Problem Set 3

Applied Stats/Quant Methods 1

Due: November 11, 2024

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on GitHub.
- This problem set is due before 23:59 on Sunday November 11, 2024. No late assignments will be accepted.

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`.

```r
# modelX will be used as name convention for regressions 1 to 5
model1 <- lm(voteshare ~ difflog, data = inc.sub)
```

```
Coefficients:
(Intercept)      difflog
    0.57903      0.04167
```

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

```
# note: the lm() function and plot() function, when specified this way, have the response and explanatory variables mirrored
plot(inc.sub$difflog, inc.sub$voteshare)
abline(lm(voteshare ~ difflog, data = inc.sub))
```

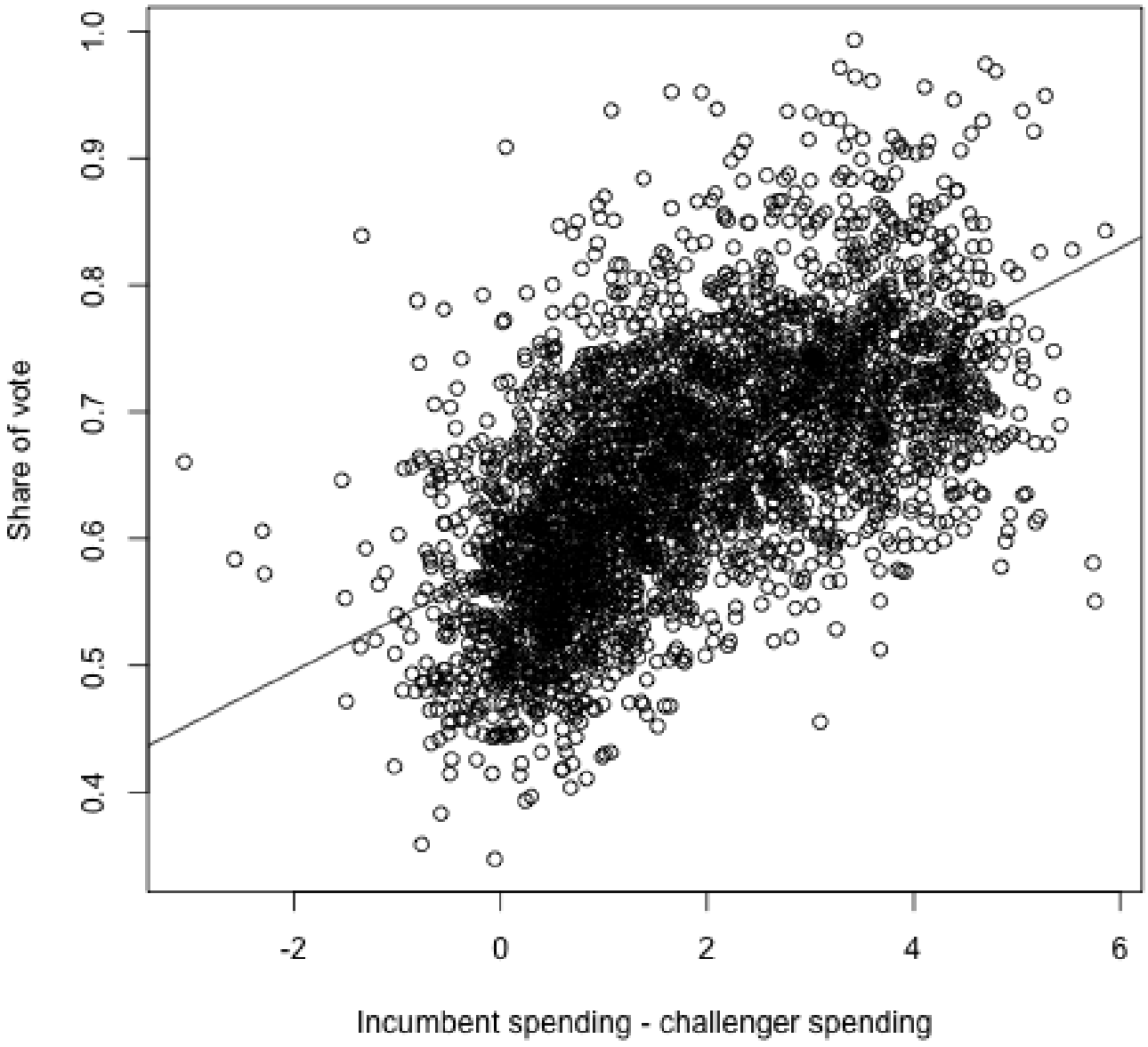
3. Save the residuals of the model in a separate object.

```
# residuals will be stored in objects resX for this and all subsequent questions
res1 <- model1$residuals
```

