

Problem Set 3

Doireanna Craven

Due: November 12, 2021

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`.

```
1 # Regression of voteshare on difflog
2 voteshare_difflog <- lm(voteshare ~ difflog, data = incumbents)
3 summary(voteshare_difflog)
```

Analysis: Every \$1m spent in excess of their opponent is associated with a 4% increase in the incumbent's voteshare. The R-squared value tells us that 37% of the variability in voteshare is associated with the difference in campaign spending between incumbent and challenger. This does not seem high, but the results are statistically significant.

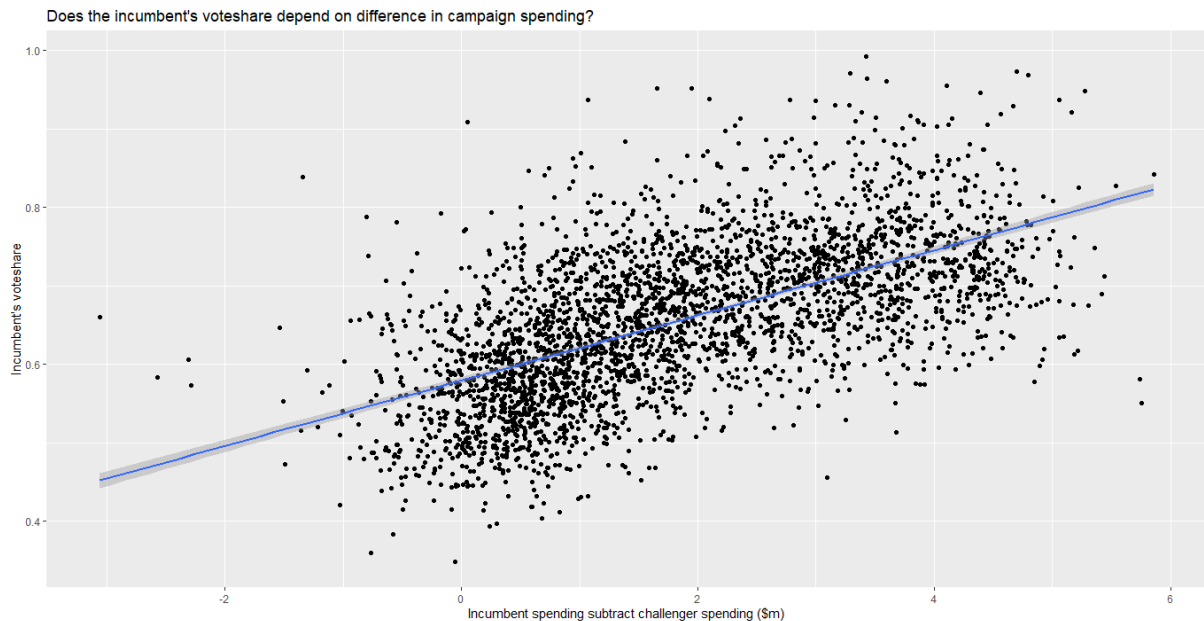
2. Make a scatterplot of the two variables and add the regression line.

```
1 # Scatterplot
2 ggplot(aes(difflog, voteshare), data = incumbents) +
3   geom_point() +
4   geom_smooth(method = "lm", formula = y ~ x) +
```

```

5 labs(title = "Does the incumbent's voteshare depend on difference in
6     campaign spending?",
7     x = "Incumbent spending subtract challenger spending ($m)",
8     y = "Incumbent's voteshare")

```



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

```

1 # Save the residuals as a separate object
2 res <- matrix(residuals(voteshare_difflog), ncol=1)

```

I have saved the residuals as a matrix. I can this as a column to the `incumbents` dataframe later.

4. Write the prediction equation.

$$V = 0.579 + 0.0417D,$$

where V = incumbent's voteshare and D = incumbent's spend - challenger's spend (\$m).

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.

