Problem Set 3

Applied Stats/Quant Methods 1

Zexi Wang

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.

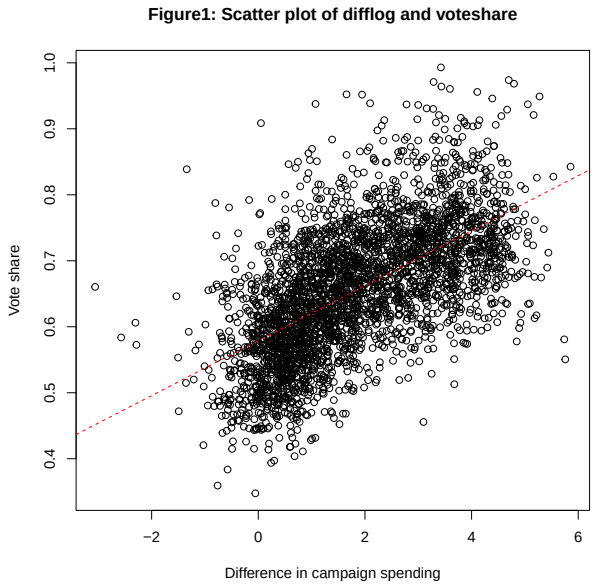
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
########### Question 1 ###########
# read in data
inc.sub <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/incumbents_subset.csv")

### 1.1 ###
# outcome variable-voteshare, explanatory variable-difflog
model_vote_dif <- lm(voteshare ~ difflog, data = inc.sub)
```

```
================================================
Dependent variable:
---------------------------
voteshare
------------------------------------------------
difflog                    0.042*** (0.001)
Constant                   0.579*** (0.002)
------------------------------------------------
Observations                     3,193
R2                               0.367
Adjusted R2                      0.367
Residual Std. Error        0.079 (df = 3191)
F Statistic          1,852.791*** (df = 1; 3191)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

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

```r
### 1.2 ###
# scatter plot
pdf("Q1_2_scatterplot.pdf")
plot( inc.sub$difflog,
      inc.sub$voteshare,
      xlab = "Difference in campaign spending", ylab = "Vote share",
      main = "Figure1: Scatter plot of difflog and voteshare")
abline(model_vote_dif, col="red", lty=2)
dev.off()
```



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

```r
### 1.3 ###
# residuals
model_vote_dif_resid <- resid(model_vote_dif)
```

4. Write the prediction equation.

$$voteshare = 0.579 + 0.042 \times \text{difflog}$$

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.

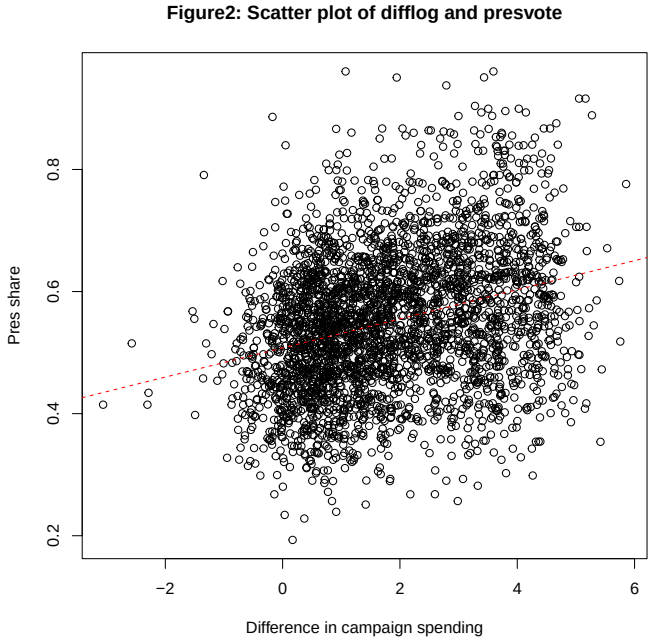
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```r
############ Question 2 ############

### 2.1 ###
# outcome variable-presvote, explanatory variable-difflog
model_pres_dif <- lm(presvote ~ difflog, data = inc.sub)
```

```
================================================
Dependent variable:
---------------------------
presvote
------------------------------------------------
difflog                    0.024*** (0.001)
Constant                   0.508*** (0.003)
------------------------------------------------
Observations                     3,193
R2                               0.088
Adjusted R2                      0.088
Residual Std. Error        0.110 (df = 3191)
F Statistic           307.715*** (df = 1; 3191)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

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

```r
### 2.2 ###
# scatter plot
pdf("Q2_2_scatterplot.pdf")
plot( inc.sub$difflog,
      inc.sub$presvote,
      xlab = "Difference in campaign spending", ylab = "Pres share",
      main = "Figure2: Scatter plot of difflog and presvote")
abline(model_pres_dif, col="red", lty=2)
dev.off()
```



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

```r
### 2.3 ###
# residuals
model_pres_dif_resid <- resid(model_pres_dif)
```

4. Write the prediction equation.

$$presvote = 0.508 + 0.024 \times \text{difflog}$$

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`.

