

# 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.

First i want to apologise because my scatterplots are all over the place. I am aware I might lose points for it but I tried many ways to input them and they kept showing up in random places. I am sorry because it looks messy.

### 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`.

Code for the regression and the summary to be able to see the results.

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

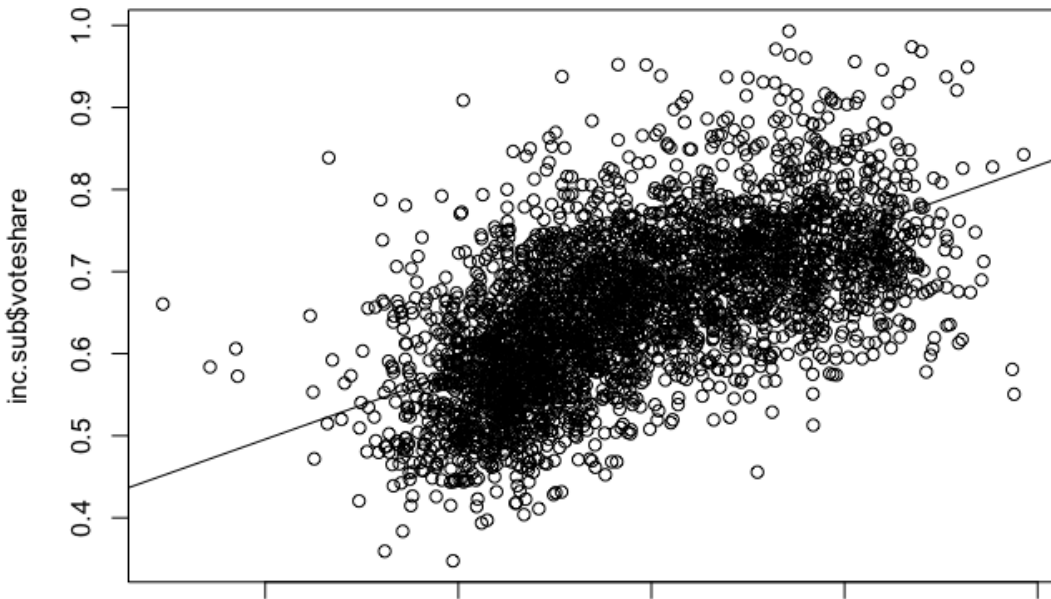


Figure 1: Voteshare X Difflog

Output:

Residuals:

Min -0.26832 1Q -0.05345 Median -0.00377 3Q 0.04780 Max 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 ***

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.

Code for plotting the two variables and then adding the regression line.

```

1 plot(x=inc.sub$difflog, y=inc.sub$voteshare)
2 abline(model)
3

```

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

I saved the residuals from the first model as `residuals_1`  
 I also printed just to make sure I had saved them.

```
1 residuals_1 <- residuals(model)
2 print(residuals_1)
3
```

4. Write the prediction equation.

Regression equation for  $x$  being the variable *difflog*,  $b_0$  the intercept and  $b_1$  the slope as seen in the output from the model's summary.

$$\hat{y} = b_0 + b_1x \quad (1)$$

$$\hat{y} = 0.58 + 0.04 * \textit{difflog} \quad (2)$$

## 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`.

Code for the regression and the summary to be able to see the results.

```
1 model_2 <- lm(presvote~difflog , data=inc.sub)
2 summary(model_2)
3
```

Residuals:

Min -0.32196 1Q-0.07407 Median-0.00102 3Q 0.07151 Max 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

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.

Code for plotting the two variables and then adding the regression line.

```
1 plot(x=inc.sub$difflog , y= inc.sub$presvote)
2 abline(model_2)
3
```

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

I saved the residuals of `model_2` as `residuals_2`

I also printed just to make sure I had saved them.

```
1 residuals_2 <- residuals(model_2)
2 print(residuals_2)
3
```

