Problem Set 3

QTM 200: Applied Regression Analysis

Due: February 17, 2020

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 the course GitHub page in `.pdf` form.
- This problem set is due at the beginning of class on Monday, February 17, 2020. No late assignments will be accepted.
- Total available points for this homework is 100.

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 (20 points)

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`.

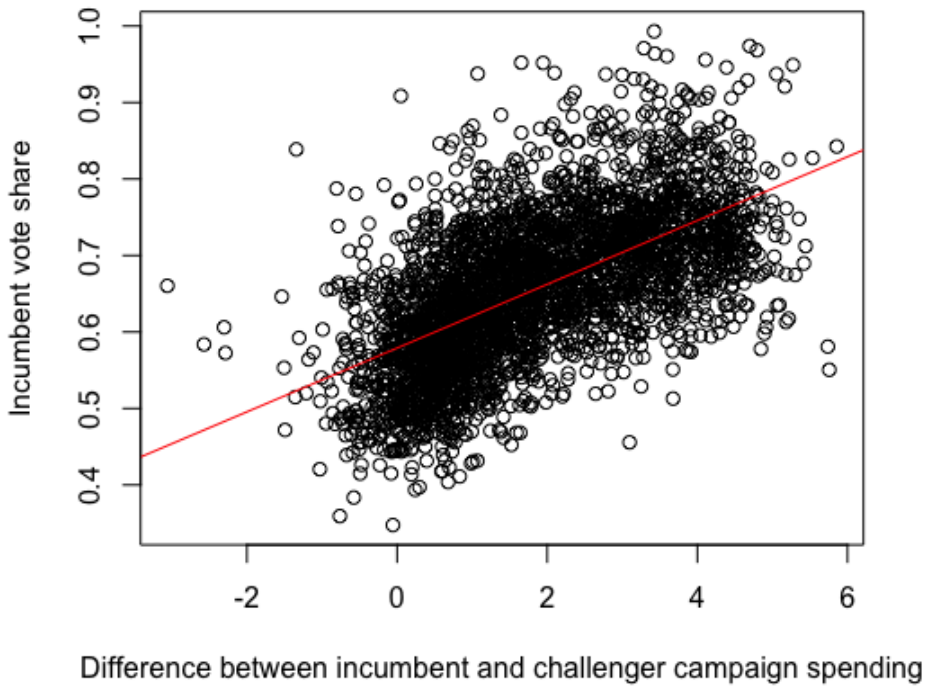
   Linear model: $y = 0.5790 + 0.0417x$

```
# x = difflog (explanatory)
# y = voteshare (outcome)

# run a regression where the outcome variable is voteshare and the
    explanatory variable is difflog.
ymean <- mean(incumbents$voteshare)
```

```r
xmean <- mean(incumbents$difflog)
ysum <- sum(incumbents$voteshare)
xsum <- sum(incumbents$difflog)
yy <- (incumbents$voteshare) - (ymean)
xx <- (incumbents$difflog) - (xmean)
yyxxsum <- sum(yy*xx)
xxsq <- (xx)^2
sumxxsq <- sum(xxsq)
betaincumbs <- yyxxsum/sumxxsq
betaincumbs
# beta = 0.0417
alphaincumbs <- ymean - (betaincumbs*xmean)
alphaincumbs
# alpha = 0.5790
# linear model: y = 0.5790 + 0.0417x
# check work
incumbreg <- lm(incumbents$voteshare ~ incumbents$difflog)
incumbreg
```

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



```
plot(incumbents$difflog, incumbents$voteshare,
     xlab="difflog", ylab="voteshare")
abline(a=0.5790, b=0.0417, col="red")
```

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

```
residscheck <- residuals(incumbreg)
residscheck
```

4. Write the prediction equation.

Question 2 (20 points)

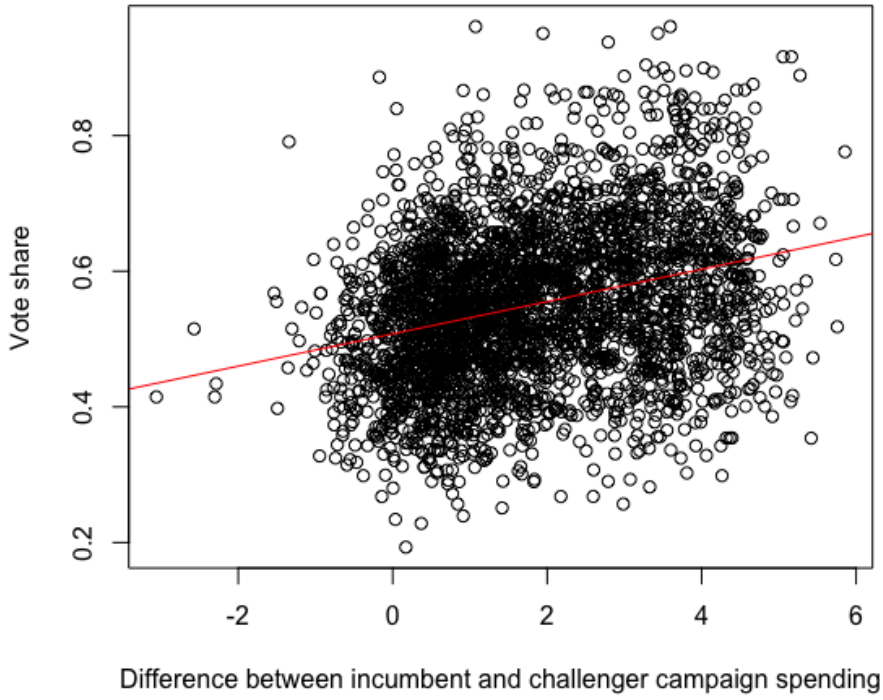
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

   Linear model: y = 0.5076 + 0.0238x

```
y2mean <- mean(incumbents$presvote)
y2sum <- sum(incumbents$presvote)
yy2 <- (incumbents$presvote) - y2mean
yyxxsum2 <- sum(xx*yy2)
betaincumbs2 <- yyxxsum2/sumxxsq
betaincumbs2
# beta = 0.0238
alphaincumbs2 <- y2mean - (betaincumbs2*xmean)
alphaincumbs2
# alpha = 0.5076
# linear model: y = 0.5076 + 0.0238x
# check work
incumbreg2 <- lm(incumbents$presvote ~ incumbents$difflog)
incumbreg2
```

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




```
plot(incumbents$difflog, incumbents$presvote,
     xlab="difflog", ylab="presvote")
abline(a=0.5076, b=0.0238, col="red")
```

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

4. Write the prediction equation.

Question 3 (20 points)

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.

Question 4 (20 points)

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.

Question 5 (20 points)

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.

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