# 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`.

```
# Running linear regression of vote share of incumbent on diff in
    campaign spending
reg_model_q1 <- lm(voteshare ~ difflog, data = incumbents)
reg_model_q1
summary(reg_model_q1) # Summary statistics for the model
```

(The question didn't specify units, so I just assumed it was %)

Using the estimated coefficients and the summary statistics from the linear regression model, we can calculate the fitted model as $\hat{y} = 0.579 + 0.042x$. The slope of the fitted model is 0.042, and we can interpret the slope as such: when the difference in campaign spending between incumbents and challengers increases by 1%, the incumbent's vote share increases by 4.2%. The intercept of the fitted model is 0.579 and we can interpret it as such: when the difference in campaign spending between incumbents and challengers is 0, the incumbent's vote share will be approximately 57.9%.

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

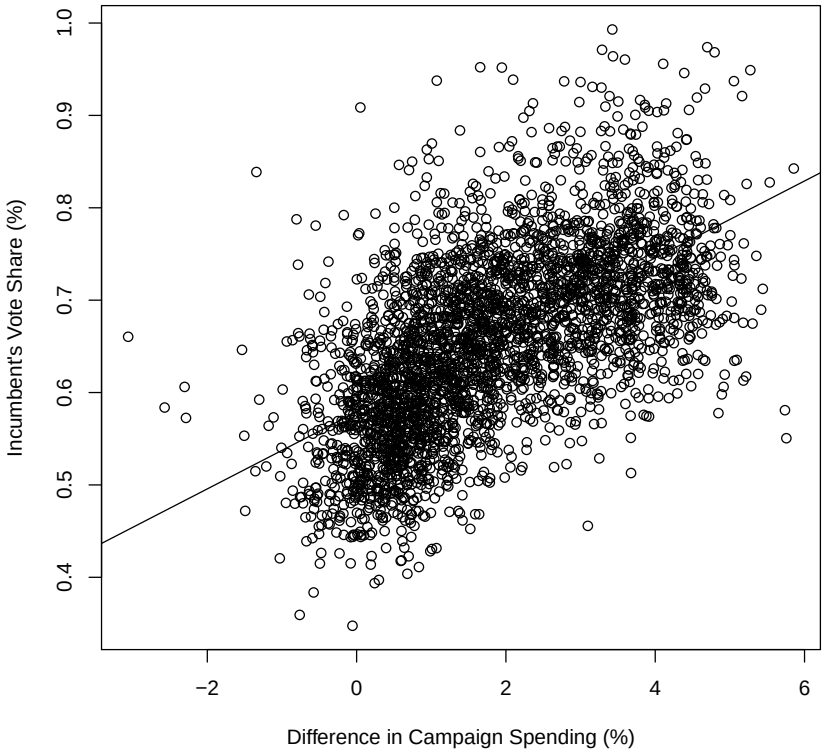


Figure 1: Scatterplot of Incumbent's Vote Share vs. Difference in Campaign Spending

```r
# Plot of diff in campaign spending vs vote share of incumbent
pdf("plotq1.pdf")
plot(incumbents$difflog, incumbents$voteshare, xlab = "Difference in
    Campaign Spending (%)", ylab = "Incumbent's Vote Share (%)")
abline(reg_model_q1) # Adding regression line
dev.off()
cor(incumbents$difflog, incumbents$voteshare) # Correlation coefficient =
    0.606
```

There seems to be a positive, moderately linear relationship between the two variables, which can be confirmed by the correlation coefficient which equals 0.606.

