

Problem Set 3 - Solution

Dino Wildi

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**.*
- 2. Make a scatterplot of the two variables and add the regression line.*
- 3. Save the residuals of the model in a separate object.*
- 4. Write the prediction equation.*

```
1 reg1 <- lm(voteshare ~ difflog, data = dataset)
2 summary(reg1)
3
4 plot1 <- ggplot(dataset, aes(x = difflog, y = voteshare)) + geom
  _point() + geom_smooth(method = "lm") +
5   labs(title = "Impact of campaign spending on vote share", x =
     "Difference in campaign spending", y = "Incumbent vote share"
  )
6 plot1
7
8 resid1 <- reg1$residuals
```

The difference in campaign spending has a significant impact with an R-squared of 0.37, explaining a good chunk of variation but not the majority of it. This is also shown in Fig. 1, which shows a general trend to higher vote share if there is more money spent by the incumbent compared to the challenger; and the margin of error is fairly narrow. However, we still see a pretty wide cloud of points around the regression line. The equation to predict the vote share y at any given difference in spending x is:

$$\hat{y} = 0.579 + 0.042x + e$$

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**.
2. Make a scatterplot of the two variables and add the regression line.
3. Save the residuals of the model in a separate object.
4. Write the prediction equation.

```
1 reg2 <- lm(presvote ~ difflog, data = dataset)
2 summary(reg2)
3
4 plot2 <- ggplot(dataset, aes(x = difflog, y = presvote)) + geom_point() +
  geom_smooth(method = "lm") +
5   labs(title = "Impact of campaign spending on incumbent party
  candidate vote share", x = "Difference in campaign spending",
  y = "Incumbent party candidate vote share")
6 plot2
7
8 resid2 <- reg2$residuals
```

Like above, the same relationship also holds when instead of an incumbent, we look at an incumbent party. However, it is a bit less pronounced, with the effect being smaller, as can be seen in Fig. 2. The scatter plot is significantly more dispersed and the trend is less obvious, but still present. Likewise, the R-squared is notably lower (0.088 rather than 0.37) than in the first model using incumbents vote shares. The equation to predict the vote share of the incumbent party y at any given difference in spending x is:

$$\hat{y} = 0.508 + 0.024x + e$$

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*.
2. Make a scatterplot of the two variables and add the regression line.
3. Write the prediction equation.

```

1 reg3 <- lm(voteshare ~ presvote, data = dataset)
2 summary(reg3)
3
4 plot3 <- ggplot(dataset, aes(x = presvote, y = voteshare)) +
  geom_point() + geom_smooth(method = "lm") +
5   labs(title = "Incumbent Party vote share and electoral success
  of incumbent party", x = "Vote share of incumbent party for
  President",
6         y = "Incumbent candidate vote shares")
7 plot3

```

This is another strong correlation, suggesting that if incumbents' parties get a high vote in the presidential election, this also increases their chance to retain their own seat. The correlation shown in Fig. 3 is relatively strong and highly significant, but does not explain all of the variation at all. The equation to predict the vote share of incumbents y at any given vote share of the incumbent's party in the presidential vote x is:

$$\hat{y} = 0.441 + 0.388x + e$$

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.
2. Make a scatterplot of the two residuals and add the regression line.
3. Write the prediction equation.

```

1 reg4 <- lm(resid1 ~ resid2)
2 summary(reg4)
3
4 plot4 <- ggplot(aes(x = resid2, y = resid1), data = NULL) + geom
  _point() + geom_smooth(method = "lm") +
5   labs(title = "Residuals of the first two regressions", x = "
     Residuals from Question 2",
6         y = "Residuals from Question 3")
7 plot4

```

The residuals are correlated to some extent, with the estimate being highly significant. This indicates that there are some unaccounted factors influencing both variables, which makes sense since factors that affect the presidential election are also likely to affect other elections taking part at the same time. Figure 4 shows that this is particularly pronounced at the high end of the range of the residuals. The prediction equation is:

$$\hat{y} = 0.257x + e$$

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*.
2. Write the prediction equation.

```

1 reg5 <- lm(voteshare ~ difflog + presvote, dataset)
2 summary(reg5)

```

The regression shows that both *difflog* and *presvote* have a significant positive effect, suggesting that both of them are related to the vote share of incumbents. The R-squared is significantly higher than in each individual regression (0.45 for both R-squared and adjusted R-squared) suggesting that this model explains more variation; once more suggesting that they are both correlated to the vote share. The prediction equation for the vote share y , predicted by the spending difference x_1 and the presidential race vote share x_2 is:

$$\hat{y} = 0.036x_1 + 0.257x_2 + e$$

*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, i.e. the portion of variation not explained by the two variables, are the same as in question 4. This might be because Question 4 uses the individual residuals, i.e. unexplained variation, of `difflog` and `presvote`. Hence, by putting both variables into the same model in question 5, I expect that we explain the same variation as in Question 4, and thus also have the same residuals.