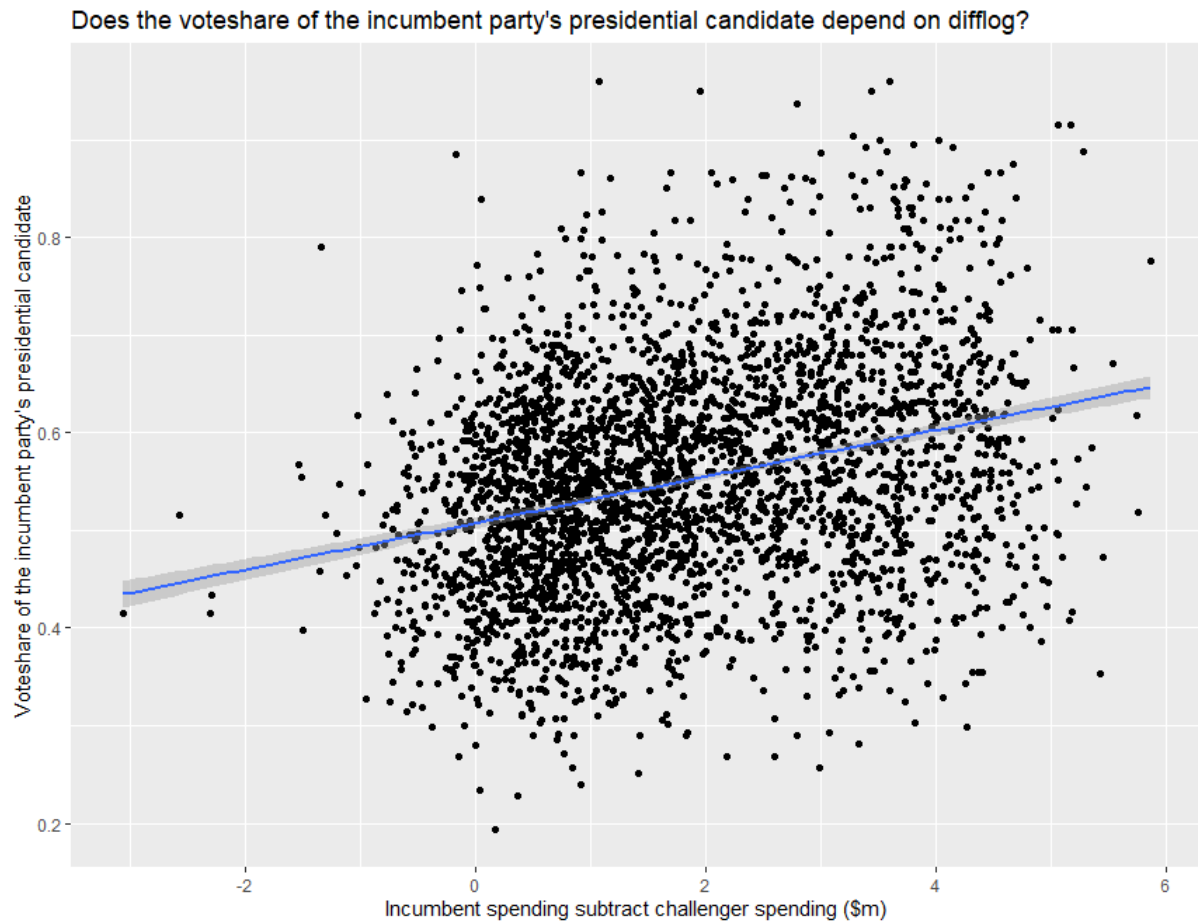
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```
1 # Regression of incumbent party's presidential candidate's voteshare on
  difflog
2 presvote ~ difflog <- lm(presvote ~ difflog, data = incumbents)
3 summary(presvote ~ difflog)
```

Analysis: Every \$1m spent by the incumbent in excess of their opponent is associated with a 2.4% increase in their presidential candidate's voteshare. While only 9% of the variation in voteshare is associated with the difference in spend, the results are statistically significant.

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

```
1 # Scatterplot
2 ggplot(aes(difflog, presvote), data = incumbents) +
3   geom_point() +
4   geom_smooth(method = "lm", formula = y ~ x) +
5   labs(title = "Does the voteshare of the incumbent party's presidential
  candidate depend on difflog?",
6         x = "Incumbent spending subtract challenger spending ($m)",
7         y = "Voteshare of the incumbent party's presidential candidate")
```



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

