Problem Set 3 Answers

Applied Stats/Quant Methods 1

Sunday 19th November, 2023

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 19, 2023. 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.

- First, I imported the data:
- ```
  inc.sub <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2023/main/datasets/incumbents_subset.csv")
  ```

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

   - Code for Regression Analysis using lm function:
   - ```
     vote_diff <- lm(voteshare ~ difflog, data=inc.sub)
     ```

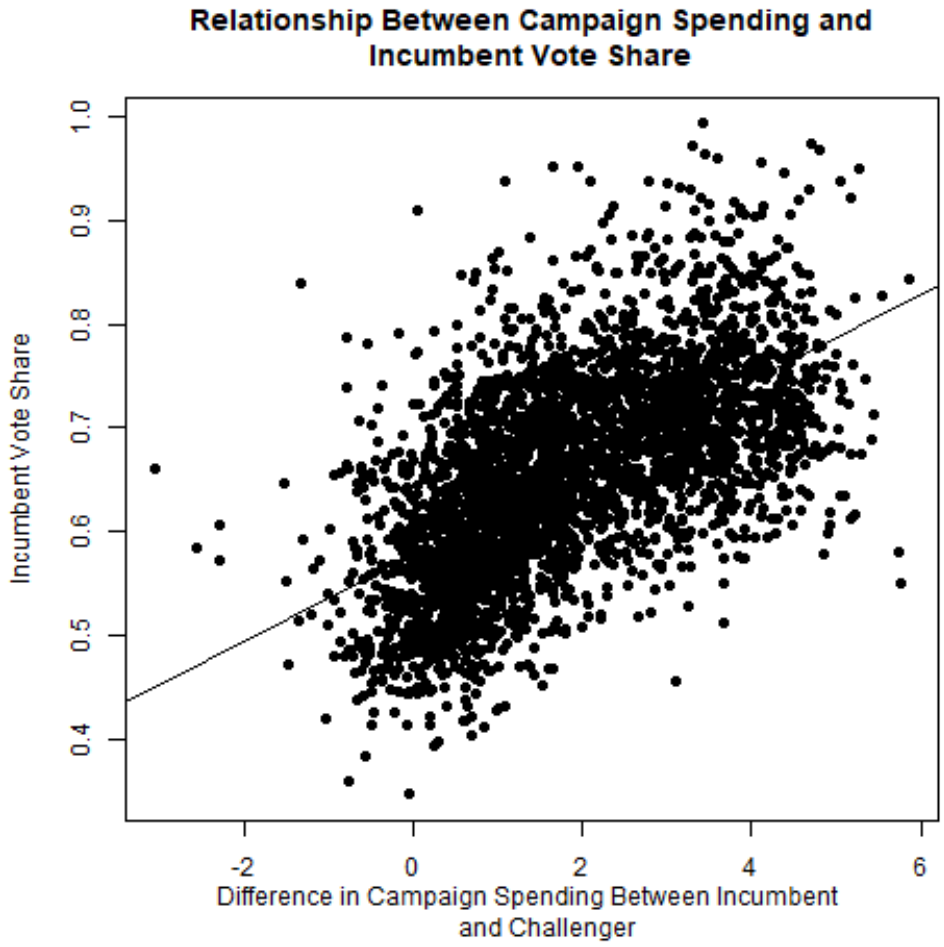
Table 1:

	Dependent variable:
	voteshare
difflog	0.042***
	(0.001)
Constant	0.579***
	(0.002)
Observations	3,193
R^2	0.367
Adjusted R^2	0.367
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,852.791*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

- The plot with the regression line:

