Problem Set 3

Luke Duggan, 16316834

November 11, 2021

Question 1

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

   The R code for this regression is the first line of the following:

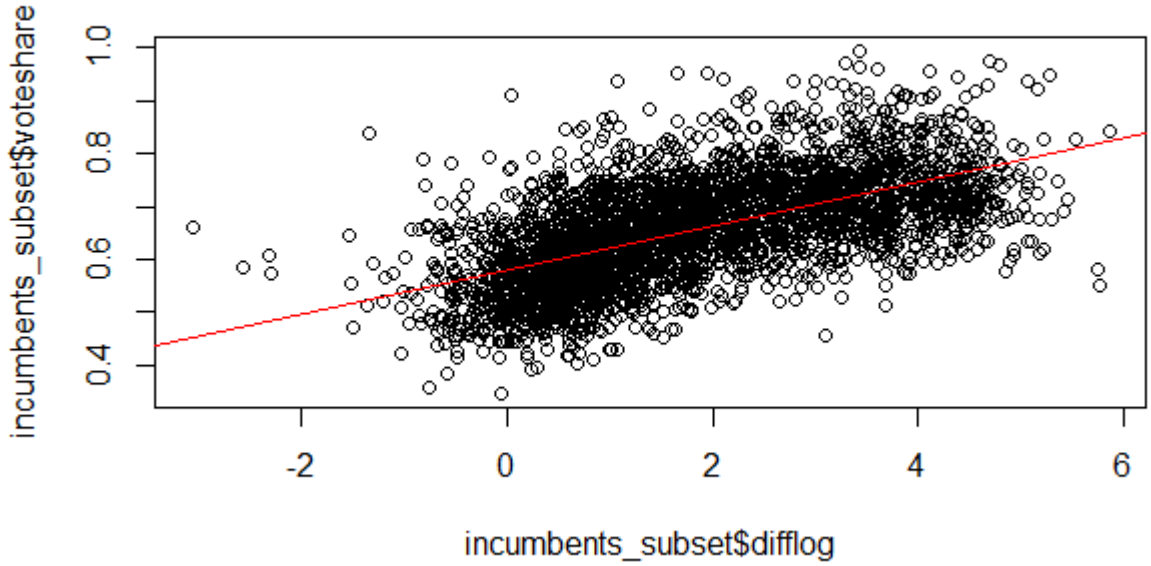
   ```
   results <- lm(voteshare ~ difflog, data = incumbents_subset)
   coef(results)
   ```

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

   The R code is as follows:

   ```
   plot(incumbents_subset$difflog, incumbents_subset$voteshare)
   abline(lm(voteshare ~ difflog, data = incumbents_subset), col = "red")
   ```

The resulting scatterplot with the regression line is:



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

   The R code is:

```
residuals <- resid(results)
```

4. Write the prediction equation.

   Using the coefficients obtained in part 1 (and rounding to two decimal places), the prediction equation is:

$$y = 0.58 + (0.042)x$$

   The interpretation is as follows. If there is no difference between the incumbent and challenger's campaign spending, the incumbent is predicted to get around 58% of the vote (although in a fully specified model the intercept term would probably be lower). For every unit increase in the explanatory variable, the incumbent is predicted to gain another 4.2% of the vote share (the explanatory variable seems to be the log of the difference in campaign expenditures).