```
1 # Save the residuals as a separate object
2 res2 <- matrix(residuals(presvote_difflog), ncol=1)
```

4. Write the prediction equation.

$$P = 0.508 + 0.0238D,$$

where P = voteshare of incumbent party's presidential candidate and D = incumbent's spend - challenger's spend (\$m).

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.

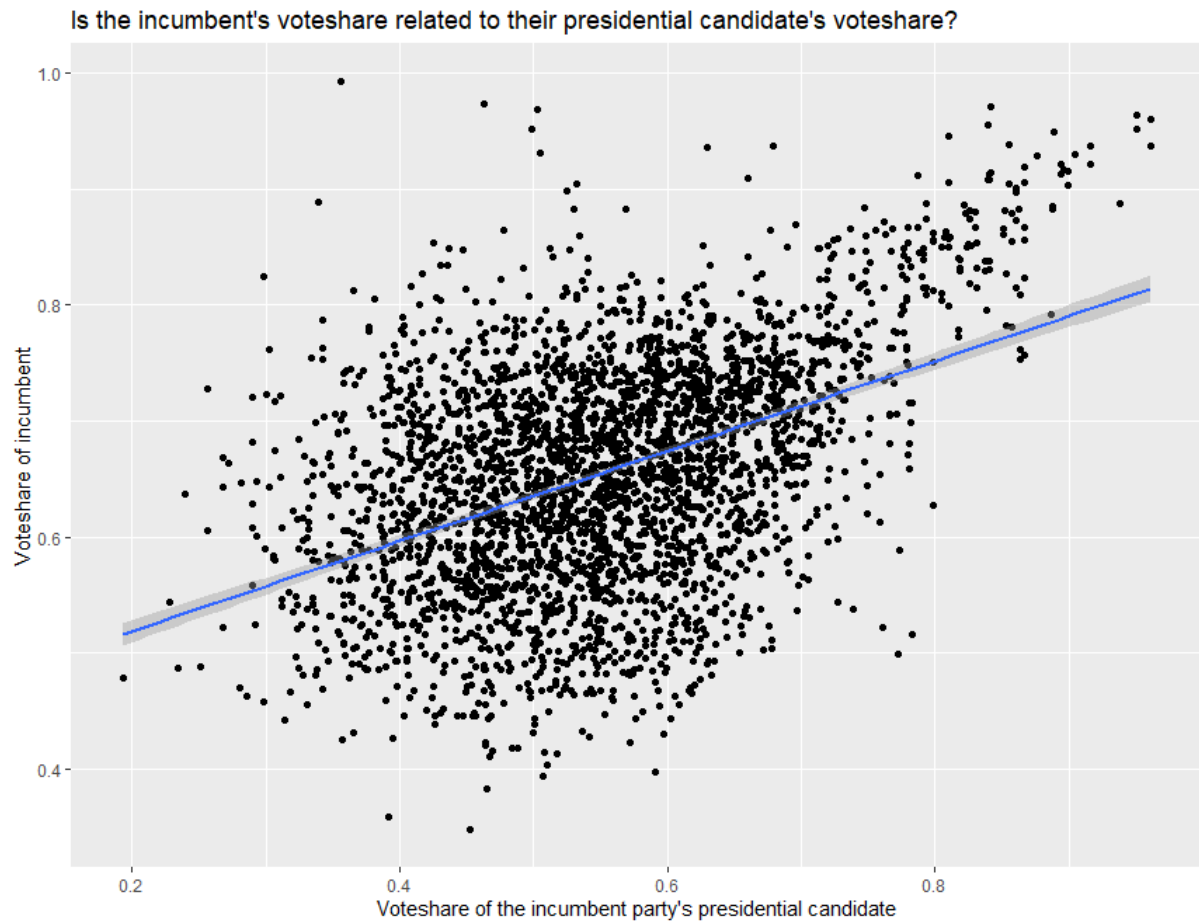
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`.