```
png(file = "difflog_on_voteshare.png")
plot(inc.sub$difflog, inc.sub$voteshare, pch=16,col = c("black"),
     main = paste(
       strwrap(
         "Relationship Between Campaign Spending and Incumbent Vote Share",
         width = 50
       ),
       collapse = "\n"
     ),
     xlab="Difference in Campaign Spending Between Incumbent and Challenger", ylab="Incumbent Vote Share")
abline(vote_diff)
dev.off()
```



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

   - `ex1resids <- vote_diff$residuals`

4. Write the prediction equation.

   - $\hat{y} = 0.579 + 0.042\beta_1$
   - OR
   - Incumbent Electoral Success = 0.579 + 0.042 x Difference in Campaign Spending

   Interpretation of Regression Results: A one-unit increase in the Difference in Campaign Spending is associated, on average, with a 0.579 unit increase in Incumbent Electoral Success.

   - The null hypotheses for such the coefficient:

  - Null Hypothesis: $\beta_1 = 0$
  - Alternative Hypothesis: $\beta_1 \neq 0$
- The null hypotheses for the intercept is as follows:
  - Null Hypothesis: $\beta_0 = 0$
  - Alternative Hypothesis: $\beta_0 \neq 0$

Because the p-value of the $\beta_1$coefficient is less than 0.05, we can reject the null hypothesis at the 0.05 level.

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`.

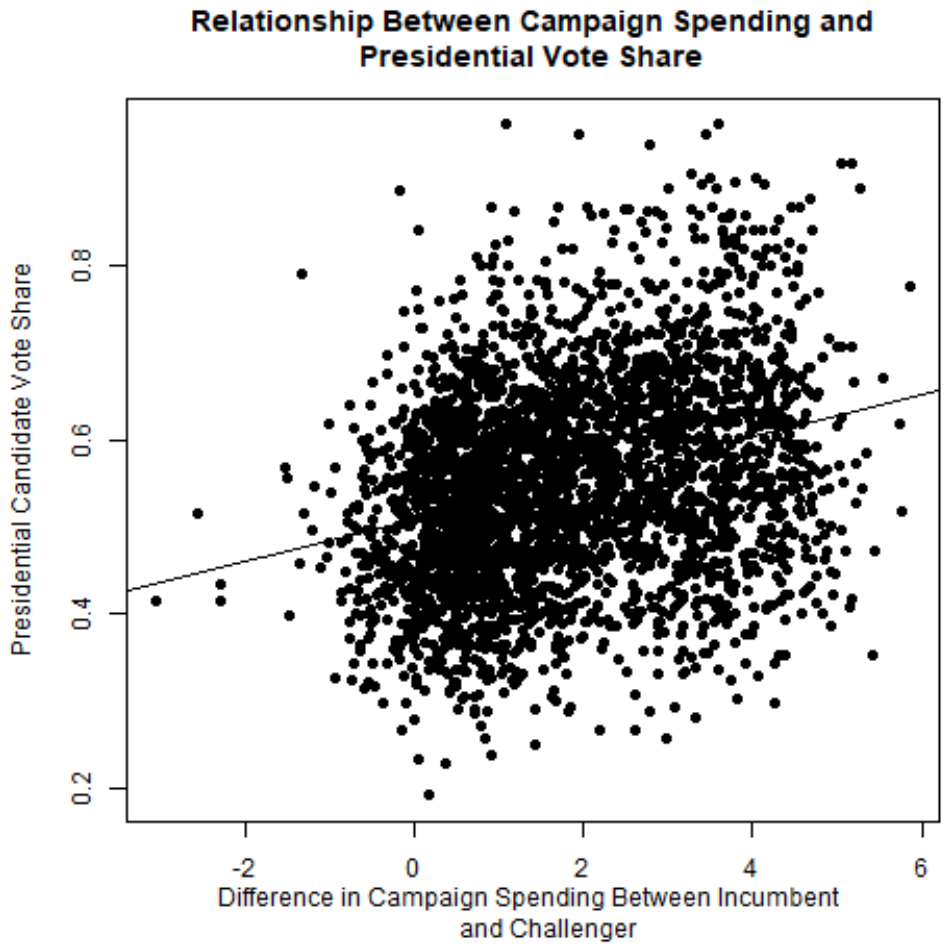
   - Code for Regression Analysis using lm function:
   - ```
     pres_diff <- lm(presvote ~ difflog, data=inc.sub)
     ```

Table 2:

	Dependent variable:
	presvote
difflog	0.024***
	(0.001)
Constant	0.508***
	(0.003)
Observations	3,193
R^2	0.088
Adjusted R^2	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01
     ```

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

```
png(file = "difflog_on_pres_diff.png")
plot(inc.sub$difflog, inc.sub$presvote, pch=16,col = c("black"),
     main=paste(
       strwrap(
         "Relationship Between Campaign Spending and Presidential
    Vote Share",
         width = 50
       ),
       collapse = "\n"
     ),
     xlab="Difference in Campaign Spending Between Incumbent
     and Challenger", ylab="Presidential Candidate Vote Share")
abline(pres_diff)
dev.off()
```



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

   - `ex2resids <- pres_diff$residuals`

4. Write the prediction equation.

   - $\hat{y} = 0.508 + 0.024\beta_1$
   - OR
   - Presidential Candidate Vote Share = 0.508 + 0.024*Difference in Campaign Spending

   Interpretation of Regression Results: A one-unit increase in the Difference in Campaign Spending is associated, on average, with a 0.023837 unit increase in Presidential Vote Share.

