

Problem Set 3

Applied Stats/Quant Methods 1
Laura McPhillips

Due: November 12, 2021

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**.

Y-intercept = 0.579. When difflog is 0, voteshare is 0.579.

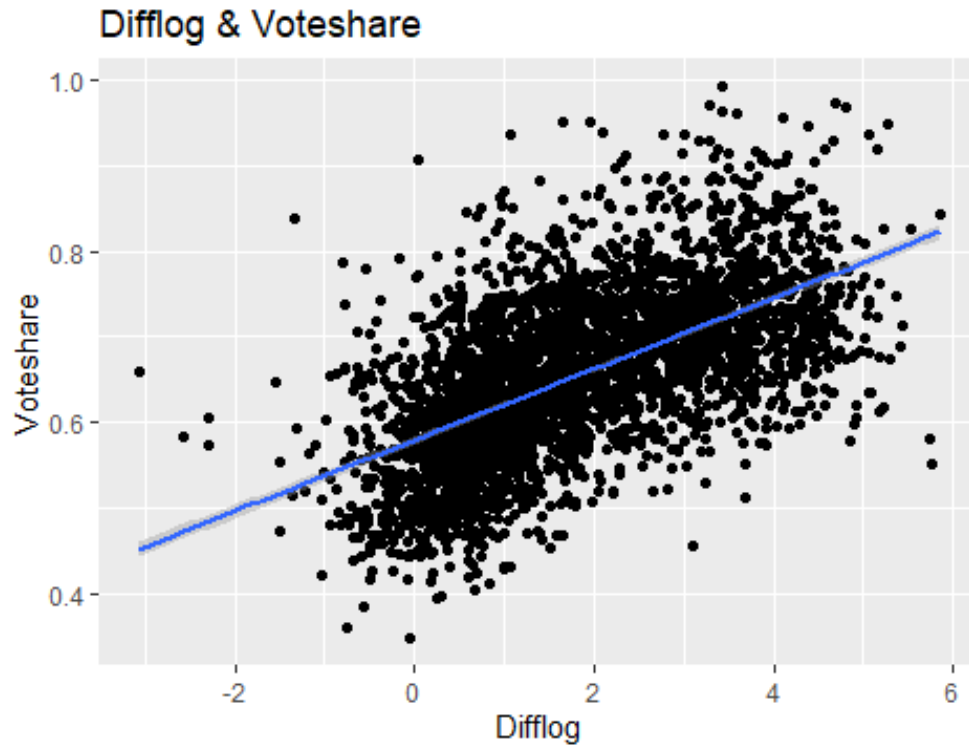
Co-efficient Estimate = 0.0416. For every 1 unit increase in difflog, there is a 0.04.26 unit increase in voteshare.

```
1 campaign <- lm(incumbents$voteshare ~ incumbents$difflog)
2 campaign
```

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

```
1 ggplot(incumbents, aes(x= difflog, y=voteshare)) +
2   geom_point() +
3   geom_smooth(method=lm, formula = y~x) +
4   ggtitle("Difflog & Voteshare") +
5   labs(y = "Voteshare") +
6   labs(x = "Difflog")
```

Figure 1: Difflog and Voteshare



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

```
1 campaign_resid <- resid(campaign)
2 campaign_resid
```

4. Write the prediction equation. $y=mx+c$

```
1 voteshare_predict <- 0.0416*difflog + 0.579
```

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`.

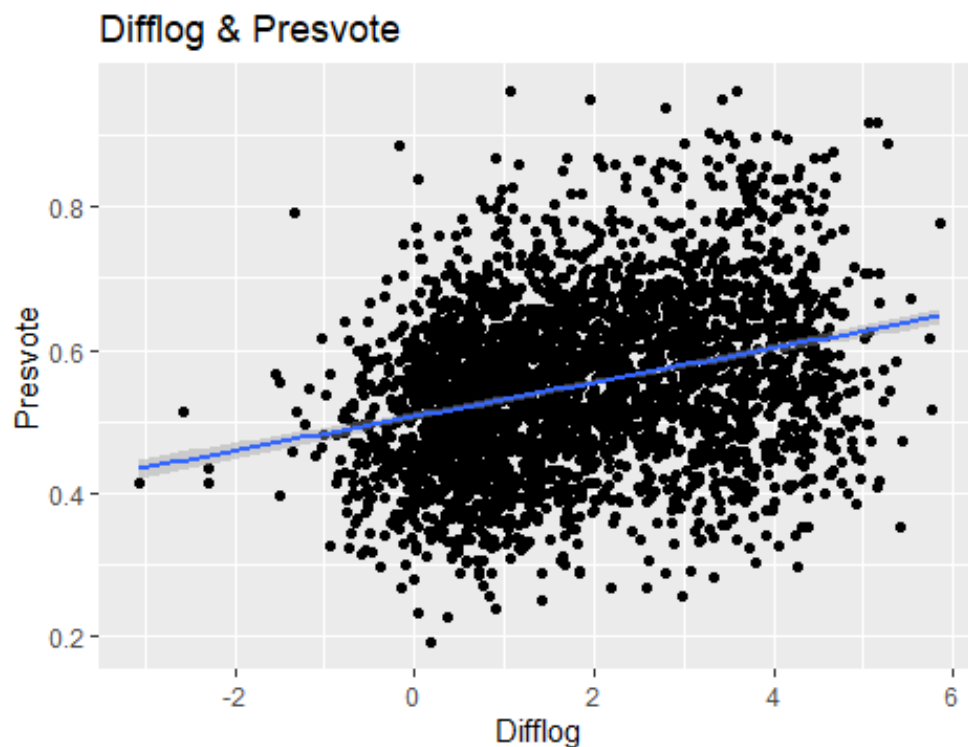
Y-intercept is 0.507. When `difflog` is 0, `presvote` is 0.507.

