

Problem Set 3

Applied Stats/Quant Methods 1

Due: November 20, 2022

Question 1

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

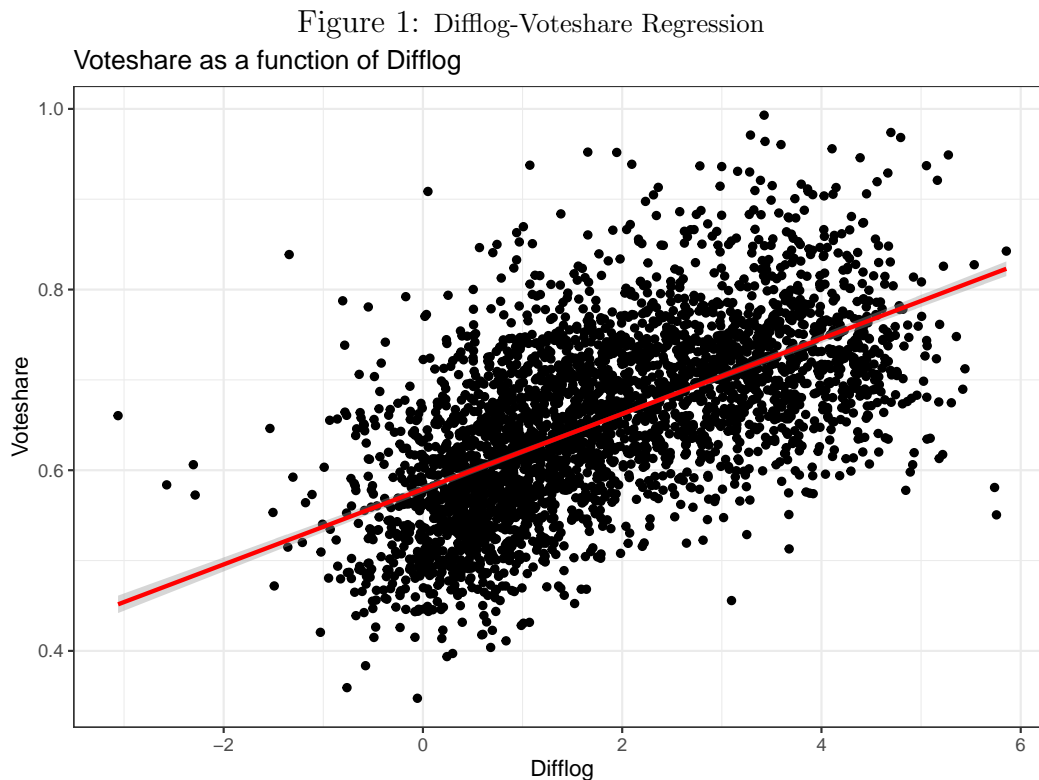
```
1 Diff_Vote_Reg <- lm(voteshare ~ difflog, data = inc.sub) # fit the model
2
3 summary(Diff_Vote_Reg)
4
5 stargazer(Diff_Vote_Reg, type = "latex", out = "Diff_Vote_Reg.tex",
6           title = "Difflog-Voteshare Regression")
```

Table 1: Difflog-Voteshare Regression

	<i>Dependent variable:</i>
	voteshare
difflog	0.042*** (0.001)
Constant	0.579*** (0.002)
Observations	3,193
R ²	0.367
Adjusted R ²	0.367
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,852.791*** (df = 1; 3191)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

```
1 ggplot(inc.sub, aes(x = difflog , y = voteshare)) +  
2   geom_point() +  
3   geom_smooth(method = "lm", color = "red") +      # regression line  
4   theme_bw() +  
5   labs(title = "Voteshare as a function of Difflog",  
6         x = "Difflog",  
7         y = "Voteshare")
```



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

```
1 inc.sub$residuals <- residuals(Diff_Vote_Reg) # Save the residual values  
2 Voteshare_res <- inc.sub$residuals
```

4. Write the prediction equation.

The estimated regression line equation can be written as follows:

$$\text{voteshare} = 0.58 + 0.04 \cdot \text{difflog}$$

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`.

```
1 Diff_Presvote_Reg <- lm(presvote ~ difflog, data = inc.sub) # fit the
  model
2
3 summary(Diff_Presvote_Reg)
4
5 stargazer(Diff_Presvote_Reg, type = "latex\\vspace{.25cm}", out = "Diff_
  Presvote_Reg.latex", title =
6           "Difflog-Presvote Regression")
```