4. Write the prediction equation.

```
# in this an all subsequent questions, I interpret "prediction equation" as the equation of the line/plane, without the error term
Y1 <- 0.04167*inc.sub$difflog + 0.57903
```

Figure 1: Model 1



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`.

```
model2 <- lm(presvote ~ difflog, data = inc.sub)
```

```
    Coefficients:
(Intercept)      difflog
    0.50758      0.02384
```

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

```
plot(inc.sub$difflog, inc.sub$presvote)
abline(lm(presvote ~ difflog, data = inc.sub))
```

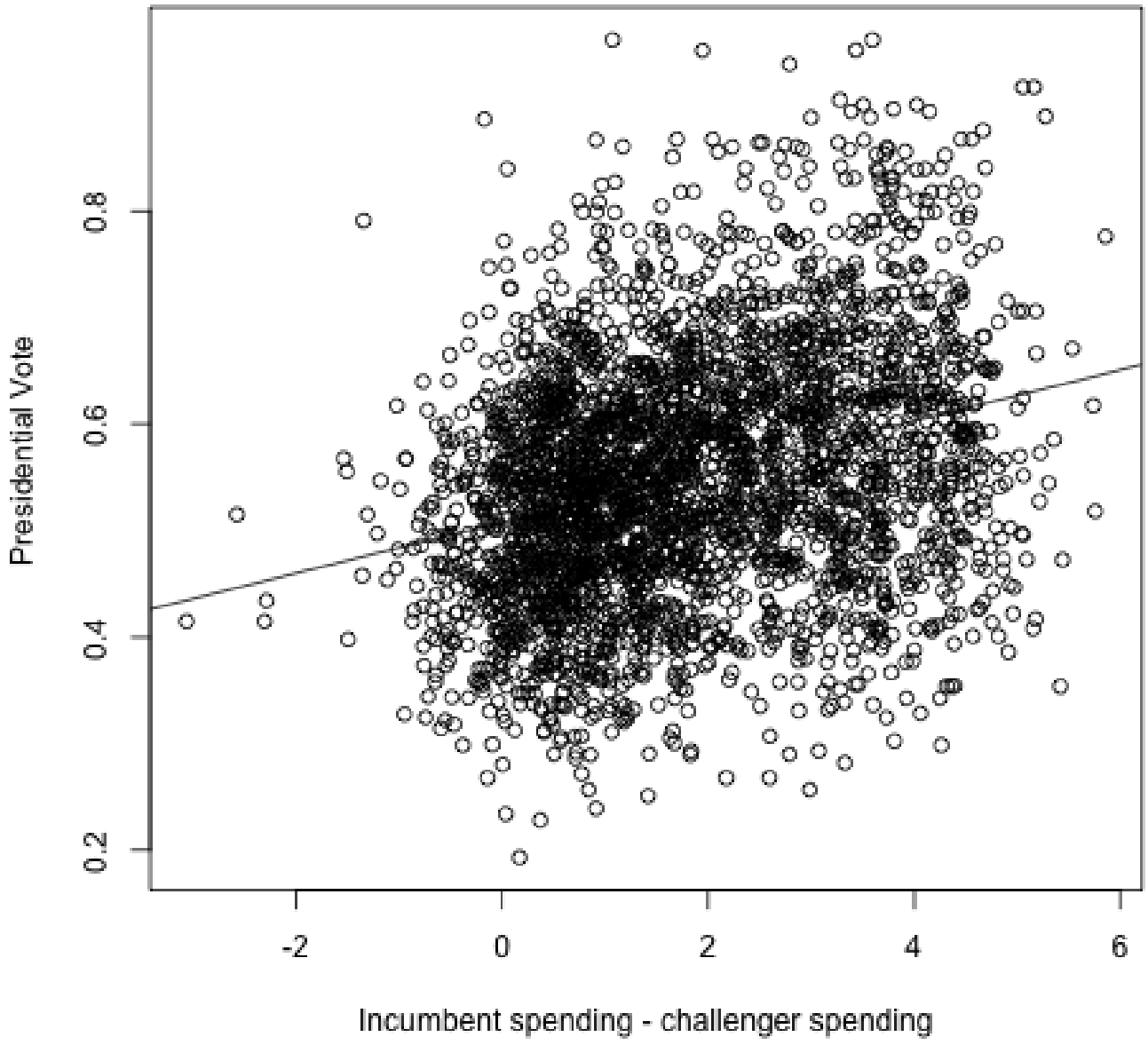
3. Save the residuals of the model in a separate object.

```
res2 <- model2$residuals
```