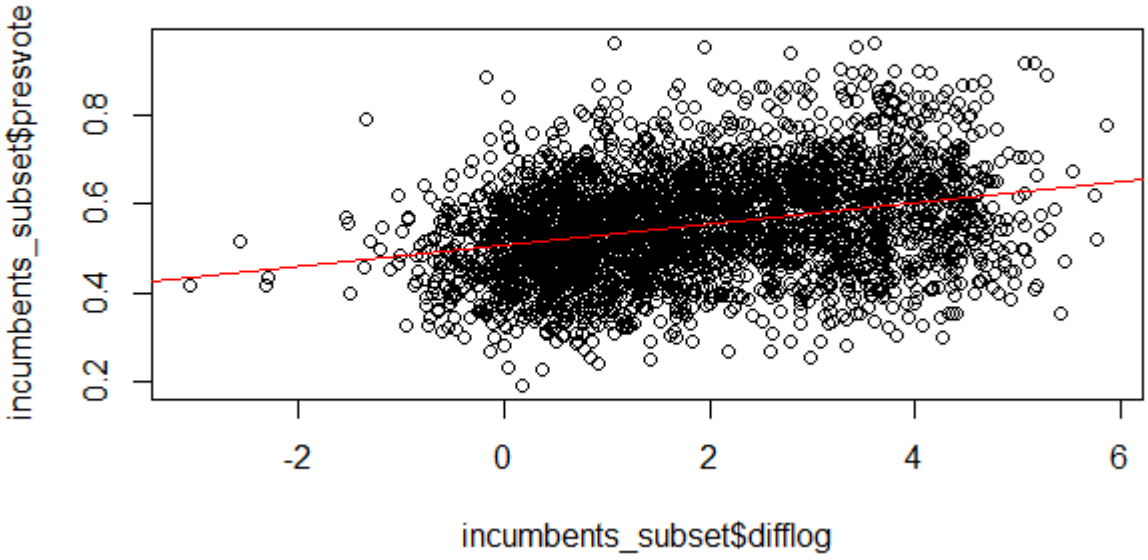
Question 2

1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`. The R code is:

```
results_1 <- lm(presvote ~ difflog, data = incumbents_subset)
```

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

   The R code is as in Question 1 (see also the attached script file). The resulting scatterplot with regression line is:



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

```
residuals_1 <- resid(results_1)
```

4. Write the prediction equation. Using the estimated coefficients from part 1 (and rounding), the prediction equation is:

$$y = 0.51 + (0.02)x$$

The interpretation is as follows. If there is a zero difference in campaign spending between the incumbent and the challenger, the predicted vote share of the presidential candidate of the incumbent's party is just over half. For each increase of one unit in the explanatory variable, the party's presidential candidate is predicted to obtain about 2% more of the vote share

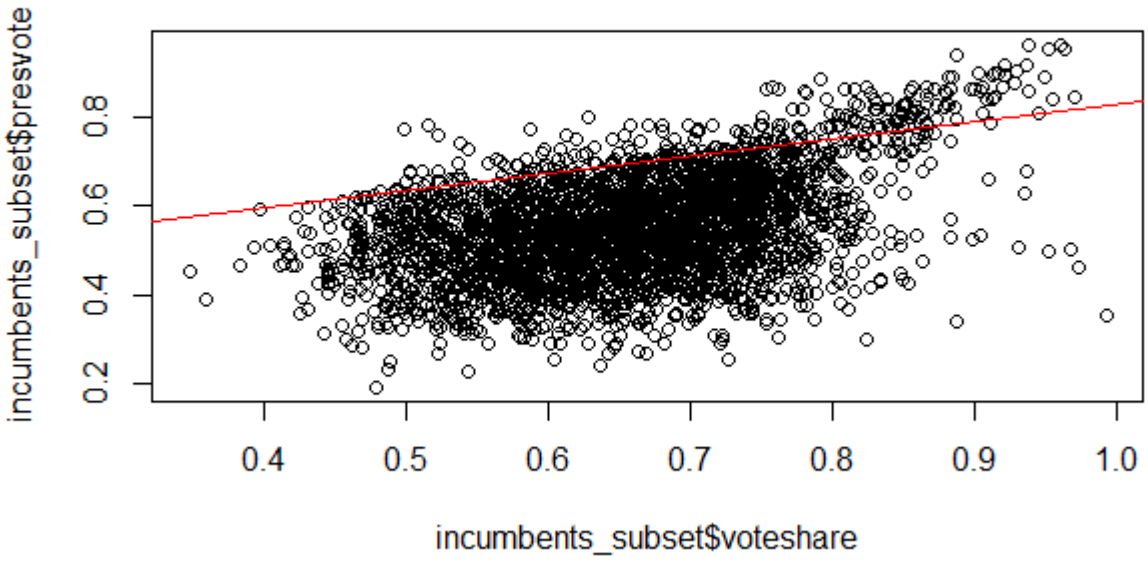
Question 3

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

   The R code is:

   ```
   results_2 <- lm(voteshare ~ presvote, data = incumbents_subset)
   ```

2. Make a scatterplot of the two variables and add the regression line. The code is as before. The scatterplot with the regression line is:



3. Write the prediction equation. Using the estimated coefficients from part 1 (and rounding), the prediction equation is:

$$y = 0.44 + (0.39)x$$

The interpretation is as follows. If the presidential candidate of the incumbent's party received 0% of the vote share (perhaps because the incumbent's party didn't field a presidential candidate) the incumbent is predicted to get around half the vote share; for each percentage point of the vote share received by the presidential candidate of the incumbent's party, the incumbent is predicted to receive around 2% more of the vote share.