Table 2: Difflog-Presvote Regression

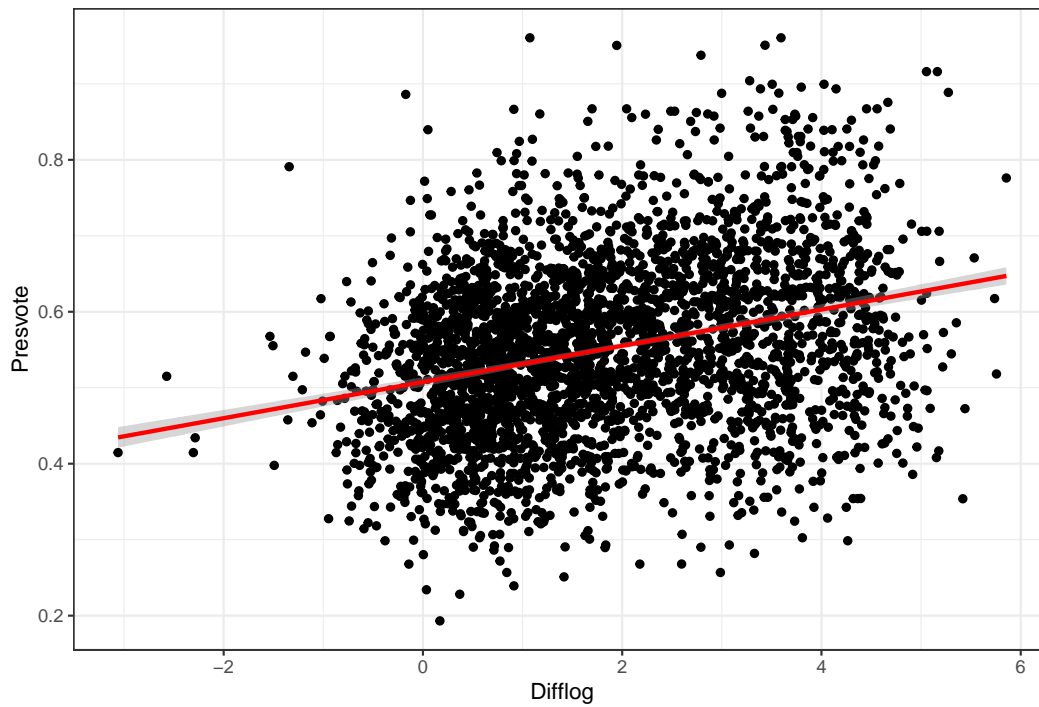
	<i>Dependent variable:</i>
	presvote
difflog	0.024*** (0.001)
Constant	0.508*** (0.003)
Observations	3,193
R ²	0.088
Adjusted R ²	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

```
1 ggplot(inc.sub, aes(x = difflog, y = presvote)) +
2   geom_point() +
3   geom_smooth(method = "lm", color = "red") + # regression line
4   theme_bw() +
5   labs(title = "Presvote as a function of Difflog",
6        x = "Difflog",
7        y = "Presvote")
```

Figure 2: Difflog-Presvote Regression

Presvote as a function of Difflog



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

```
1 inc.sub$residuals <- residuals(Diff_Presvote_Reg) # Save the residual  
  values  
2 Presvote_res <- inc.sub$residuals
```

4. Write the prediction equation.

The estimated regression line equation can be written as follows:

$$\text{presvote} = 0.51 + 0.02 \cdot \text{difflog}$$

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.**

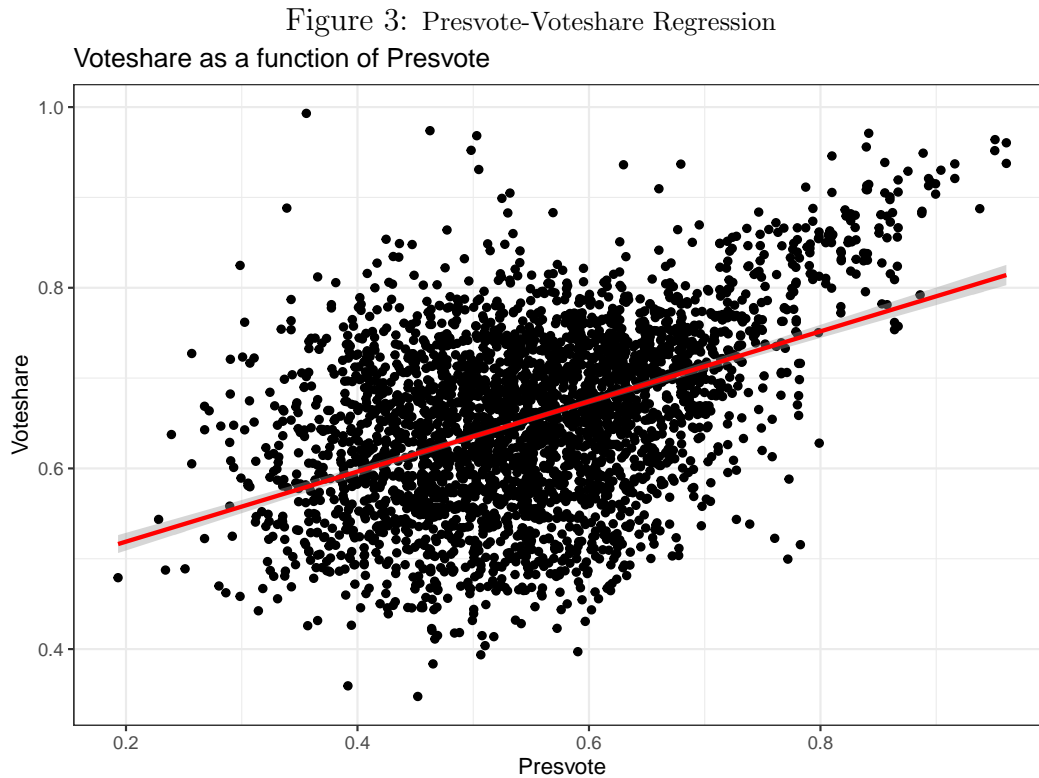
```
1 Presvote_Voteshare_Reg <- lm(voteshare ~ presvote, data = inc.sub) # fit
  the model
2
3 summary(Presvote_Voteshare_Reg)
4
5 stargazer(Presvote_Voteshare_Reg, type = "latex", out = "Presvote_
  Voteshare_Reg.tex", title =
6           "Presvote-Voteshare Regression")
```

Table 3: Presvote-Voteshare Regression

	<i>Dependent variable:</i>
	voteshare
presvote	0.388*** (0.013)
Constant	0.441*** (0.008)
Observations	3,193
R ²	0.206
Adjusted R ²	0.206
Residual Std. Error	0.088 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

```
1 ggplot(inc.sub, aes(x = presvote, y = voteshare)) +  
2   geom_point() +  
3   geom_smooth(method = "lm", color = "red") +      # regression line  
4   theme_bw() +  
5   labs(title = "Voteshare as a function of Presvote",  
6         x = "Presvote",  
7         y = "Voteshare")
```



3. Write the prediction equation.

The estimated regression line equation can be written as follows:

$$\text{voteshare} = 0.44 + 0.39 \cdot \text{presvote}$$

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.