4. Write the prediction equation.

```
Y2 <- 0.02384*inc.sub$difflog + 0.50758
```

Figure 2: Model 2



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`.

```
model3 <- lm(voteshare ~ presvote, data = inc.sub)
```

```
Coefficients:
(Intercept)     presvote
  0.4413        0.3880
```

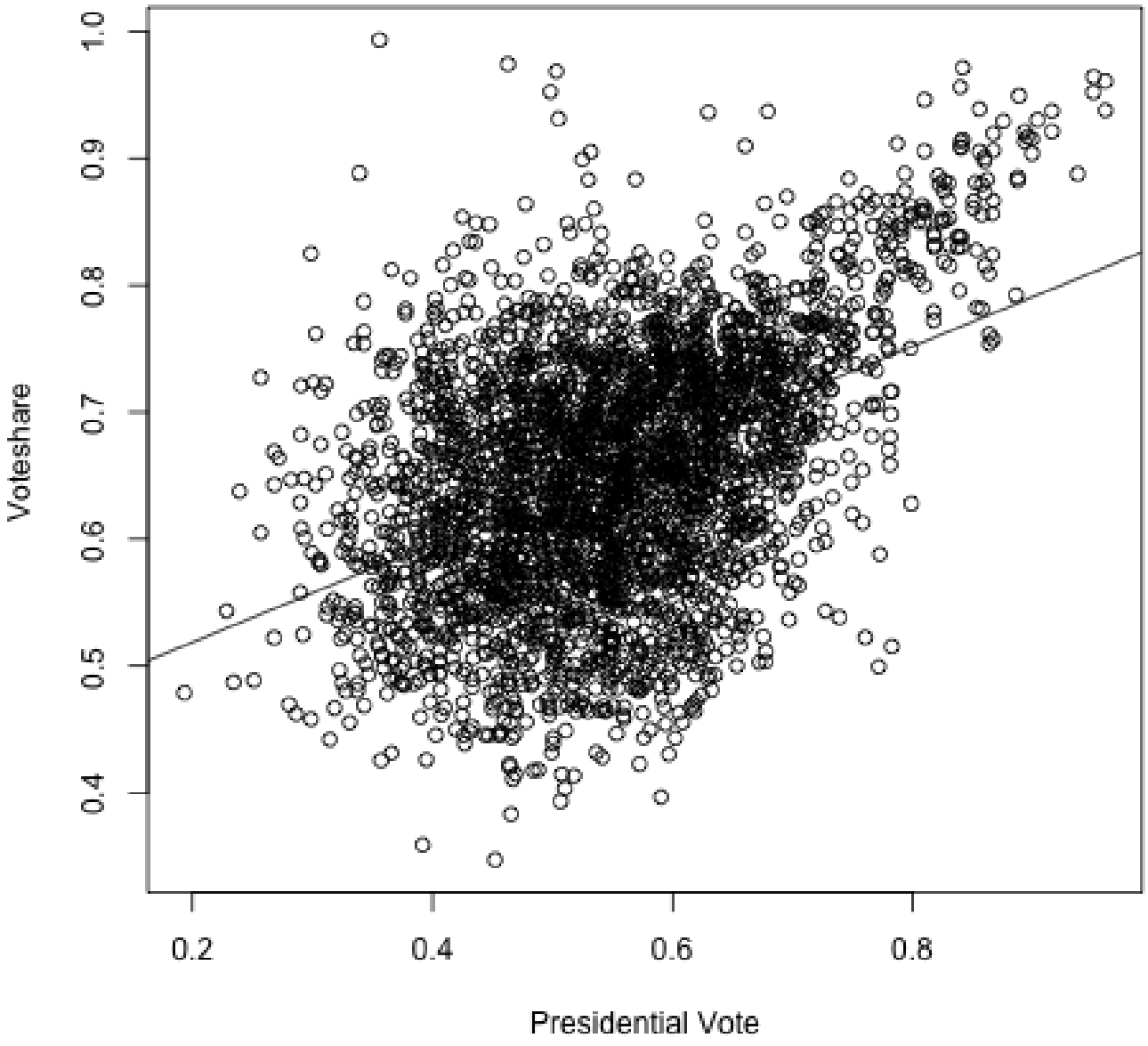
2. Make a scatterplot of the two variables and add the regression line.

```
plot(inc.sub$presvote, inc.sub$voteshare)
abline(lm(voteshare ~ presvote, data = inc.sub))
```

3. Write the prediction equation.

```
Y3 <- 0.3880*inc.sub$presvote + 0.4413
```

Figure 3: Model 3



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.

```
model4 <- lm(res1~res2)
```

```
Coefficients:
(Intercept)         res2
  -1.942e-18    2.569e-01
```