Coefficient-estiamte is 0.0238. For every unit increase in `difflog`, `presvote` increases by 0.0238.

```
1 vote <- lm(incumbents$presvote ~ incumbents$difflog)
2 vote
```

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

Figure 2: Difflog and Presvote



```

1 ggplot(incumbents, aes(x= difflog, y=presvote)) +
2   geom_point() +
3   geom_smooth(method=lm, formula = y ~ x)+
4   ggtitle("Difflog & Presvote") +
5   labs(y = "Presvote") +
6   labs(x = "Difflog")
7

```

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

```

1 vote_resid <- resid(vote)
2 vote_resid

```

4. Write the prediction equation.

$$y = mx + c$$

```

1 presvote_predict <- 0.023*difflog + 0.507

```

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**.

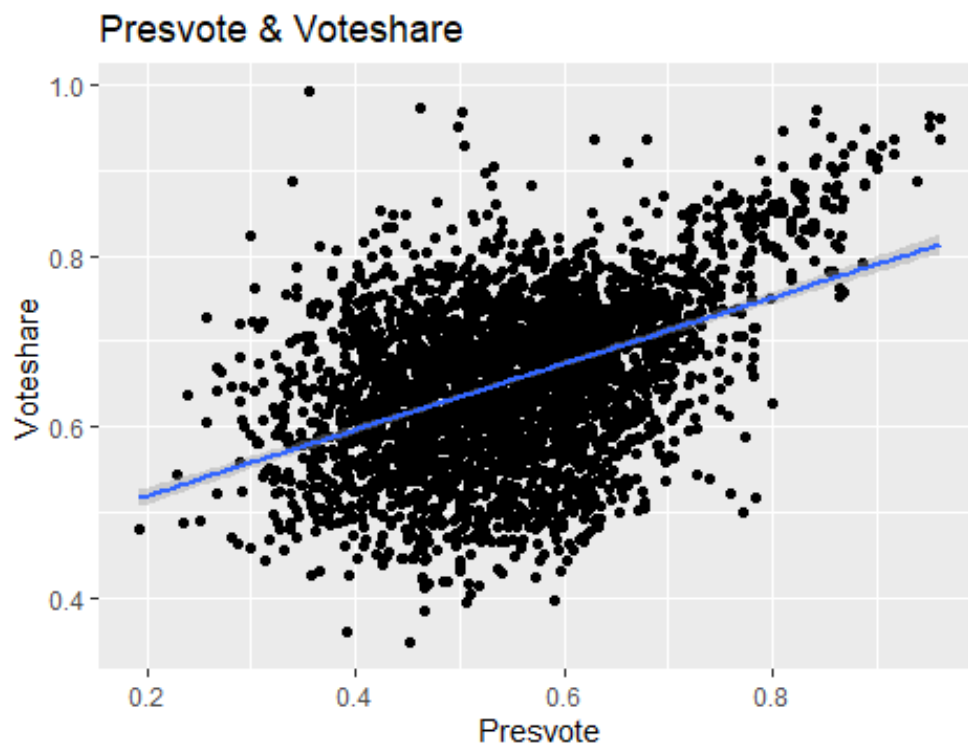
Y-intercept = 0.4413. When **presvote** is 0, **voteshare** is 0.4413.

