Problem Set 3

Applied Stats/Quant Methods 1

Due: November 20, 2022

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 20, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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.

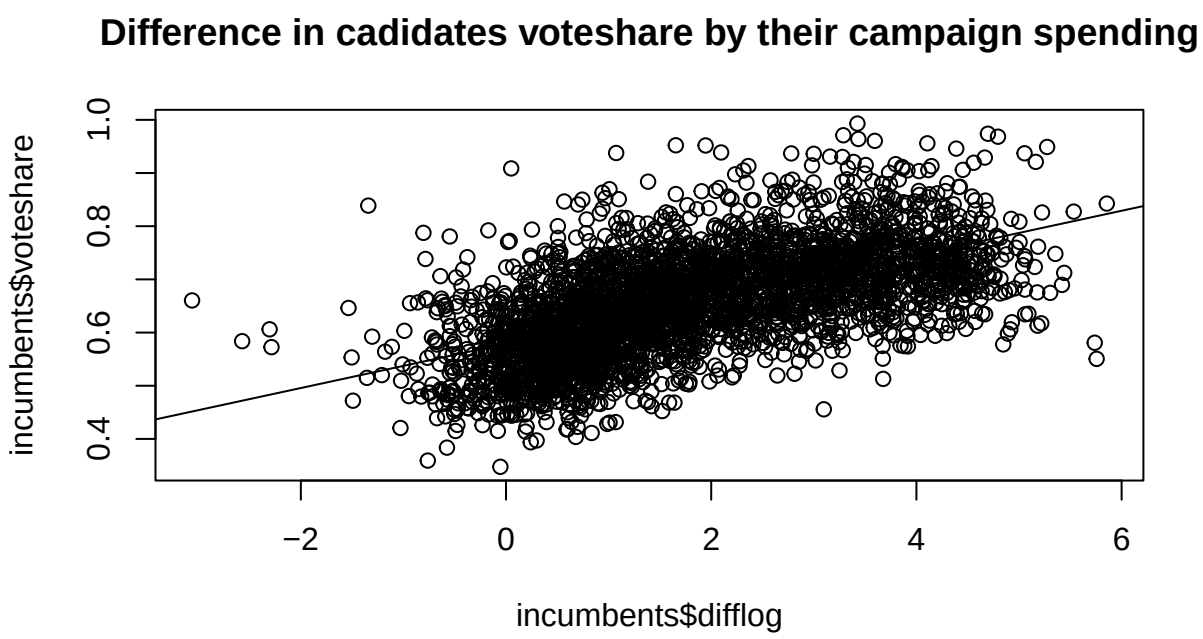
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
lm(dat$voteshare ~ dat$difflog)
dlvs.lm <- lm(dat$voteshare ~ dat$difflog)
summary(dlvs.lm)
stargazer(dlvs.lm)
```

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.



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

```
dlvs.resid <- resid(dlvs.lm)
dlvs.resid
```

4. Write the prediction equation. Y = B0 + B1xi where Y is predicted value of the OV, B0 is the intercept, B1 is the slope of the regression line, and xi a value of the the EV. Using the regression coefficient; Y = 0.58 + 0.04xi.

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.

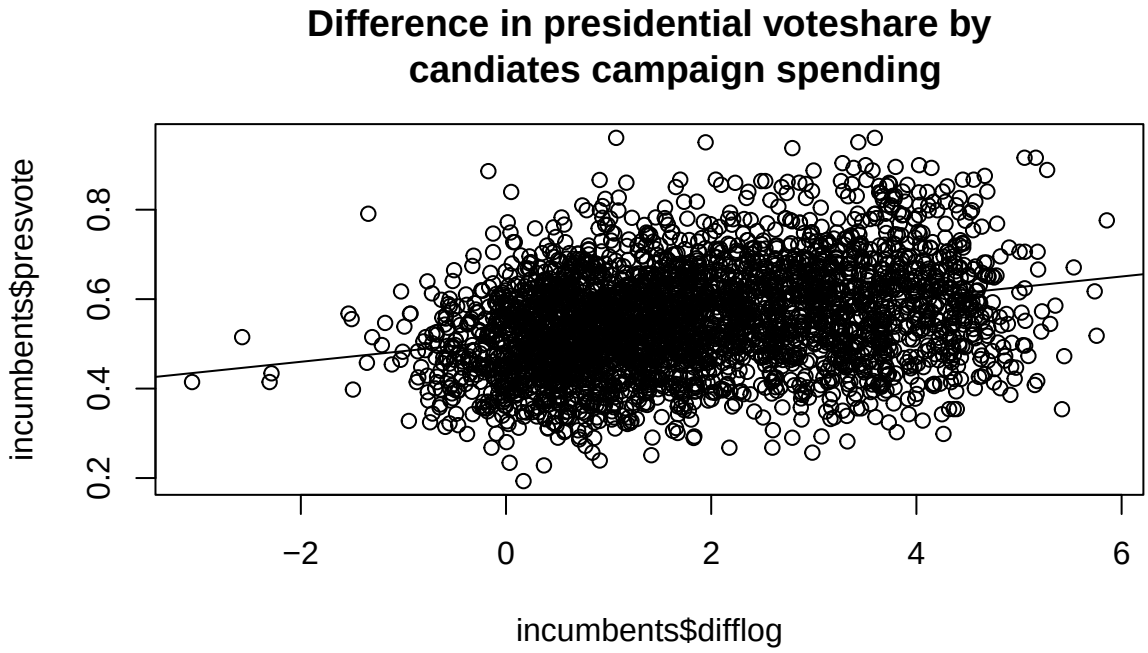
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

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

