")
7  abline(q2_regression, col = "red")
8  dev.off()
```


scatterplot of presvote & difflog
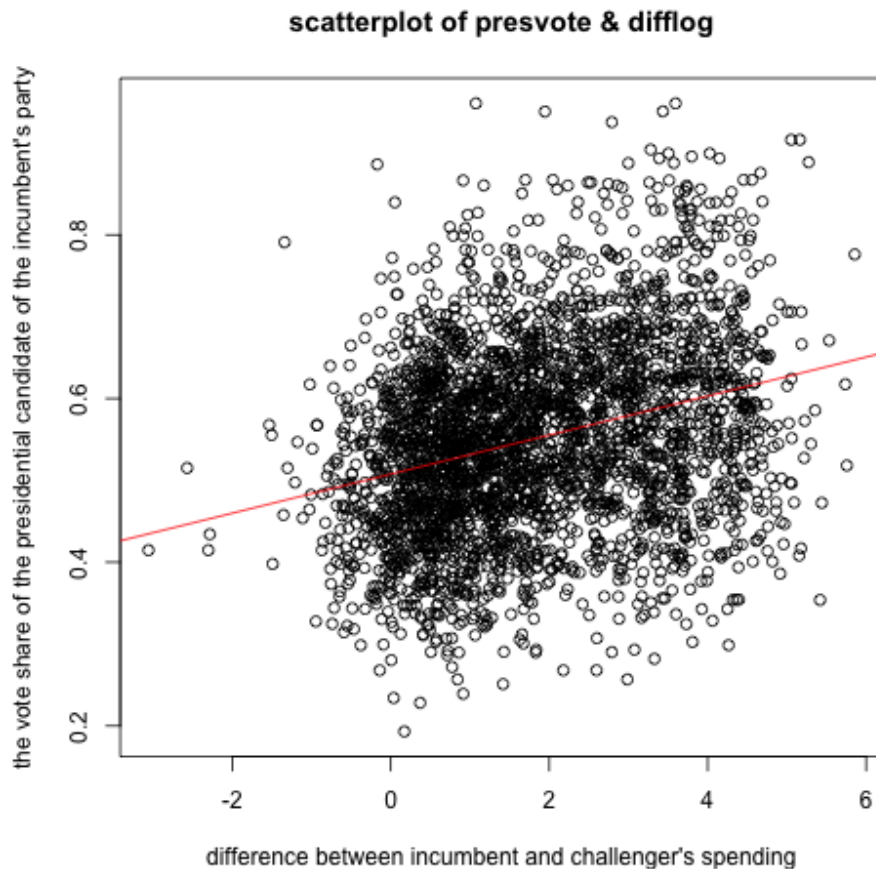
3. Save the residuals of the model in a separate object.

```
1  #2.3
2  q2_residuals <- q2_regression$residuals
3  q2_residuals
```

4. Write the prediction equation.

```
Results:
presvote = intercept + coefficient * difflog
presvote = 0.507583 + 0.023837 * 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`.

```
#3.1
q3_regression <- lm(voteshare ~ presvote, data = inc.sub)
summary(q3_regression)
```

```
Results:
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


This regression shows that the vote share of the presidential
candidate has a significant positive effect on the incumbent's
vote share. On average, each unit increase in prevote increases
the incumbent's vote share by about 0.388.
```

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

```
#3.2
png(file = "q3_regression.png")
plot(inc.sub$presvote, inc.sub$voteshare,
    xlab = "the vote share of the presidential candidate of the
    incumbent's party",
```

```
5        ylab = "the  incumbents  electoral  success",
6        main = "scatterplot  of  voteshare  &  presvote")
7  abline(q3_regression,  col = "red")
8  dev.off()
```

scatterplot of voteshare & presvote

the vote share of the presidential candidate of the incumbent's party
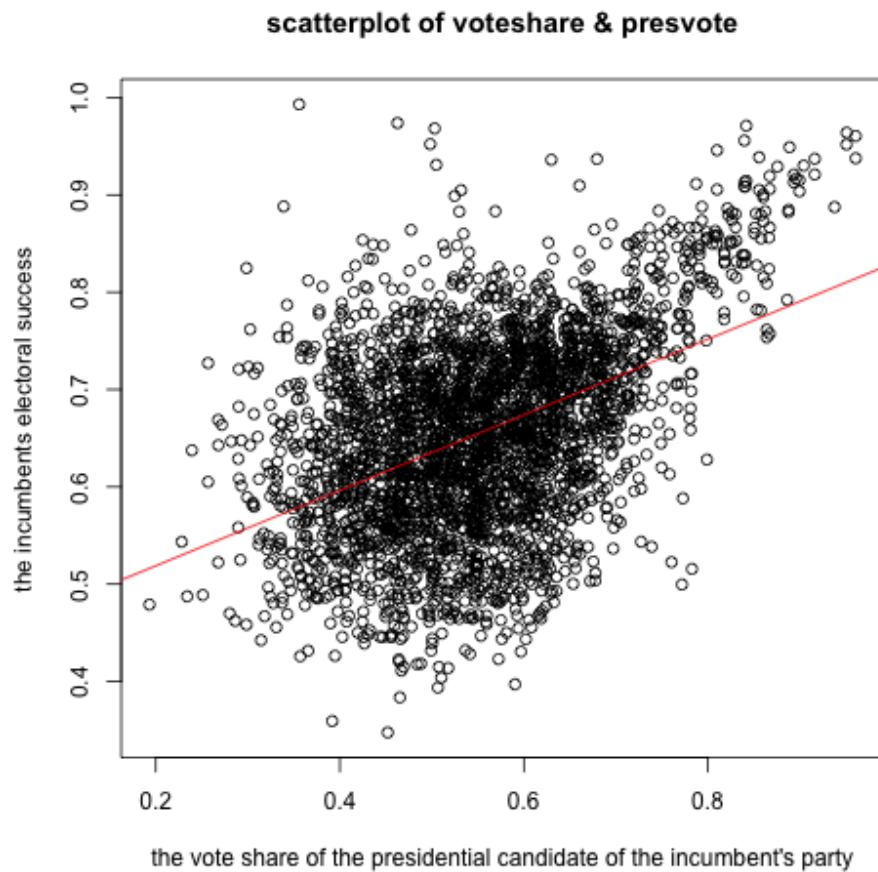
3. Write the prediction equation.

```
Results:
voteshare = intercept + coefficient * presvote
voteshare = 0.441330 + 0.388018 * presvote
```