Co-efficient estimate is 0.388. For every unit increase in **presvote**, there is a 0.388 increase in **voteshare**.

```
1 voteshare <- lm(incumbents$voteshare ~ incumbents$presvote)
2 voteshare
```

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

Figure 3: Difflog and Voteshare



```

1 ggplot(incumbents, aes(x=presvote, y=votesshare)) +
2   geom_point() +
3   geom_smooth(method=lm, formula = y ~ x)+
4   ggtitle("presvote & Voteshare") +
5   labs(y = "Voteshare") +
6   labs(x = "Presvote")

```

3. Write the prediction equation.

$y=mx+c$

```

1 voteshare_predict <- 0.38*presvote + 0.44

```

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.

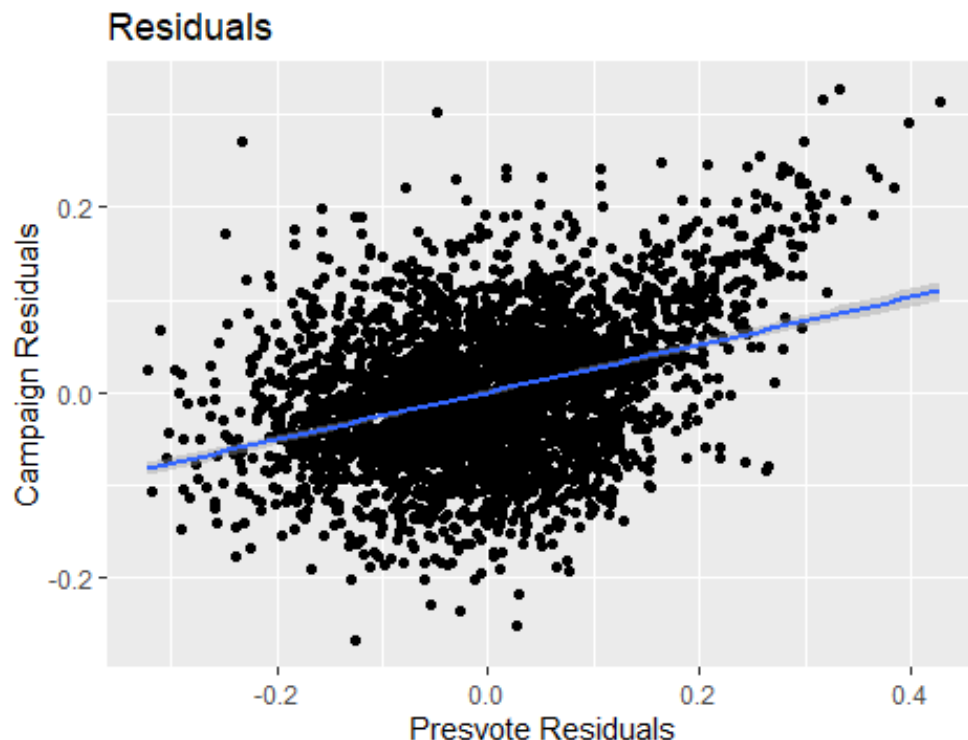
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

Y-intercept is -4.860e-18. When Vote Residuals is 0, Campaign Residuals is -4.860e-18. Coefficient-estimate is 2.569e-01. For every unit increase in Vote Residuals, Campaign residuals increases by 2.569e-01.

```
1 resid_regress <- lm(campaign_resid ~ vote_resid, data = incumbents)
2 resid_regress
```

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

Figure 4: Residuals



```

1 ggplot(incumbents, aes(x=vote_resid, y=campaign_resid)) +
2   geom_point() +
3   geom_smooth(method=lm, formula = y ~ x)+
4   ggtitle("Residuals") +
5   labs(y="Campaign Residuals") +
6   labs(x="Presvote Residuals")

```

3. Write the prediction equation.

$y=mx+c$

```

1 resid_predict <- 2.569e-01*vote_resid + -4.860e-18
2

```


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`.

Y-intercept is 0.448. When `difflog` and `presvote` are 0, `voteshare` is 0.448 .

Coefficient estimates are 0.03554 and 0.25688 .

For every unit increase in `difflog`, `voteshare` increases by 0.3554.

For every unit increase in `presvote`, `voteshare` increases by 0.25688.

```
1 multivar <- lm(voteshare ~ difflog + presvote, data = incumbents)
2 multivar
```

2. Write the prediction equation.

$y = mx_1 + mx_2 + c$

```
1 multivar$predict <- 0.035*difflog + 0.256*presvote + 0.448
```

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 residuals from Question 4 and Question 5 are the same. This is because the residuals account for what is not explained by the explanatory variable.

So, the variation when the residuals are plotted against each other together, will be the same as the variation when the two variables are plotted against each other. The variation remains the same.