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

```
q1_resid <- reg_model_q1$residuals # Storing residuals as separate object
```

4. Write the prediction equation.

   $\hat{y} = 0.579 + 0.042x$

# 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`.

```
# Running linear regression of vote share of the presidential candidate
    of the incumbent's party on diff in campaign spending
reg_model_q2 <- lm(presvote ~ difflog, data = incumbents)
reg_model_q2
summary(reg_model_q2) # Summary statistics for the model
```

   Using the estimated coefficients and the summary statistics from the linear regression model, we can calculate the fitted model as $\hat{y} = 0.508 + 0.024x$. The slope of the fitted model is 0.024 and we can interpret the slope as such: when the difference in campaign spending between incumbents and challengers increases by 1%, the vote share of the presidential candidate of the incumbent's party increases by 2.4%. The intercept of the fitted model is 0.508 and we can interpret it as such: when the difference in campaign spending between incumbents and challengers is 0, the presidential candidate's vote share will be approximately 50.8%

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

```
# Plot of diff in campaign spending vs vote share of the presidential
    candidate of the incumbent's party
pdf("plotq2.pdf")
plot(incumbents$difflog, incumbents$presvote, xlab = "Difference in
    Campaign Spending (%)", ylab = "Vote Share of Pres. Candidate of
    Incumbent's Party (%)")
abline(reg_model_q2) # Adding regression line
dev.off()
cor(incumbents$difflog, incumbents$presvote) # Correlation coefficient =
    0.297
```

   There seems to be a positive, weak linear relationship between the two variables, which can be confirmed by the correlation coefficient which is approximately 0.297.

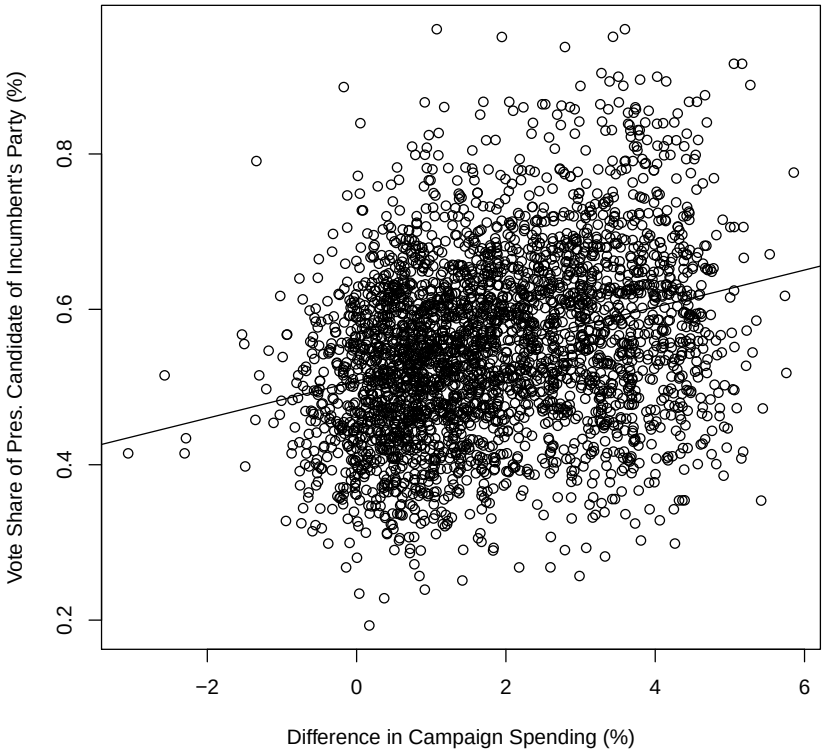


Figure 2: Scatterplot of Vote Share of Presidential Candidate in Incumbent's Party vs. Difference in Campaign Spending

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

```
q2_resid <- reg_model_q2$residuals # Storing residuals as separate object
```

4. Write the prediction equation.

   $\hat{y} = 0.508 + 0.024x$

# 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`.

```
# Running linear regression of incumbent's vote share on vote share of
    the presidential candidate of the incumbent's party
reg_model_q3 <- lm(voteshare ~ presvote, data = incumbents)
```

```
reg_model_q3
summary(reg_model_q3) # Summary statistics for the model
```

Using the estimated coefficients and the summary statistics from the linear regression model, we can calculate the fitted model as $\hat{y} = 0.441 + 0.388x$. The slope of the fitted model is 0.388, and we can interpret the slope as such: when the vote share of the presidential candidate of the incumbent's party increases by 1%, the incumbent's vote share increases by 38.8%. The intercept of the fitted model is 0.441 and we can interpret it as such: when vote share of the presidential candidate is zero, the incumbent's vote share will be approximately 44.1%.