4. Write the prediction equation.

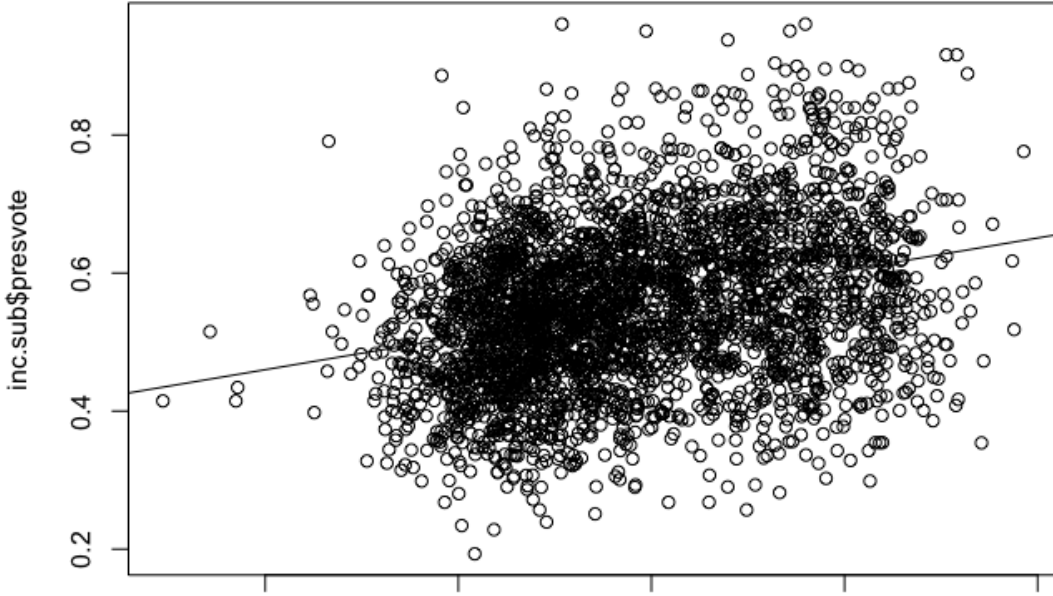


Figure 2: Presvote X Difflog

Regression equation for x being the variable difflog, b0 the intercept and b1 the slope as seen in the output from the model\_2's summary.

$$\hat{y} = b_0 + b_1x \quad (3)$$

$$\hat{y} = 0.50 + 0.02 * difflog \quad (4)$$

## 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`.

Code for the regression and the summary to be able to see the results.

```
1 model_3 <- lm(voteshare~presvote , data=inc.sub)
2 summary(model_3)
3
```

Residuals:

Min -0.27330 1Q -0.05888 Median 0.00394 3Q 0.06148 Max 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 ***

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.

Code for plotting the two variables and then adding the regression line.

```
1 plot(x=inc.sub$presvote , y=inc.sub$voteshare)
2 abline(model_3)
3
```

3. Write the prediction equation.

Regression equation for  $x$  being the variable `presvote`,  $b_0$  the intercept and  $b_1$  the slope as seen in the output from the `model_3`'s summary.

