

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: $y = 0.5790 + 0.0417x$

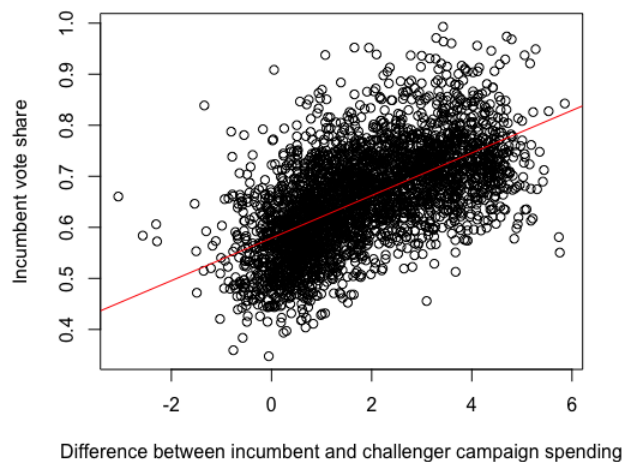
```
1 # x = difflog (explanatory)
2 # y = voteshare (outcome)
3
4 # run a regression where the outcome variable is voteshare and the
  explanatory variable is difflog.
5 ymean <- mean(incumbents$voteshare)
```

```

6 xmean <- mean(incumbents$difflog)
7 ysum <- sum(incumbents$voteshare)
8 xsum <- sum(incumbents$difflog)
9 yy <- (incumbents$voteshare) - (ymean)
10 xx <- (incumbents$difflog) - (xmean)
11 yyxxsum <- sum(yy*xx)
12 xxsq <- (xx)^2
13 sumxxsq <- sum(xxsq)
14 betaincumbs <- yyxxsum/sumxxsq
15 betaincumbs
16 # beta = 0.0417
17 alphaincumbs <- ymean - (betaincumbs*xmean)
18 alphaincumbs
19 # alpha = 0.5790
20 # linear model: y = 0.5790 + 0.0417x
21 # check work
22 incumbreg <- lm(incumbents$voteshare ~ incumbents$difflog)
23 incumbreg

```

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



```

1 plot(incumbents$difflog, incumbents$voteshare,
2       xlab="Difference between incumbent and challenger campaign spending",
3       ylab="Incumbent vote share")
3 abline(a=0.5790, b=0.0417, col="red")

```

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

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

4. Write the prediction equation.

Question 2 (20 points)

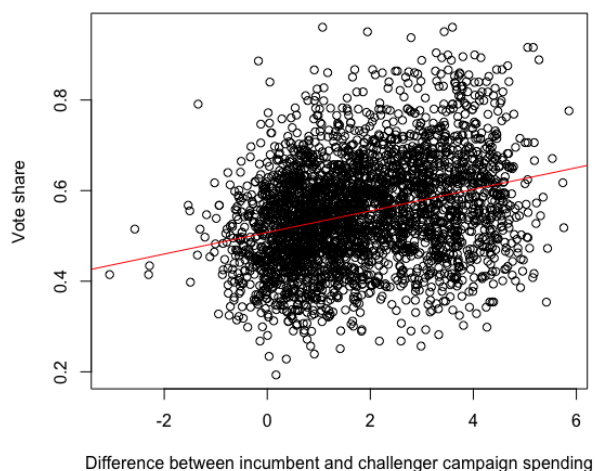
We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

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

Linear model: $y = 0.5076 + 0.0238x$

```
1 y2mean <- mean(incumbents$presvote)
2 y2sum <- sum(incumbents$presvote)
3 yy2 <- (incumbents$presvote) - y2mean
4 yyxxsum2 <- sum(xx*yy2)
5 betaincumbs2 <- yyxxsum2/sumxxsq
6 betaincumbs2
7 # beta = 0.0238
8 alphaincumbs2 <- y2mean - (betaincumbs2*xmean)
9 alphaincumbs2
10 # alpha = 0.5076
11 # linear model: y = 0.5076 + 0.0238x
12 # check work
13 incumbreg2 <- lm(incumbents$presvote ~ incumbents$difflog)
14 incumbreg2
```

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



```
1 plot(incumbents$difflog, incumbents$presvote,
```

```
2      xlab="Difference between incumbent and challenger campaign spending"  
      , ylab="Vote share")  
3 abline(a=0.5076, b=0.0238, col="red")
```

