Problem Set 3 - My Answer

Junjie LIU[1,*]

20 November, 2023

[1] Department of Political Science, Trinity College Dublin, 2 Clare, Street, Dublin 2, Ireland

[*] Correspondence: Junjie LIU <liuj13@tcd.ie>

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.
- Total available points for this homework is 80.

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`.
2. Make a scatterplot of the two variables and add the regression line.
3. Save the residuals of the model in a separate object.
4. Write the prediction equation.

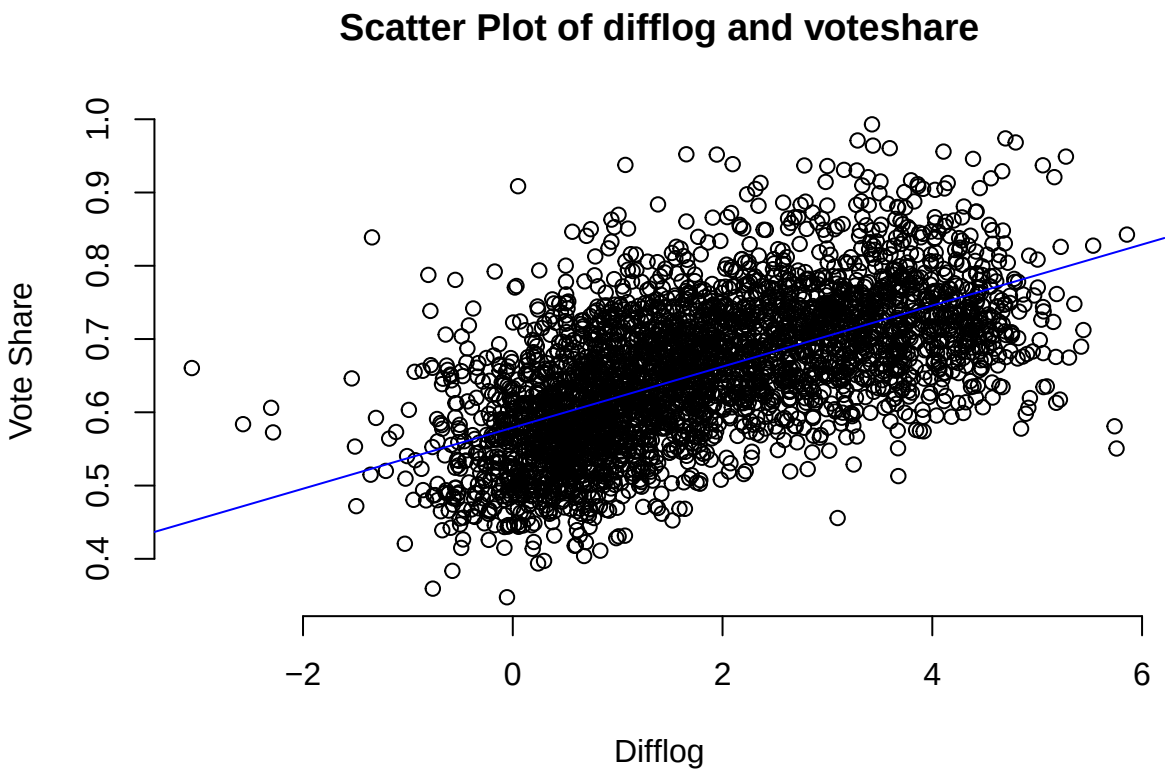
```
data <- read.csv("../../../datasets/incumbents_subset.csv")
model <- lm(data$voteshare~data$difflog)
print(model)
```

```
##
## Call:
## lm(formula = data$voteshare ~ data$difflog)
##
## Coefficients:
##  (Intercept)  data$difflog
##      0.57903       0.04167
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = data$voteshare ~ data$difflog)
##
## 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 ***
## data$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
```

```r
plot(data$difflog, data$voteshare, main = "Scatter Plot of difflog and voteshare",
     xlab = "Difflog", ylab = "Vote Share", frame = FALSE)
abline(lm(model), col = "blue")
```




```
residual <- resid(model)
```

Based on the result given by the regression model, we can find that there is a positive relation between `voteshare` and `difflog` and the slope is 0.04167 means one unit increase of `difflog` is related to 0.04167 increase in `voteshare`. The model can be noted as:

$$\hat{y} = 0.57903 + 0.04167 \cdot x$$

where x is `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. Make a scatterplot of the two variables and add the regression line.
3. Save the residuals of the model in a separate object.
4. Write the prediction equation.