```
1 # Regression of incumbent party's presidential candidate's voteshare on
  difflog
2 voteshare_presvote <- lm(voteshare ~ presvote, data = incumbents)
3 summary(voteshare_presvote)
```

Analysis: A 1% increase in the voteshare of the incumbent's presidential candidate is associated with a 0.4% increase in the incumbent's voteshare. 21% of the variation can be explained by this relationship, and the results are statistically significant.

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

```
1 # Scatterplot
2 ggplot(aes(presvote, voteshare), data = incumbents) +
3   geom_point() +
4   geom_smooth(method = "lm", formula = y ~ x) +
5   labs(title = "Is the incumbent's voteshare related to their presidential
  candidate's voteshare?",
6         x = "Voteshare of the incumbent party's presidential candidate",
7         y = "Voteshare of incumbent")
```



3. Write the prediction equation.

$$V = 0.441 + 0.388P,$$

where V = voteshare of incumbent and P = voteshare of incumbent party's presidential candidate.

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.

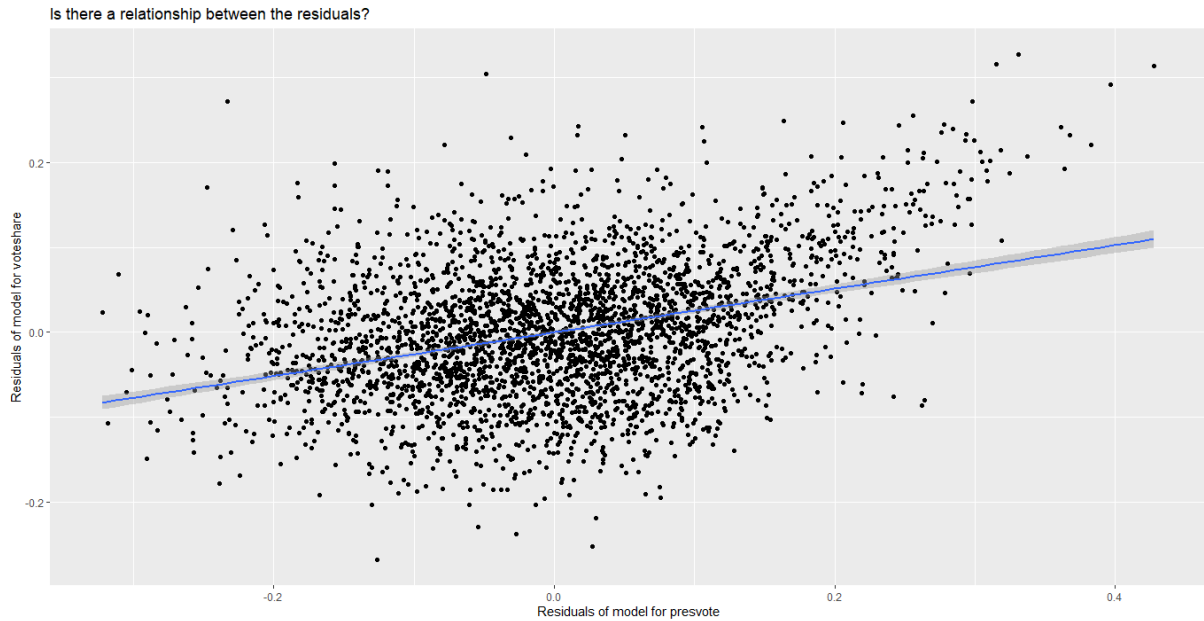
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
1 # Regression of res on res2
2 # Add the residuals to the incumbents df
3 incumbents_res <- cbind(incumbents, res, res2)
4 res_res2 <- lm(res ~ res2, data = incumbents)
5 summary(res_res2)
```

There is a statistically significant association between the incumbent's voteshare unexplained by spending, and their presedential candidate's voteshare unexplained by spending.

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

```
1 # Scatterplot
2 ggplot(aes(res2, res), data = incumbents) +
3   geom_point() +
4   geom_smooth(method = "lm", formula = y ~ x) +
5   labs(title = "Is there a relationship between the residuals?",
6         x = "Residuals of model for presvote",
7         y = "Residuals of model for voteshare")
```



3. Write the prediction equation.

The y-intercept is practically zero. $Y = 0.257X$,

where Y = residuals from spending model for incumbent's voteshare and X = residuals from spending model for voteshare of incumbent party's presidential candidate. This is the effect of `presvote` on `voteshare` after adjusting for `difflog`.