$$\hat{y} = b_0 + b_1x \quad (5)$$

$$\hat{y} = 0.44 + 0.38 * presvote \quad (6)$$

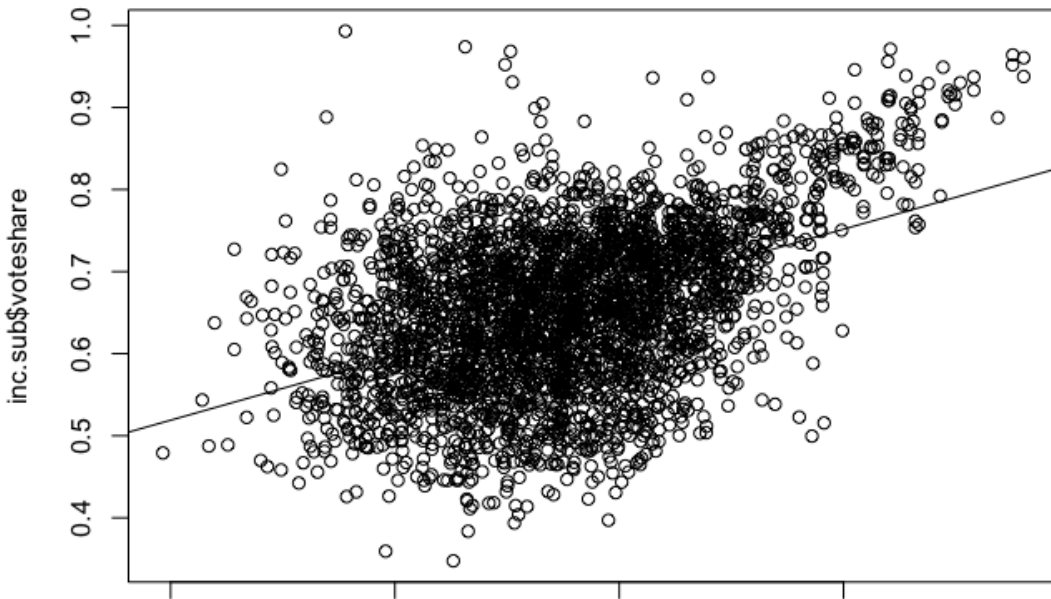


Figure 3: voteshare X presvote

## 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.

Code for the regression using the saved objects with residuals from the first model and `model_2` and the summary to be able to see the results.

```
1 model_4 <- lm(residuals_1~residuals_2)
2 summary(model_4)
3
```

Residuals:

Min -0.25928 1Q -0.04737 Median -0.00121 3Q 0.04618 Max 0.33126

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.942e-18	1.299e-03	0.00	1

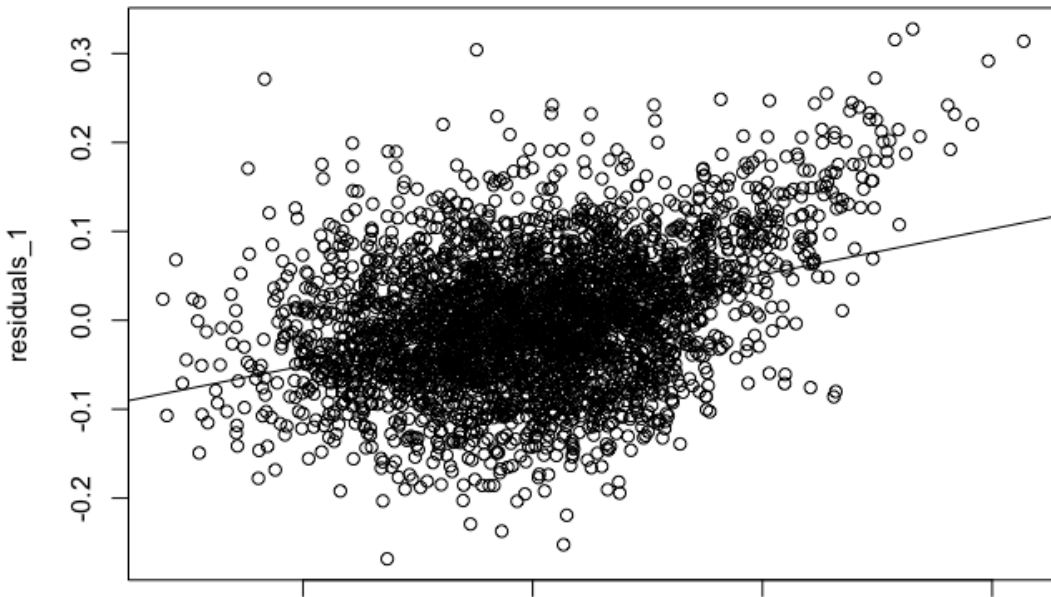


Figure 4: residuals model 1 X residuals model 2

```
residuals_2  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.

Code for plotting the two residuals and then adding the regression line.

```
1 plot(x=residuals_2 , y=residuals_1)
2 abline(model_4)
3
```

3. Write the prediction equation.

Regression equation for x being residuals\_2, b0 the intercept and b1 the slope as seen in the output from the model\_4's summary.

$$\hat{y} = b_0 + b_1 x \quad (7)$$

$$\hat{y} = -1.94 + 2.56 * residuals_2 \quad (8)$$



## 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`.

Code for the regression and the summary to be able to see the results.

```
1 model_5 <- lm(voteshare~difflog+presvote ,data=inc.sub)
2 summary(model_5)
3
```

Residuals:

Min -0.25928 1Q -0.04737 Median -0.00121 3Q 0.04618 Max 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 ***

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.

Regression equation for  $x_1$  being the variable `difflog` and  $x_2$  being the variable `presvote`,  $b_0$  the intercept,  $b_1$  the slope for `difflog` and  $b_2$  the slope for `presvote` as seen in the output from the `model_5`'s summary.

$$\hat{y} = b_0 + b_1 * difflog + b_2 * presvote \quad (9)$$

$$\hat{y} = 0.45 + 0.03 * difflog + 0.25 * presvote \quad (10)$$

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 two models have the same residual standard error which means that they both have the same level of fitness to the dataset, it could indicate multicollinearity.