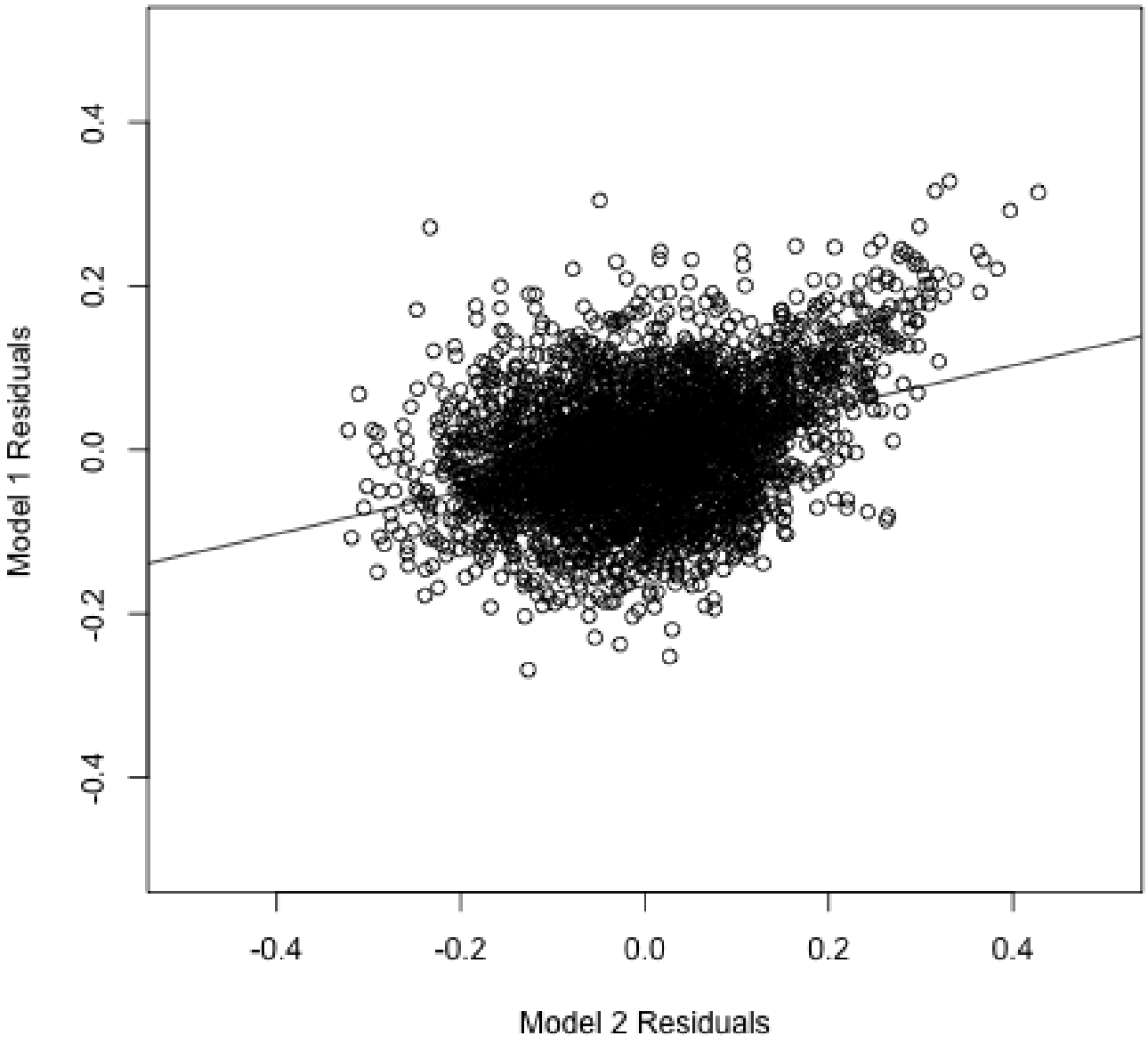
2. Make a scatterplot of the two residuals and add the regression line.

```
plot(res2, res1, ylim = c(-0.5,0.5), xlim = c(-0.5,0.5))
abline(model4)
```

3. Write the prediction equation.

```
Y4 <- 2.569e-01*res2 - 1.942e-18
```

Figure 4: Model 4



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`.

```
model5 <- lm(voteshare ~ difflog + presvote, data = inc.sub)

install.packages("scatterplot3d") # import package for 3D scatter-plots
library(scatterplot3d)
scatterplot3d(inc.sub$difflog, inc.sub$presvote, inc.sub$voteshare)
```