```
lm(dat$voteshare ~ dat$difflog)
dlvs.lm <- lm(dat$voteshare ~ dat$difflog)
summary(dlvs.lm)
stargazer(dlvs.lm)
```

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



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

```
pvdl.resid <- resid(pvdl.lm)
```

4. Write the prediction equation. Y = B0 + B1x where Y is the predicted value of the OV, B0 is the intercept, B1 is the slope of the regression line, and x a given value of the the EV; Y = 0.51 + 0.02xi.

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.

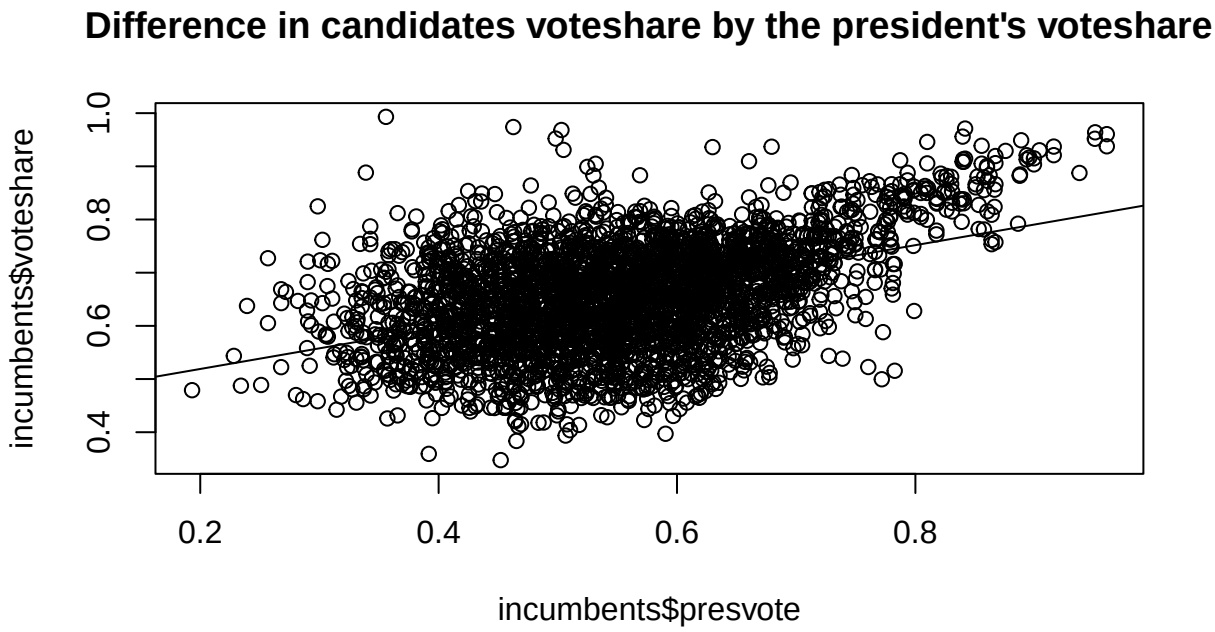
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`.

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