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

```
1 resids2 <- residuals(incumbreg2)  
2 resids2
```

4. Write the prediction equation.

Question 3 (20 points)

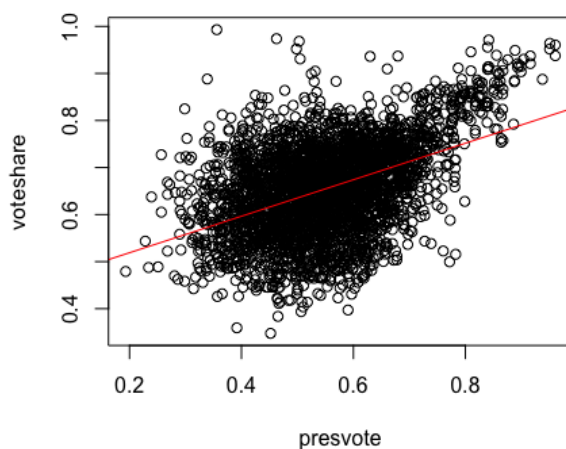
We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

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

Linear model: $y = 0.4413 + 0.3880x$

```
1 xmean3 <- mean(incumbents$presvote)
2 ymean3 <- mean(incumbents$voteshare)
3 xsum3 <- sum(incumbents$presvote)
4 ysum3 <- sum(incumbents$voteshare)
5 yy3 <- (incumbents$voteshare) - (ymean3)
6 xx3 <- (incumbents$presvote) - (xmean3)
7 yyxxsum3 <- sum(yy3*xx3)
8 xxsq3 <- (xx3)^2
9 sumxxsq3 <- sum(xxsq3)
10 beta3 <- yyxxsum3/sumxxsq3
11 beta3
12 # beta = 0.3880
13 alpha3 <- ymean3 - (beta3*xmean3)
14 alpha3
15 # alpha = 0.4413
16 # linear model: y= 0.4413 + 0.3880x
17 # check work
18 reg3 <- lm(incumbents$voteshare~incumbents$presvote)
19 reg3
```

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



```
1 plot(incumbents$presvote, incumbents$voteshare,
2       xlab="presvote", ylab="voteshare")
3 abline(a=0.4413, b=0.3880, col="red")
```

3. Write the prediction equation.

Question 4 (20 points)

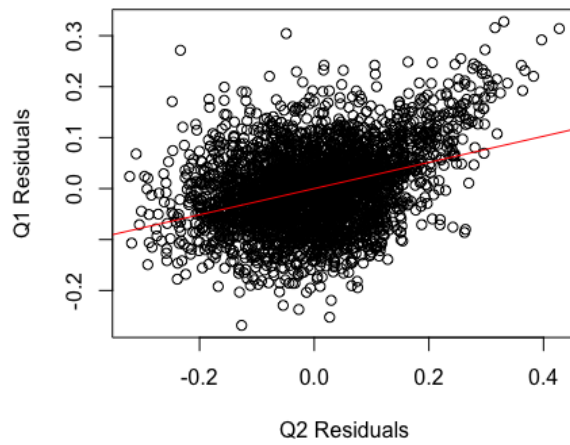
The residuals from part (a) tell us how much of the variation in **voteshare** is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in **presvote** is *not* explained by the difference in spending between incumbent 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.

Linear model: $y = -4.860 \times 10^{-18} + 0.2569x$

```
1 xmean4 <- mean(resids2)
2 ymean4 <- mean(residscheck)
3 xsum4 <- sum(resids2)
4 ysum4 <- sum(residscheck)
5 yy4 <- residscheck - (ymean4)
6 xx4 <- resids2 - (xmean4)
7 yyxxsum4 <- sum(yy4*xx4)
8 xxsq4 <- (xx4)^2
9 sumxxsq4 <- sum(xxsq4)
10 beta4 <- yyxxsum4/sumxxsq4
11 beta4
12 # beta = 0.2569
13 alpha4 <- ymean4 - (beta4*xmean4)
14 alpha4
15 # alpha = -4.939 x 10^(-18)
16 # linear model: y = -4.939 x 10^(-18) + 0.2569x
17 # check work
18 reg4 <- lm(residscheck ~ resids2)
19 reg4
20 summary(reg4)
```