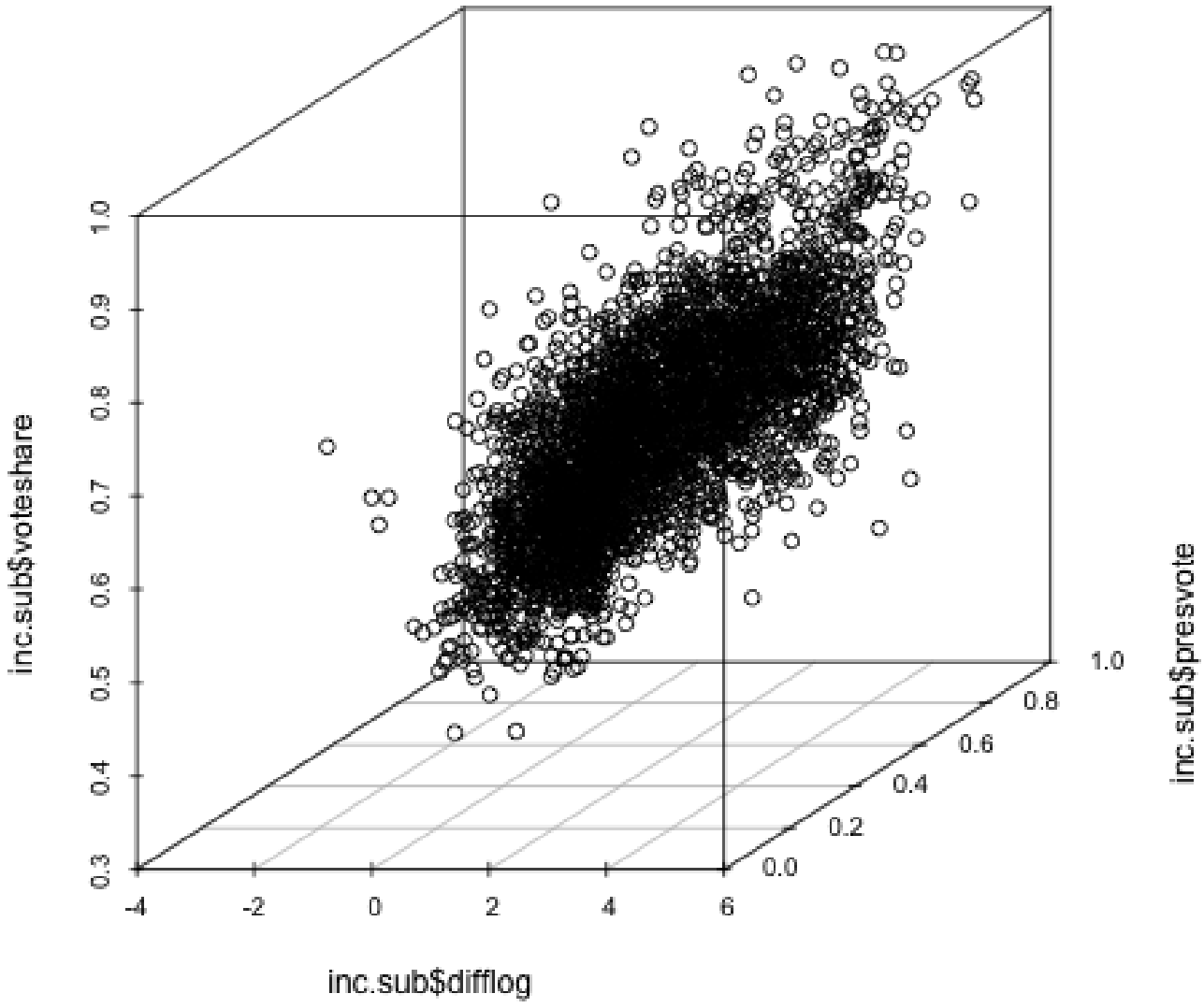
```
Coefficients:
(Intercept)      difflog     presvote
    0.44864      0.03554      0.25688
```

2. Write the prediction equation.

```
Y5 = 0.03554*inc.sub$difflog + 0.25688*inc.sub$presvote + 0.44864
```

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

Figure 5: Model 5



Upon closer inspection of the coefficients found for Question 5 and Question 4, we can see that the coefficient modifying `presvote` in Question 5, 0.25688, is identifical to the coefficient modifying the residuals of Question 2, i.e. the variance in `presvote` not explained

by `difflog`. This is because this coefficient captures the relationship between variance in `voteshare` and `presvote` that cannot be explained by/is independent of `difflog`. This makes sense in the context of our multiple linear regression equation, where this coefficient should have nothing to do with `difflog`. After digging, I discovered this is true because of the Frisch-Waugh-Lovell theorem, often used in Econometrics.