```
lm(dat$voteshare ~ dat$presvote)
pvvs.lm <- lm(dat$voteshare ~ dat$presvote)
summary(pvvs.lm)
stargazer(pvvs.lm)
```

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



3. Write the prediction equation. Y = B0 + B1xi where Y is predicted value of the OV, B0 is the intercept, B1 is the slope of the regression line, and xi a given value of the the EV; Y = 0.44 + 0.39xi.

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.

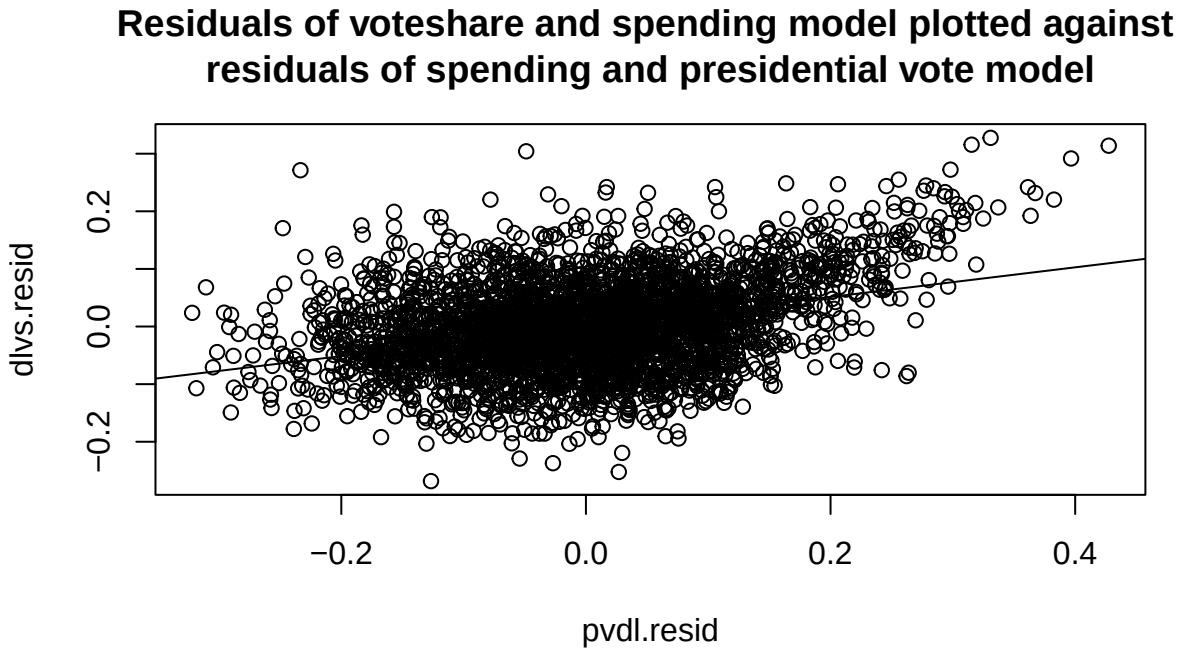
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

Table 4:

	Dependent variable:
	dlvs.resid
pvdl.resid	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

```
lm(dlvs.resid ~ pvdl.resid)
resid.lm <- lm(dlvs.resid ~ pvdl.resid)
summary(resid.lm)
stargazer(resid.lm)
```

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



3. Write the prediction equation. Y = B0 + B1xi where Y is predicted value of the OV, B0 is the intercept, B1 is the slope of the regression line, and xi is a given value of the the EV. Using the regression coefficient: Y = -5.2 + 2.56xi.

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`.

Table 5:

	Dependent variable:
	voteshare
difflog	0.036***
	(0.001)
presvote	0.257***
	(0.012)
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

```
lm(dat$voteshare ~ dat$difflog + dat$presvote)
dlpvvs.mm <- lm(dat$voteshare ~ dat$difflog + dat$presvote)
summary(dlpvvs.mm)
stargazer(dlpvvs.mm)
```

2. Write the prediction equation. Y = B0 + B1xi + B1xii where Y is predicted value of the OV, B0 is the intercept, B1 is the slope of the regression line, and xi and xii are given values of the the EVs. Y = -0.45 + 0.03xi + 0.25xii.

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