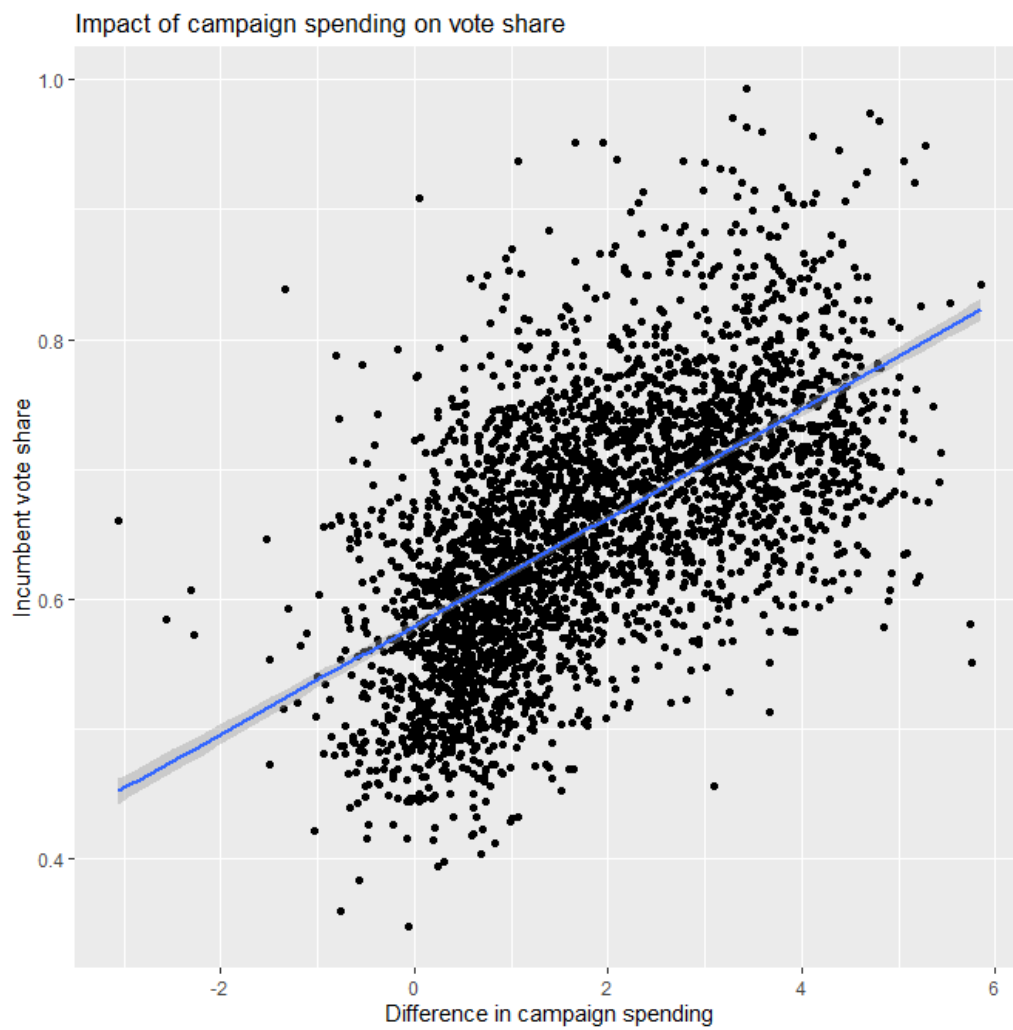


Figure 1: Impact of campaign spending on the vote share of an incumbent

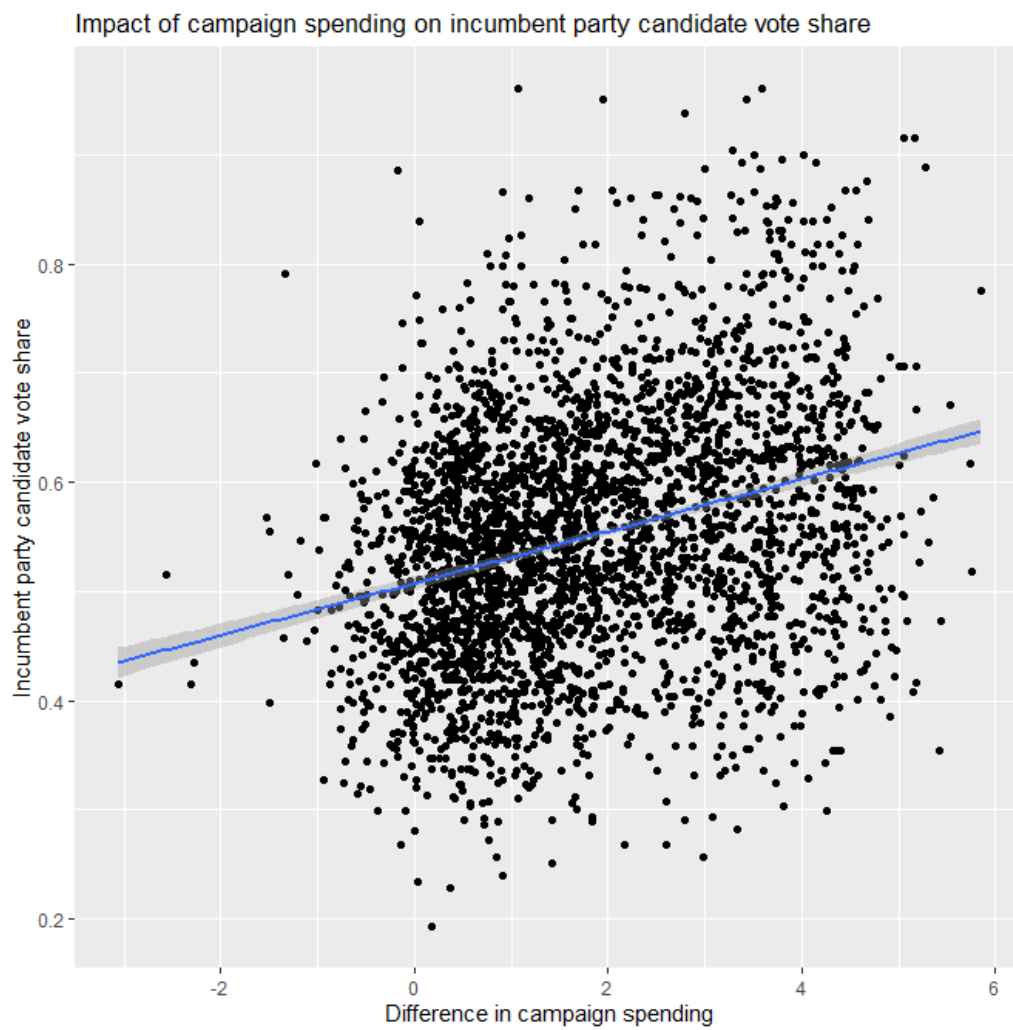