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



```
1 plot(resids2, residscheck,
2       xlab="Q2 Residuals", ylab="Q1 Residuals")
3 abline(reg4, col="red")
```

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?

```
1 #####  
2 # load libraries
```

```

3 # set wd
4 # clear global .envir
5 #####
6
7 # remove objects
8 rm(list=ls())
9 # detach all libraries
10 detachAllPackages <- function() {
11   basic.packages <- c("package:stats", "package:graphics", "package:grDevices",
12     "package:utils", "package:datasets", "package:methods", "package:base")
13   package.list <- search()[ifelse(unlist(gregexpr("package:", search()))==1,
14     TRUE, FALSE)]
15   package.list <- setdiff(package.list, basic.packages)
16   if (length(package.list)>0) for (package in package.list) detach(package,
17     character.only=TRUE)
18 }
19 detachAllPackages()
20
21 # load libraries
22 pkgTest <- function(pkg){
23   new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
24   if (length(new.pkg))
25     install.packages(new.pkg, dependencies = TRUE)
26   sapply(pkg, require, character.only = TRUE)
27 }
28
29 # here is where you load any necessary packages
30 # ex: stringr
31 # lapply(c("stringr"), pkgTest)
32
33 # set working directory, import datasets
34 setwd("~/GitHub/QT200Spring2020/problem_sets/PS3")
35 incumbents <- read.csv("incumbents_subset.csv")
36 #####
37 # Problem 1:
38 #####
39 # x = difflog (explanatory)
40 # y = voteshare (outcome)
41
42 # run a regression where the outcome variable is voteshare and the explanatory
43   variable is difflog.
44 ymean <- mean(incumbents$voteshare)
45 xmean <- mean(incumbents$difflog)
46 ysum <- sum(incumbents$voteshare)
47 xsum <- sum(incumbents$difflog)
48 yy <- (incumbents$voteshare) - (ymean)
49 xx <- (incumbents$difflog) - (xmean)
50 yyxxsum <- sum(yy*xx)

```

```

50 xxsq <- (xx)^2
51 sumxxsq <- sum(xxsq)
52 betaincumbs <- yyxxsum/sumxxsq
53 betaincumbs
54 # beta = 0.0417
55 alphaincumbs <- ymean - (betaincumbs*xmean)
56 alphaincumbs
57 # alpha = 0.5790
58 # linear model: y = 0.5790 + 0.0417x
59 # check work
60 incumbreg <- lm(incumbents$voteshare ~ incumbents$difflog)
61 incumbreg
62
63 # make a scatterplot of the two variables and add the regression line
64 plot(incumbents$difflog, incumbents$voteshare,
65      xlab="Difference between incumbent and challenger campaign spending",
66      ylab="Incumbent vote share")
67 abline(a=0.5790, b=0.0417, col="red")
68
69 # save the residuals of the model in a separate object
70 residscheck <- residuals(incumbreg)
71 residscheck
72
73 # write the prediction equation
74
75 #####
76 # Problem 2:
77 #####
78 # x = difflog (explanatory)
79 # y = presvote (outcome)
80
81 # run a regression where the outcome variable is presvote and the explanatory
    variable is difflog
82 y2mean <- mean(incumbents$presvote)
83 y2sum <- sum(incumbents$presvote)
84 yy2 <- (incumbents$presvote) - y2mean
85 yyxxsum2 <- sum(xx*yy2)
86 betaincumbs2 <- yyxxsum2/sumxxsq
87 betaincumbs2
88 # beta = 0.0238
89 alphaincumbs2 <- y2mean - (betaincumbs2*xmean)
90 alphaincumbs2
91 # alpha = 0.5076
92 # linear model: y = 0.5076 + 0.0238x
93 # check work
94 incumbreg2 <- lm(incumbents$presvote ~ incumbents$difflog)
95 incumbreg2
96
97 # make a scatterplot of the two variables and add the regression line
98 plot(incumbents$difflog, incumbents$presvote,