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.

```
#4.1
q4_regression <- lm(q1_residuals ~ q2_residuals, data = inc.sub)
summary(q4_regression)
```

```
Call:
lm(formula = q1_residuals ~ q2_residuals, data = inc.sub)

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)  -1.942e-18  1.299e-03    0.00        1
q2_residuals  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


This regression shows that question2 residuals has a significant
positive effect on question1 residuals. On average, each unit
increase in question2 residuals increases question1 residuals by
about 0.2569.
```

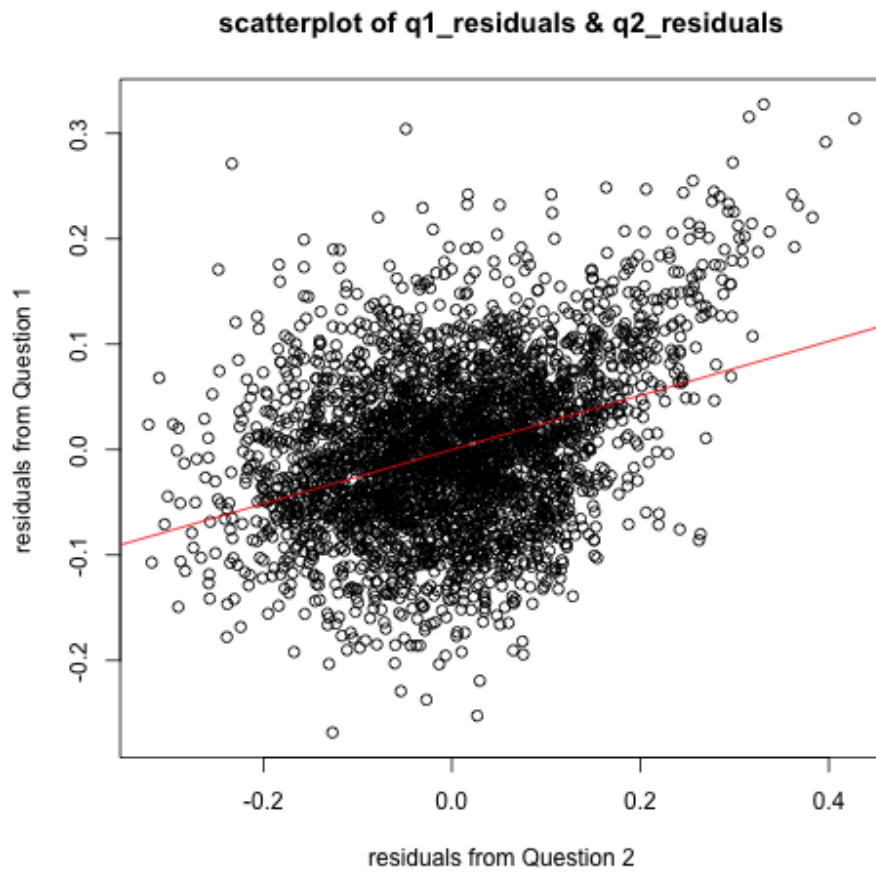2. Make a scatterplot of the two residuals and add the regression line.

```
#4.2
png(file = "q4_regression.png")
plot(q2_residuals, q1_residuals,
     xlab = "residuals from Question 2",
```

```
5        ylab = "residuals from Question 1",
6        main = "scatterplot of q1_residuals & q2_residuals")
7  abline(q4_regression, col = "red")
8  dev.off()
```

scatterplot of q1_residuals & q2_residuals

3. Write the prediction equation.

```
Results:
q1_residuals = intercept + coefficient * q2_residuals
q1_residuals = (-1.942e-18) + (2.569e-01) * q2_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 #5.1
2 q5_regression <- lm(voteshare ~ difflog + presvote, data = inc.sub)
3 summary(q5_regression)
```

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

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 ***
---
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


This regression is a multivariate regression, with both difflog
and prevote as independent variables. On average, when difflog
is held constant, each unit increase in prevote increases the
incumbent's vote share by approximately 0.2569. When prevote
is held constant, each unit increase in difflog increases the
incumbent's vote share increases by approximately 0.0355.
Both explanatory variables have a significant positive effect on
the incumbent's vote share.
```

2. Write the prediction equation.

   ```
   Results:
   voteshare = intercept + coefficient1 * difflog + coefficient2 * presvote
   voteshare = 0.448644 + 0.035543 * difflog + 0.256887 * presvote
   ```

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

   ```
   Results:
   The coefficient of presvote in the q5_regression in Question 5 is the
   same as the coefficient of q2_residuals in the q4_regression in
   Question 4, which is 0.2569. The reason for this is that in the
   q5_regression, difflog is used as an explanatory variable along with
   presvote. And in q4_regression, the residuals of voteshare ~ difflog
   and presvote ~ difflog are used for the regression. So there is an
   independent effect of presvote on vote share after controlling for
   difflog.
   ```