   - The null hypotheses for such the coefficient:
     - Null Hypothesis: $\beta_1 = 0$
     - Alternative Hypothesis: $\beta_1 \neq 0$
   - The null hypotheses for the intercept is as follows:
     - Null Hypothesis: $\beta_0 = 0$
     - Alternative Hypothesis: $\beta_0 \neq 0$

   Because the p-value of the $\beta_1$coefficient is less than 0.05, we can reject the null hypothesis at the 0.05 level.

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`.

   - Code for Regression Analysis using lm function:
   - `vote_pres <- lm(voteshare ~ presvote, data=inc.sub)`

Table 3:

	Dependent variable:
	voteshare
presvote	0.388***
	(0.013)
Constant	0.441***
	(0.008)
Observations	3,193
R^2	0.206
Adjusted R^2	0.206
Residual Std. Error	0.088 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

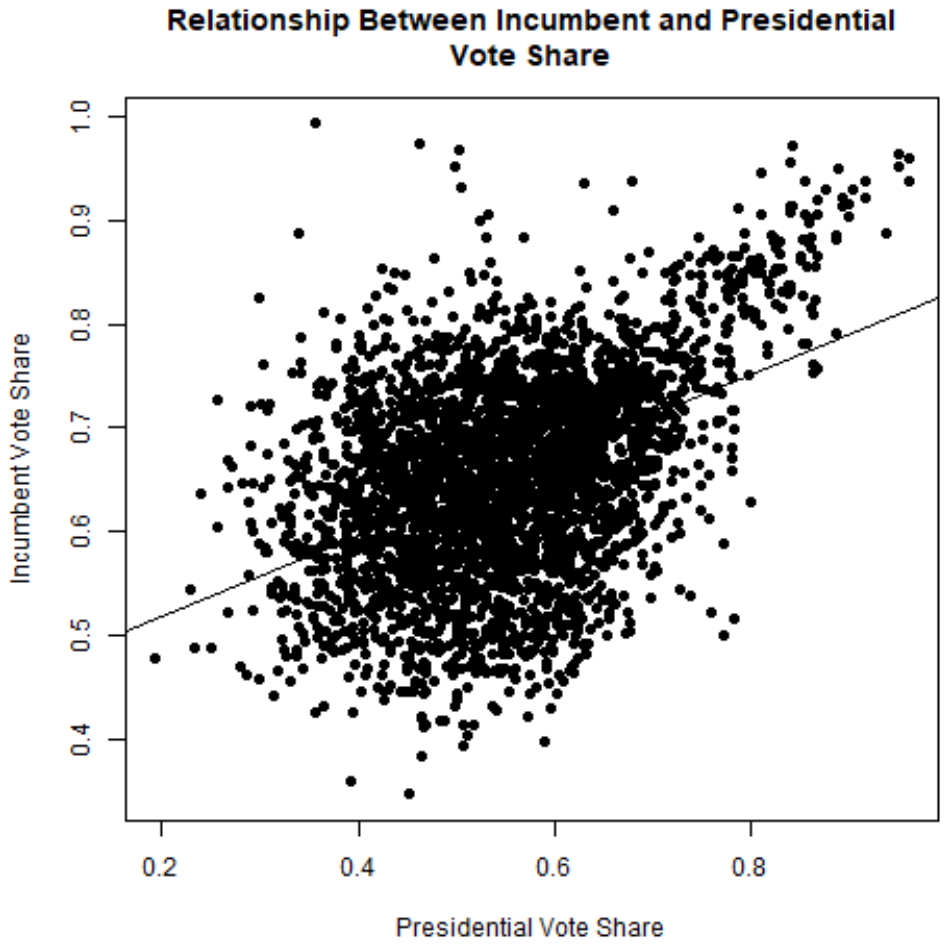
2. Make a scatterplot of the two variables and add the regression line.