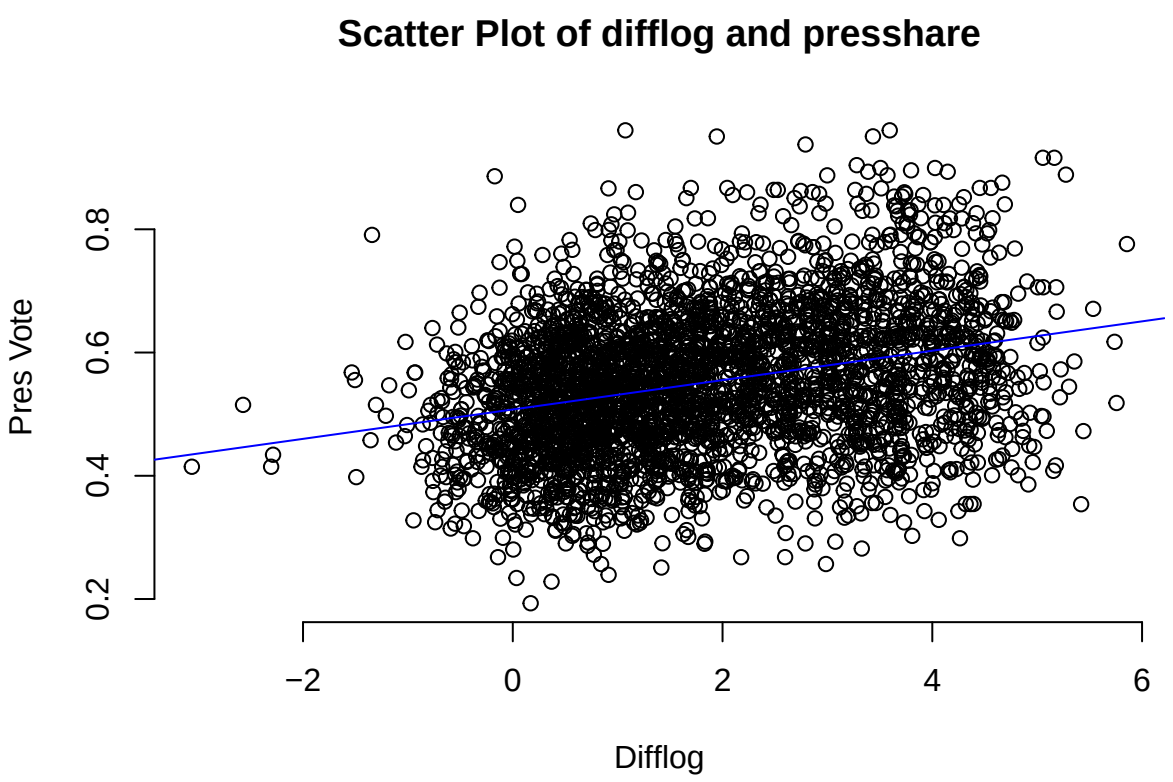
```
data <- read.csv("../../../datasets/incumbents_subset.csv")
model <- lm(data$presvote~data$difflog)
print(model)
```

```
##
## Call:
## lm(formula = data$presvote ~ data$difflog)
##
## Coefficients:
##  (Intercept)  data$difflog
##      0.50758       0.02384
```

```
summary(model)
```

```
##
## Call:
## lm(formula = data$presvote ~ data$difflog)
##
## 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 ***
## data$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
```

```
plot(data$difflog, data$presvote, main = "Scatter Plot of difflog and presshare",
     xlab = "Difflog", ylab = "Pres Vote", frame = FALSE)
abline(lm(model), col = "blue")
```




```
residual <- resid(model)
```

Based on the model's summary, we can see that there is a positive relationship between the `difflog` and `presvote`, where given one unit of `difflog` may leds `presvote` increases 0.02384.

$$\hat{y} = 0.50758 + 0.02384 \cdot x$$

where x is `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`.
2. Make a scatterplot of the two variables and add the regression line.
3. Write the prediction equation.