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 # Regression of voteshare on difflog and presvote
2 voteshare_presvote_and_difflog <- lm(voteshare ~ difflog + presvote, data
   = incumbents)
3 summary(voteshare_presvote_and_difflog)
```

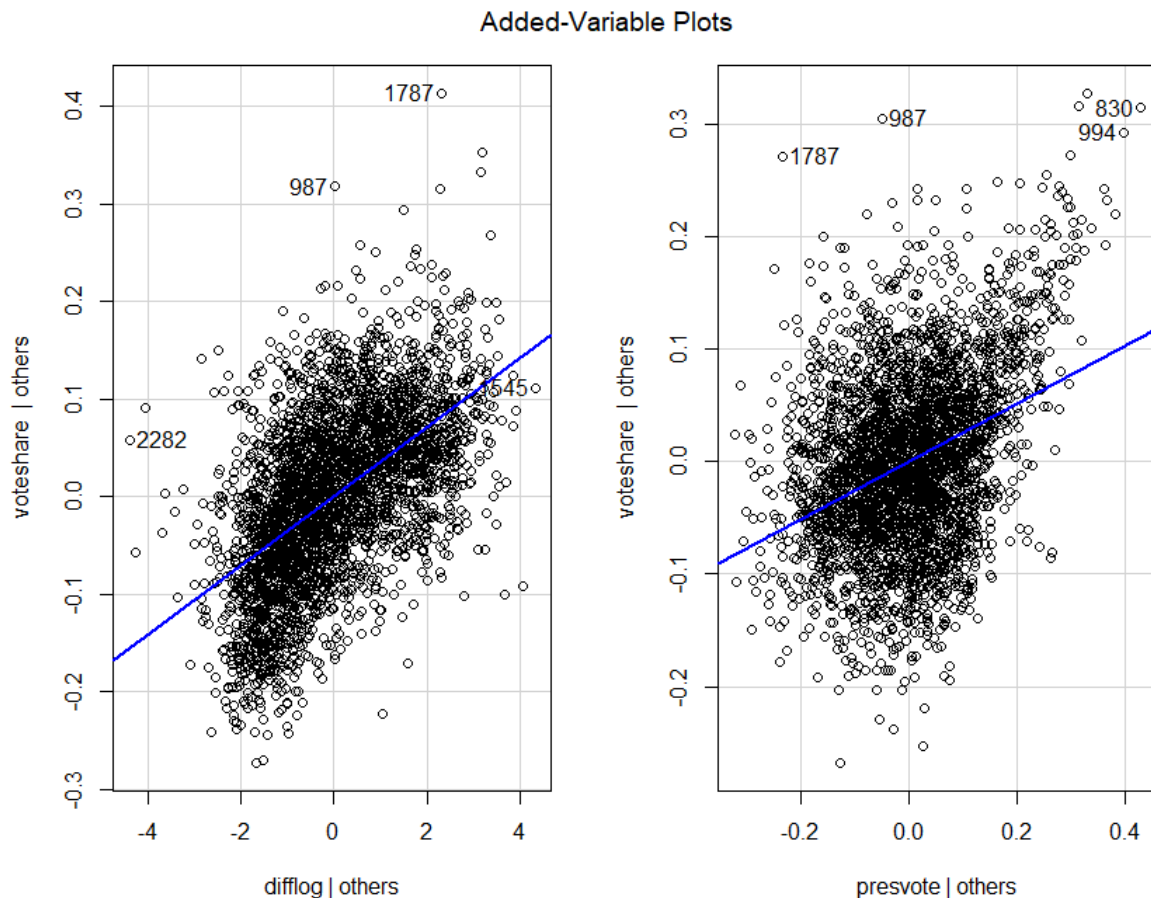
2. Write the prediction equation.

$$V = 0.449 + 0.0355D + 0.257P,$$

where V = incumbent's voteshare, D = incumbent's spend - challenger's spend (\$m), and P = voteshare of incumbent's presidential candidate.

The added variable plots below show how `voteshare` varies with `difflog` when other variables are held constant, and `presvote` when other variables are held constant.

```
1 avPlots(voteshare_presvote_and_difflog)
```



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

	β_0	β_1	β_2
Model Q1: $V =$	0.579 +	0.0417D	
Model Q3: $V =$	0.441 +		+ 0.388P
Model Q5: $V =$	0.449 +	0.0355D	+ 0.257P

β_2 in Q5 is equivalent to the slope in Q4.

The partial regression coefficient of **presvote** tells us that each unit of **presvote** adds 0.257 to **voteshare** when **difflog** is held fixed. This matches the ratio of residuals found in Q4, which is the effect of **presvote** on **voteshare** after adjusting for **difflog**.