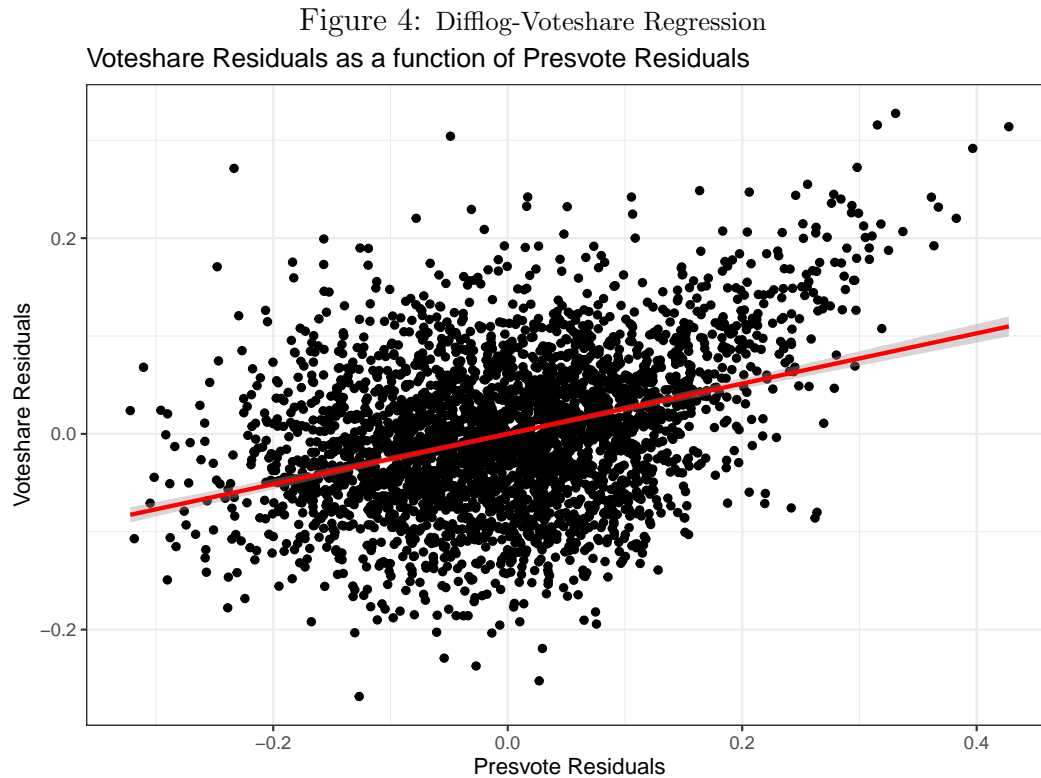
```
1 Q1Resid_Q2Resid_Reg <- lm(Voteshare_res ~ Presvote_res, data = inc.sub) #
   fit the model
2
3 summary(Q1Resid_Q2Resid_Reg)
4
5 stargazer(Q1Resid_Q2Resid_Reg, type = "latex", out = "Q1Resid_Q2Resid_Reg
   .latex", title =
6 "Q1 Residuals-Q2 Residuals Regression")
```

Table 4: Q1 Residuals-Q2 Residuals Regression

	Dependent variable:
	Voteshare_res
Presvote_res	0.257*** (0.012)
Constant	-0.000 (0.001)
Observations	3,193
R ²	0.130
Adjusted R ²	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.

```
1 ggplot(inc.sub, aes(x = Presvote_res, y = Voteshare_res)) +  
2   geom_point() +  
3   geom_smooth(method = "lm", color = "red") +      # regression line  
4   theme_bw() +  
5   labs(title = "Voteshare Residuals as a function of Presvote Residuals",  
6         x = "Presvote Residuals",  
7         y = "Voteshare Residuals")
```



3. Write the prediction equation.

The estimated regression line equation can be written as follows:

$$Q1Resid = -4.860e-18 + 2.569e-01 * Q2Resid$$

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 Diff_Presvote_Voteshare_Reg <- lm(voteshare ~ difflog + presvote, data =
  inc.sub) # fit the model
2
3 summary(Diff_Presvote_Voteshare_Reg)
4
5 stargazer(Diff_Presvote_Voteshare_Reg, type = "latex", out = "Q5.latex",
  title =
6           "Difflog & Presvote - Voteshare Regression")
```

Table 5: Difflog and Presvote - Voteshare Regression

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

2. **Write the prediction equation.**

The estimated regression line equation can be written as follows:

$$\text{voteshare} = 0.45 + (0.04 \cdot \text{difflog}) + (0.26 \cdot \text{presvote})$$

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

The Residual Std. Error in the output for this question is identical to the Residual Std. Error in the output for Q4.

This is because running a regression of the residuals of $\text{voteshare} \sim \text{difflog}$ against the residuals of $\text{presvote} \sim \text{difflog}$ as we do in Q4, tells us how much of the unexplained variation in voteshare is influenced by presvote .

In Q5, we are essentially showing the same thing in a different way, by running a regression of voteshare against difflog AND presvote , we can see how much of the unexplained variance in $\text{voteshare} \sim \text{difflog}$ is explained by presvote , and this means that the Residual Std. Error will be the same.