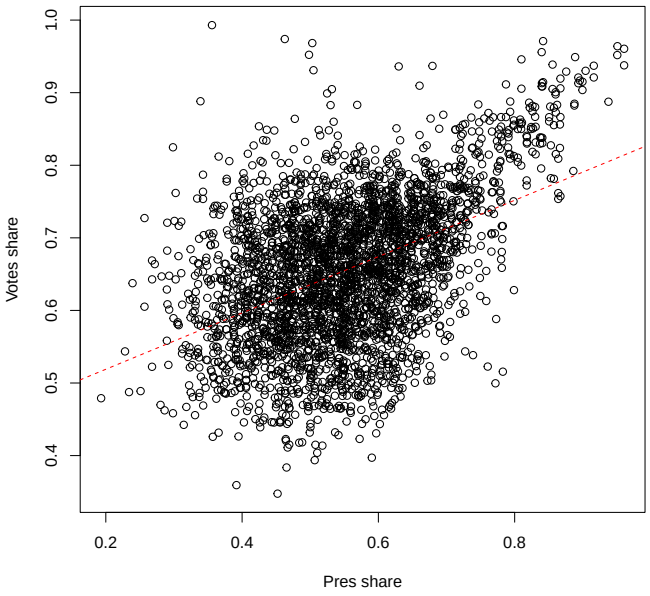
```
########### Question 3 ###########

### 3.1 ###
# outcome variable-voteshare, explanatory variable-presvote
model_vote_pres <- lm(voteshare ~ presvote, data = inc.sub)
```

```
================================================
Dependent variable:
---------------------------
voteshare
------------------------------------------------
presvote                   0.388*** (0.013)
Constant                   0.441*** (0.008)
------------------------------------------------
Observations                     3,193
R2                               0.206
Adjusted R2                      0.206
Residual Std. Error        0.088 (df = 3191)
F Statistic           826.950*** (df = 1; 3191)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

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

```
### 3.2 ###
# scatter plot
pdf("Q3_2_scatterplot.pdf")
plot( inc.sub$presvote,
      inc.sub$voteshare,
      xlab = "Pres share",ylab = "Votes share",
      main = "Figure3: Scatter plot of presvote and voteshare")
abline(model_vote_pres, col="red", lty=2)
dev.off()
```

Figure3: Scatter plot of presvote and voteshare



3. Write the prediction equation.

$$voteshare = 0.441 + 0.388 \times presvote$$

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.

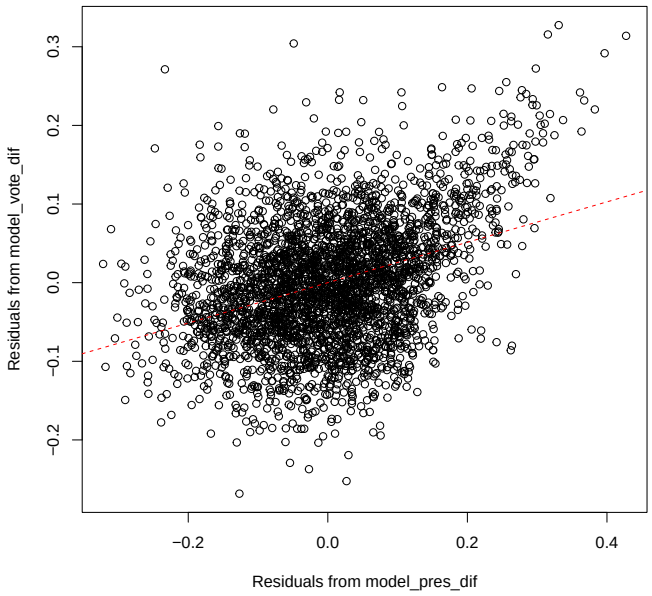
```
############ Question 4 ############

### 4.1 ###
# residuals model
residuals_model <- lm(model_vote_dif_resid ~ model_pres_dif_resid)
```

```
================================================
Dependent variable:
---------------------------
model_vote_dif_resid
------------------------------------------------
model_pres_dif_resid      0.257*** (0.012)
Constant                   -0.000 (0.001)
------------------------------------------------
Observations                    3,193
R2                              0.130
Adjusted R2                     0.130
Residual Std. Error       0.073 (df = 3191)
F Statistic           476.975*** (df = 1; 3191)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

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

```
### 4.2 ###
# scatter plot
pdf("Q4_2_scatterplot.pdf")
plot( model_pres_dif_resid,
      model_vote_dif_resid,
      xlab = "Residuals from model_pres_dif", ylab = "Residuals from
   model_vote_dif",
      main = "Figure 4: Scatter plot of residuals")
abline(residuals_model, col="red", lty=2)
dev.off()
```

Figure 4: Scatter plot of residuals



3. Write the prediction equation.

$$\texttt{model\_pres\_dif\_residuals} = 0 + 0.257 \times \texttt{model\_vote\_dif\_residuals}$$

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`.

```
########### Question 5 ###########

### 5.1 ###
# outcome variable-voteshare, explanatory variables-difflog & presvote
model_vote_diff_pres <- lm(voteshare ~ difflog + presvote, data= inc.sub)
```

```
================================================
Dependent variable:
---------------------------
voteshare
------------------------------------------------
difflog                    0.036*** (0.001)
presvote                   0.257*** (0.012)
Constant                   0.449*** (0.006)
------------------------------------------------
Observations                     3,193
R2                               0.450
Adjusted R2                      0.449
Residual Std. Error        0.073 (df = 3190)
F Statistic           1,302.947*** (df = 2; 3190)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

2. Write the prediction equation.

$$voteshare = 0.449 + 0.036 \times difflog + 0.257 \times presvote$$

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

   This phenomenon occurs because the two coefficients actually describe the same relationship - the residual association between presvote and voteshare after controlling for difflog.