   - The plot with the regression line:

```
png(file = "presvote_on_voteshare.png")
plot(inc.sub$presvote, inc.sub$voteshare, pch=16,col = c("black"),
     main = paste(
       strwrap(
         "Relationship Between Incumbent and Presidential Vote Share",
         width = 50
       ),
       collapse = "\n"
     ),
     xlab="Presidential Vote Share", ylab="Incumbent Vote Share")
abline(vote_pres)
dev.off()
```

3. Write the prediction equation.

   - $\hat{y} = 0.441 + 0.388\beta_1$
   - OR
   - Incumbent Electoral Success = 0.441 + 0.388*Vote Share of Presidential Candidate



Interpretation of Regression Results: A one-unit increase in the Incumbent Electoral Success is associated with, on average, a 0.388 unit increase in Presidential Vote Share.

- The null hypotheses for such the coefficient:
  - Null Hypothesis: $\beta_1 = 0$
  - Alternative Hypothesis: $\beta_1 \neq 0$
- The null hypotheses for the intercept is as follows:
  - Null Hypothesis: $\beta_0 = 0$
  - Alternative Hypothesis: $\beta_0 \neq 0$

Because the p-value of the $\beta_1$coefficient is less than 0.05, we can reject the null hypothesis at the 0.05 level.

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.

   - Code for Regression Analysis using lm function:
   - ```
     resid_reg <- lm(ex1resids ~ ex2resids)
     ```

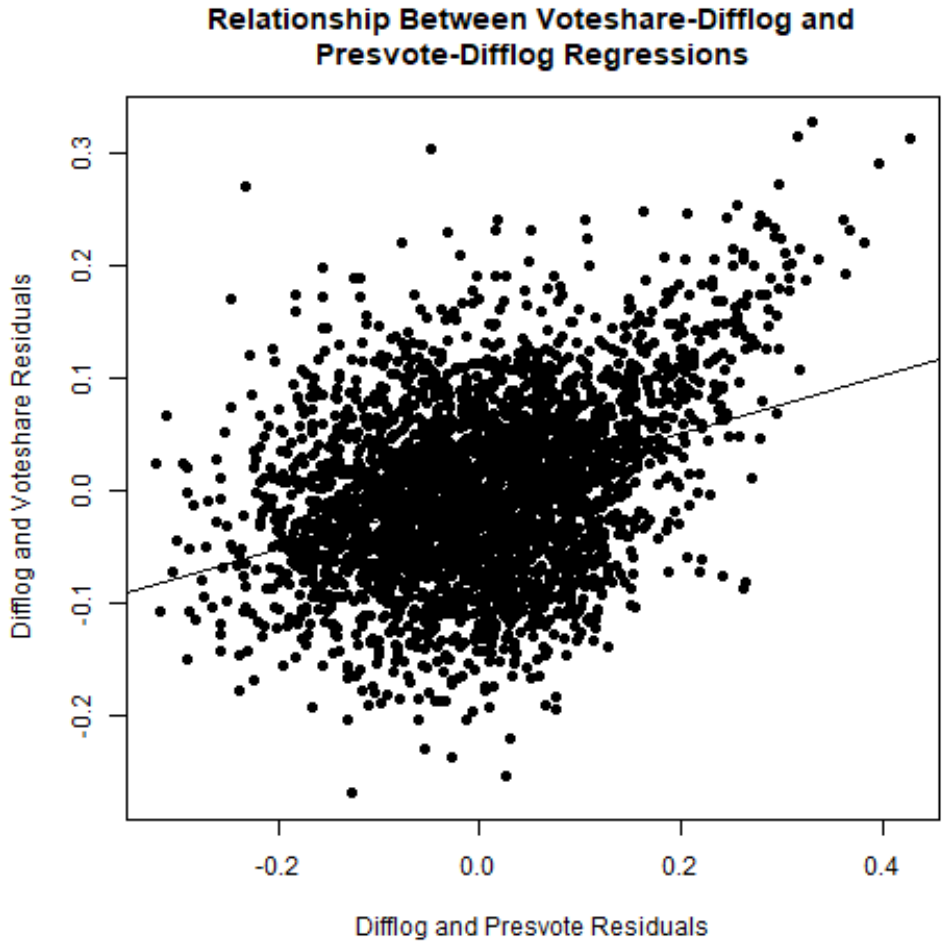
Table 4:

	Dependent variable:
	ex1resids
ex2resids	0.257***
	(0.012)
Constant	−0.000
	(0.001)
Observations	3,193
R^2	0.130
Adjusted R^2	0.130
Residual Std. Error	0.073 (df = 3191)
F Statistic	476.975*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

   - The plot with the regression line:
   - ```
     png(file = "resids_regression.png")
     plot(ex2resids, ex1resids, pch=16,col = c("black"),
          main = paste(
            strwrap(
              "Relationship Between Voteshare-Difflog and Presvote-Difflog Regressions",
              width = 50
     ```

```
        ),
        collapse = "\n"
      ),
      xlab="Difflog and Presvote Residuals", ylab="Difflog and
    Voteshare Residuals")
abline(resid_reg)
dev.off()
```