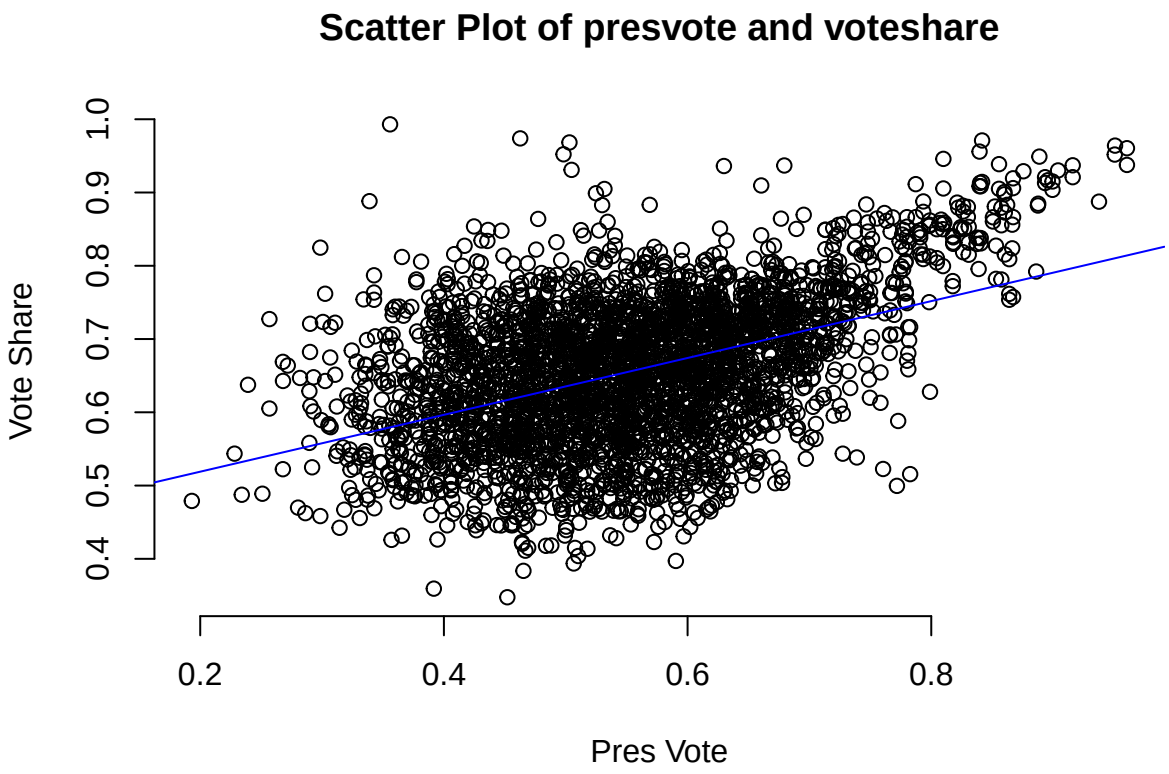
```
data <- read.csv("../../../datasets/incumbents_subset.csv")
model <- lm(data$voteshare~data$presvote)
print(model)
```

```
##
## Call:
## lm(formula = data$voteshare ~ data$presvote)
##
## Coefficients:
##   (Intercept)  data$presvote
##        0.4413         0.3880
```

```
summary(model)
```

```
##
## Call:
## lm(formula = data$voteshare ~ data$presvote)
##
## 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 ***
## data$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
```

```
plot(data$presvote, data$voteshare, main = "Scatter Plot of presvote and voteshare",
     xlab = "Pres Vote", ylab = "Vote Share", frame = FALSE)
abline(lm(model), col = "blue")
```




```
residual <- resid(model)
```

Based on the model's summary, we can see that there is a positive relationship between the `presvote` and `voteshare`, where given one unit of `presvote` may leds `voteshare` increases 0.388.

$$\hat{y} = 0.4413 + 0.3880 \cdot x$$

where x is `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.
2. Make a scatterplot of the two residuals and add the regression line.
3. Write the prediction equation.

```
data <- read.csv("../../../datasets/incumbents_subset.csv")
model_one <- lm(data$voteshare~data$difflog)
model_two <- lm(data$presvote~data$difflog)

model <- lm(resid(model_one)~resid(model_two))
print(model)
```

```
##
## Call:
## lm(formula = resid(model_one) ~ resid(model_two))
##
## Coefficients:
##      (Intercept)  resid(model_two)
##       -2.242e-20         2.569e-01
```

```
summary(model)
```

```
##
## Call:
## lm(formula = resid(model_one) ~ resid(model_two))
##
## 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)      -2.242e-20  1.299e-03    0.00        1
## resid(model_two)  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
```

```r
plot(resid(model_two), resid(model_one), main = "Scatter Plot of two Residual",
     xlab = "Residual of Model two", ylab = "Residual of Model one", frame = FALSE)
abline(lm(model), col = "blue")
```


```r
residual <- resid(model)
```

Based on the model's summary, we can see that there is a positive relationship between the residual of model one and residual of model one, where given one unit of residual of model two may leds residual of model one increases 0.2569.

$$\hat{y} = -2.242^{-20} + 0.2569 \cdot x$$

where x is residual of model two.

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?

```
data <- read.csv("../../../datasets/incumbents_subset.csv")
model <- lm(data$voteshare~(data$difflog+data$presvote))
print(model)
```

```
##
## Call:
## lm(formula = data$voteshare ~ (data$difflog + data$presvote))
##
## Coefficients:
##   (Intercept)   data$difflog  data$presvote
##       0.44864        0.03554        0.25688
```

```
summary(model)
```

```
##
## Call:
## lm(formula = data$voteshare ~ (data$difflog + data$presvote))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25928 -0.04737 -0.00121  0.04618  0.33126
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4486442  0.0063297   70.88   <2e-16 ***
## data$difflog  0.0355431  0.0009455   37.59   <2e-16 ***
## data$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
```

$$\hat{y} = 0.44864 + 0.03554 \cdot x_0 + 0.25688 \cdot x_1$$

where x_0 is `difflog` and x_1 is `presvote`. Based on the Q4 and Q5's result, we can find that the coefficient of $x1$ in Q5's model and x in Q4's model are approximately equal. In other words, the co-variation between `presvote` and `voteshare` cannot been explained by the `difflog`.