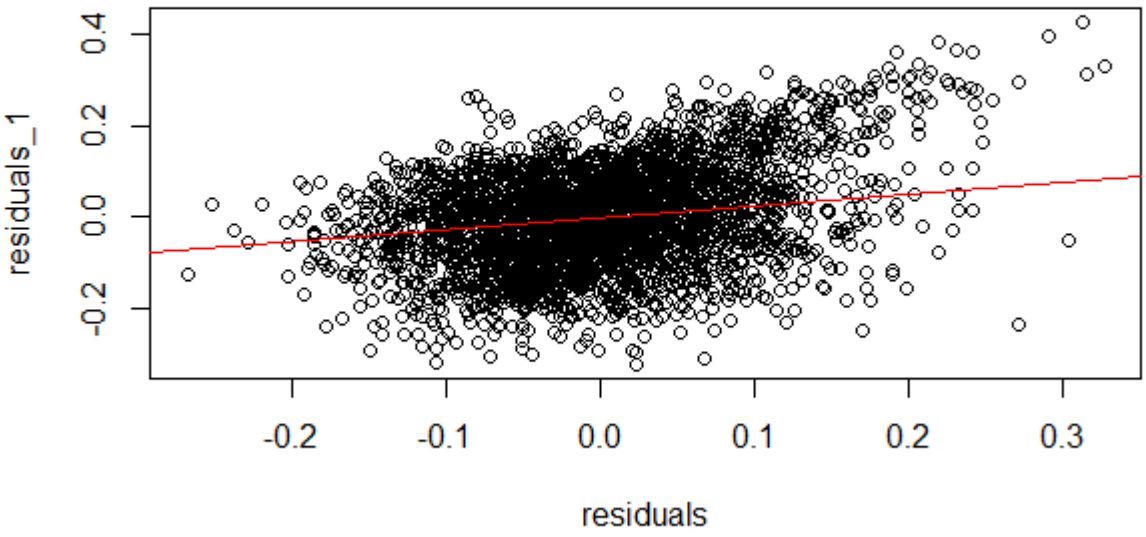
Question 4

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

   The R code is:

```
results_3 <- lm(residuals ~ residuals_1, data = incumbents_subset)
```

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



3. Write the prediction equation. Using the estimates from part 1 and, for once, not rounding:

$$y = (-4.859631e - 18) + (0.25687701)x$$

The intercept (which is expressed by R in scientific notation) is very, very small, and not significantly different from 0; this suggests that it might be worthwhile rerunning the above regression and instructing R to suppress the intercept.

Question 5

1. Run a regression where the outcome variable is the incumbent's `voteshare` and the explanatory variables are `difflog` and `presvote`.

   The R code is:

   ```
   results_4 <- lm(voteshare ~ difflog + presvote, data = incumbents_subset)
   ```

2. Write the prediction equation.

   Using the estimates from part 1:

   $$y = (0.44864422) + (0.03554309)x_1 + (0.25687701)x_2$$

   Where x_1 is the explanatory variable `difflog` and x_2 the variable `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?

   We see that the estimated coefficient on `presvote` is identical to the estimated coefficient on the explanatory variable in the previous model; which was the residuals from question 2, i.e., the amount of variation in `presvote` not explained by the difference in campaign spending between the incumbent and challenger.

   Why might this be the case? In the Question Four model, the outcome variable was that part of the variation in vote share not explained by differences in campaign spending. In the Question Five model, we decompose vote share into: (a) an intercept, (b) that due to differences in campaign spending, and (c) that due to the `presvote` variable.

   Each of these gives the "partial effect" of the explanatory variable on the outcome variable, i.e., the effect on the outcome of increasing the explanatory variable while holding all other explanatory variables constant. This is just another way of saying that the estimated coefficient gives the variation in the outcome associated solely with that particular explanatory variable, since the other variables are being held constant. Thus, we find that the two estimates are identical.