3. Write the prediction equation.

   - $\hat{y}$ = -0.000 + 0.257β_1
   - OR
   - Residuals for difflog on voteshare(reg1) = -0.000 + 0.257*Residuals for difflog on presvote(reg2)

Interpretation of Regression Results: A one-point increase in error on the second regression is associated with, on average, is associated with a a 0.257 point increase on the first regression.

- The null hypotheses for such the coefficient:
  - Null Hypothesis: $\beta_1 = 0$
  - Alternative Hypothesis: $\beta_1 \neq 0$
- The null hypotheses for the intercept is as follows:
  - Null Hypothesis: $\beta_0 = 0$
  - Alternative Hypothesis: $\beta_0 \neq 0$

Because the p-value of the $\beta_1$coefficient is less than 0.05, we can reject the null hypothesis at the 0.05 level.

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`.
   - Code for Regression Analysis using lm function:
   - ```
     multireg <- lm(voteshare ~ presvote + difflog , data=inc.sub)
     ```
2. Write the prediction equation.
   - $\hat{y}$ = -0.449 + 0.257β_1 + 0.356β_2
   - OR
   - Incumbent Vote Share = -0.449 + 0.257*Presidential Vote Share + 0.036*Difference in Spending
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 β_1 coefficient from Question 4 is the same as the β_1 coefficient for presvote in the regression analysis from Question 5 (They are both 0.257).
     ```

Table 5:

	Dependent variable:
	voteshare
presvote	0.257***
	(0.012)
difflog	0.036***
	(0.001)
Constant	0.449***
	(0.006)
Observations	3,193
R^2	0.450
Adjusted R^2	0.449
Residual Std. Error	0.073 (df = 3190)
F Statistic	1,302.947*** (df = 2; 3190)
Note:	*p<0.1; **p<0.05; ***p<0.01

- This is because the residuals saved from question 1 represent the leftover/unexplained variation of difflog on voteshare. Question 4 shows that the unexplained variation in model 1 is associated with unexplained variation in model 2. In multiple linear regression we are calculating the partial effect, or rather then amount of covariance between an outcome and an explanatory variable that is not explained by the other variables in the model. So, the coefficient for presvote in Question 5 represents the variance explained by presvote **that is not explained by difflog**. In question 4, we see that the coefficient for the residuals (everything not explained by difflog) is 0.257.
- Overall, the results are the same because conceptually, the coefficient of presvote in question 5 represents variance that is not explained by difflog, which conceptually is the exact same thing we are measuring in the regression model in question 4.

Interpretation of Regression Results: A 1 unit increase in presvote is associated with, on average, a 0.257 point increase in voteshare, holding all other variables constant. A 1 unit increase in difflog is associated with, on average, a 0.036 point increase in voteshare, holding all other variables constant.

- The null hypotheses for such the coefficient:

  - Null Hypothesis: $\beta_1 = 0$
  - Alternative Hypothesis: $\beta_1 \neq 0$
- The null hypotheses for the intercept is as follows:
  - Null Hypothesis: $\beta_0 = 0$
  - Alternative Hypothesis: $\beta_0 \neq 0$

Because the p-value of the β_1 and β_2 coefficients are less than 0.05, we can reject the null hypothesis at the 0.05 level.