Figure 2: Impact of campaign spending on the vote share of an incumbent

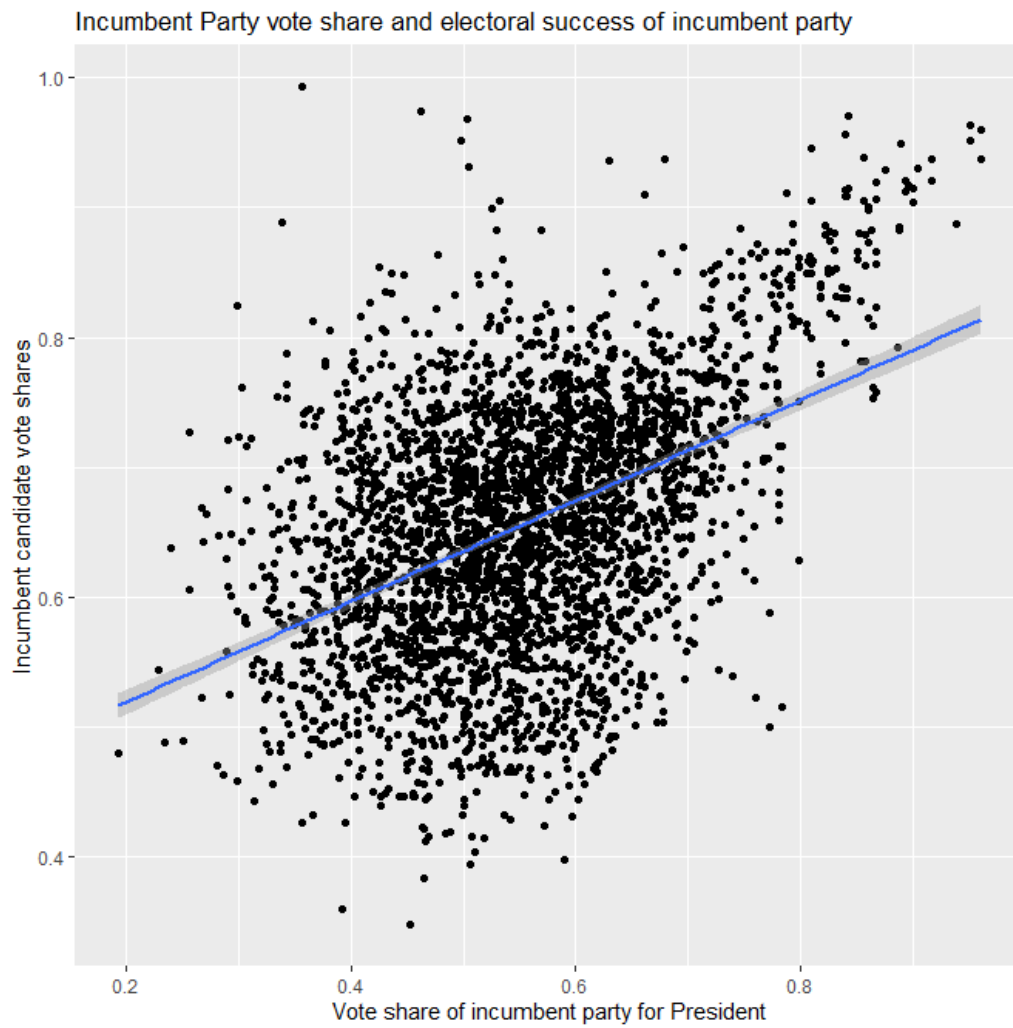


Figure 3: Impact of campaign