```

```

99     xlab="Difference between incumbent and challenger campaign spending",
100     ylab="Vote share")
101 abline(a=0.5076, b=0.0238, col="red")
102 #save the residuals of the model in a separate object
103 resids2 <- residuals(incumbreg2)
104 resids2
105
106 # write the prediction equation
107
108 #####
109 # Problem 3:
110 #####
111 # x = presvote (explanatory)
112 # y = voteshare (outcome)
113
114 # run a regression where the outcome variable is voteshare and the explanatory
    variable is presvote
115 xmean3 <- mean(incumbents$presvote)
116 ymean3 <- mean(incumbents$voteshare)
117 xsum3 <- sum(incumbents$presvote)
118 ysum3 <- sum(incumbents$voteshare)
119 yy3 <- (incumbents$voteshare) - (ymean3)
120 xx3 <- (incumbents$presvote) - (xmean3)
121 yyxxsum3 <- sum(yy3*xx3)
122 xxsq3 <- (xx3)^2
123 sumxxsq3 <- sum(xxsq3)
124 beta3 <- yyxxsum3/sumxxsq3
125 beta3
126 # beta = 0.3880
127 alpha3 <- ymean3 - (beta3*xmean3)
128 alpha3
129 # alpha = 0.4413
130 # linear model: y= 0.4413 + 0.3880x
131 # check work
132 reg3 <- lm(incumbents$voteshare~incumbents$presvote)
133 reg3
134
135 # make a scatterplot of the two variables and add the regression line
136 plot(incumbents$presvote, incumbents$voteshare,
137      xlab="presvote", ylab="voteshare")
138 abline(a=0.4413, b=0.3880, col="red")
139
140 # write the prediction equation
141
142 #####
143 # Problem 4:
144 #####
145 # x = resids2 (explanatory)
146 # y = residscheck (outcome)
147

```

```

148 # run a regression where the outcome variable is the residuals from Q1 and the
      explanatory variable is the residuals from Q2
149 xmean4 <- mean(resids2)
150 ymean4 <- mean(residscheck)
151 xsum4 <- sum(resids2)
152 ysum4 <- sum(residscheck)
153 yy4 <- residscheck - (ymean4)
154 xx4 <- resids2 - (xmean4)
155 yyxxsum4 <- sum(yy4*xx4)
156 xxsq4 <- (xx4)^2
157 sumxxsq4 <- sum(xxsq4)
158 beta4 <- yyxxsum4/sumxxsq4
159 beta4
160 # beta = 0.2569
161 alpha4 <- ymean4 - (beta4*xmean4)
162 alpha4
163 # alpha = -4.939 x 10^(-18)
164 # linear model: y = -4.939 x 10^(-18) + 0.2569x
165 # check work
166 reg4 <- lm(residscheck ~ resids2)
167 reg4
168 summary(reg4)
169
170 # make a scatterplot of the two residuals and add the regression line
171 plot(resids2, residscheck,
172      xlab="Q2 Residuals", ylab="Q1 Residuals")
173 abline(reg4, col="red")
174
175 # write the prediction equation
176
177 #####
178 # Problem 5:
179 #####
180 # x = difflog, presvote (explanatory)
181 # y = voteshare (outcome)
182
183 # run a regression where the outcome variable is the incumbent's voteshare and
      the explanatory variables are difflog and presvote
184 lmbyhand <- function(inputDF, covariates, outcome){
185   X2 <- as.matrix(cbind(rep(1, dim[inputDF][1]), inputDF[, covariates]))
186   Y2 <- inputDF[, outcome]
187   betas2 <- solve((t(X2)%*%X2)) %*% (t(X2)%*%Y2)
188   rownames(betas2)[1] <- "Intercept"
189   n <- dim(inputDF)[1]
190   k <- ncol(X2)}
191 lmbyhand(incumbents, c("difflog", "presvote"), "voteshare")
192 reg_results
193 # check work
194 lm(incumbents$voteshare ~ incumbents$difflog + incumbents$presvote)
      $coefficients
195 multireg2 <- lm(incumbents$voteshare ~ incumbents$difflog +

```

```

    incumbents$presvote)
196 summary(multireg2)
197 # linear model:  $y = 0.4486 + 0.0355x_{\text{difflog}} + 0.2569x_{\text{presvote}}$ 
198
199 # write the prediction equation
200
201 # what is it in this output that is identical to the output in Q4? why do you
    think this is the case?
202 # the slope of the regression of voteshare on (presvote and difflog) is the
    same as the slope of of the regression of Q1 residuals on Q2 residuals.
203 # i think this is the case because the regression of Q1 residuals on Q2
    residuals tells us how much variation in presvote
204 # *and* voteshare is not explained by difflog.
205
206 # the relationship between variation in presvote/voteshare that is not
    explained by difflog.

```