spending on the vote share of an incumbent

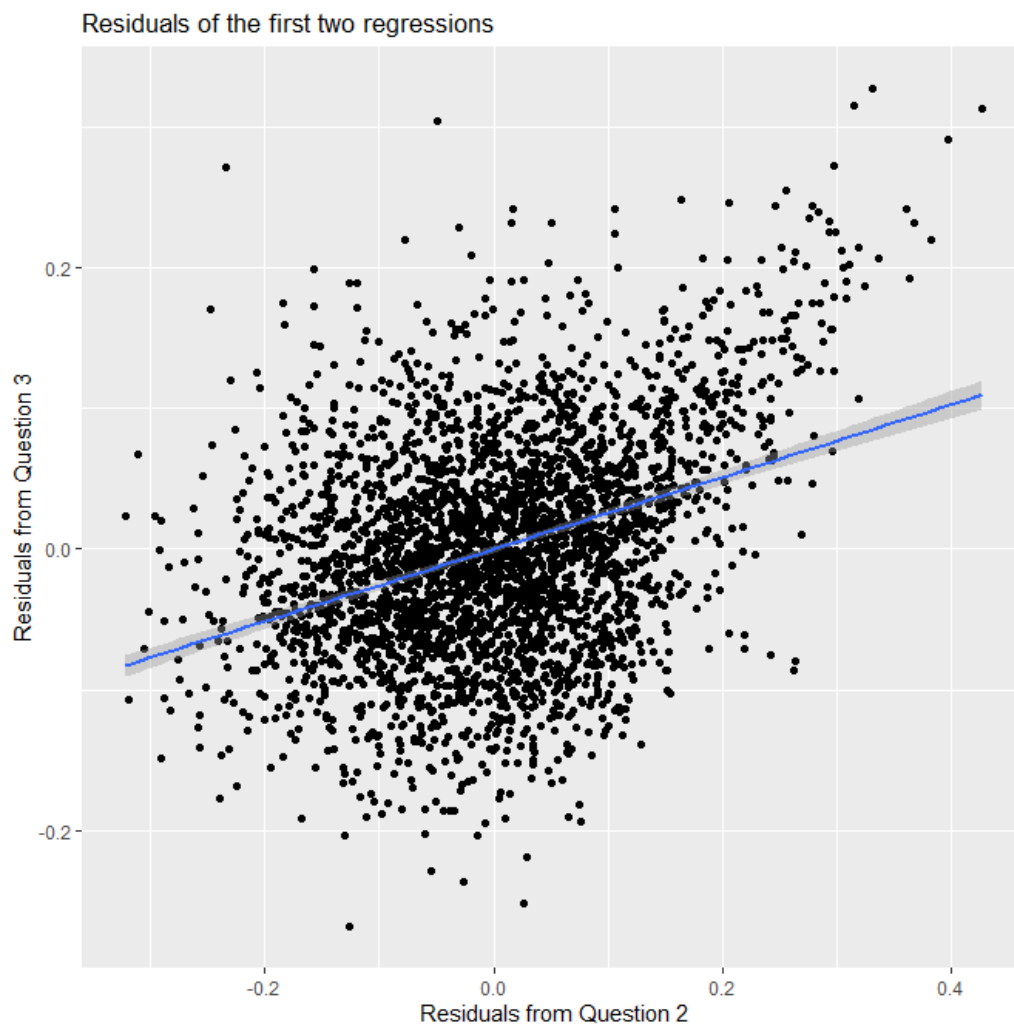


Figure 4: Impact of campaign spending on the vote share of an incumbent