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

```
# Plot of vote share of the presidential candidate of the incumbent's
    party vs incumbent's vote share
pdf("plotq3.pdf")
plot(incumbents$voteshare, incumbents$presvote, xlab = "Vote Share of
    Pres. Candidate of Incumbent's Party (%)", ylab = "Incumbent's Vote
    Share (%)")
abline(reg_model_q3) # Adding regression line
dev.off()
cor(incumbents$voteshare, incumbents$presvote) # Correlation coefficient
    = 0.454
```

There seems to be a positive, moderate linear relationship between the two variables, which can be confirmed by the correlation coefficient which is approximately 0.454.

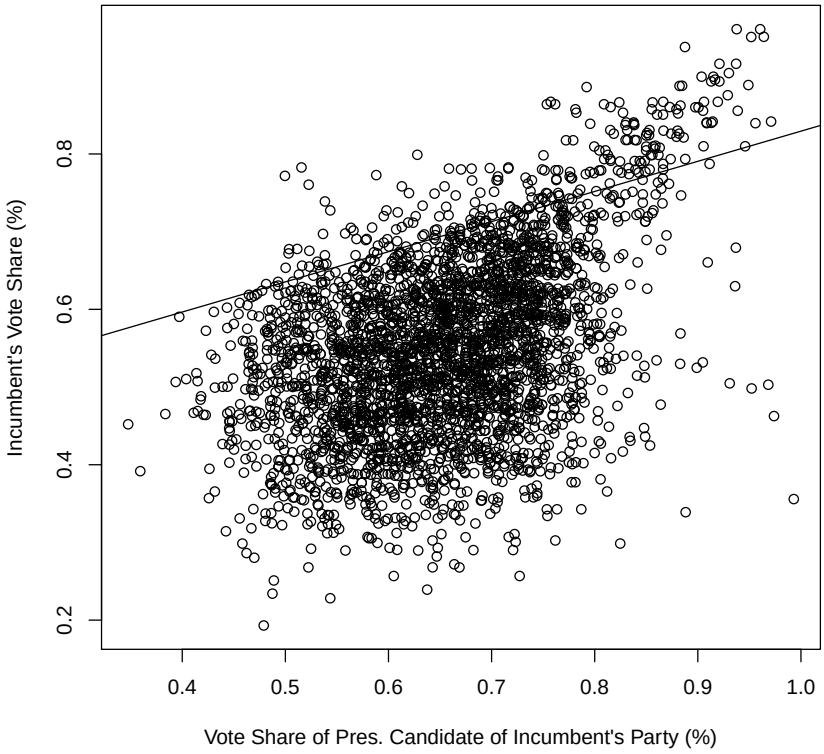


Figure 3: Scatterplot of Incumbent's Vote Share vs. Vote Share of Presidential Candidate in Incumbent's Party

3. Write the prediction equation.

   $\hat{y} = 0.441 + 0.388x$

# 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.

```r
# Running linear regression of residuals from Q1 on residuals from Q2
reg_model_q4 <- lm(q1_resid ~ q2_resid, data = incumbents)
reg_model_q4
summary(reg_model_q4) # Summary statistics for the model
```

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

```
# Plot of the two residuals
pdf("plotq4.pdf")
plot(q2_resid, q1_resid, xlab = "Residuals from Question 2", ylab = "
    Residuals from Question 1")
abline(reg_model_q4) # Adding regression line
dev.off()
cor(q2_resid, q1_resid) # Correlation coefficient = 0.361
```

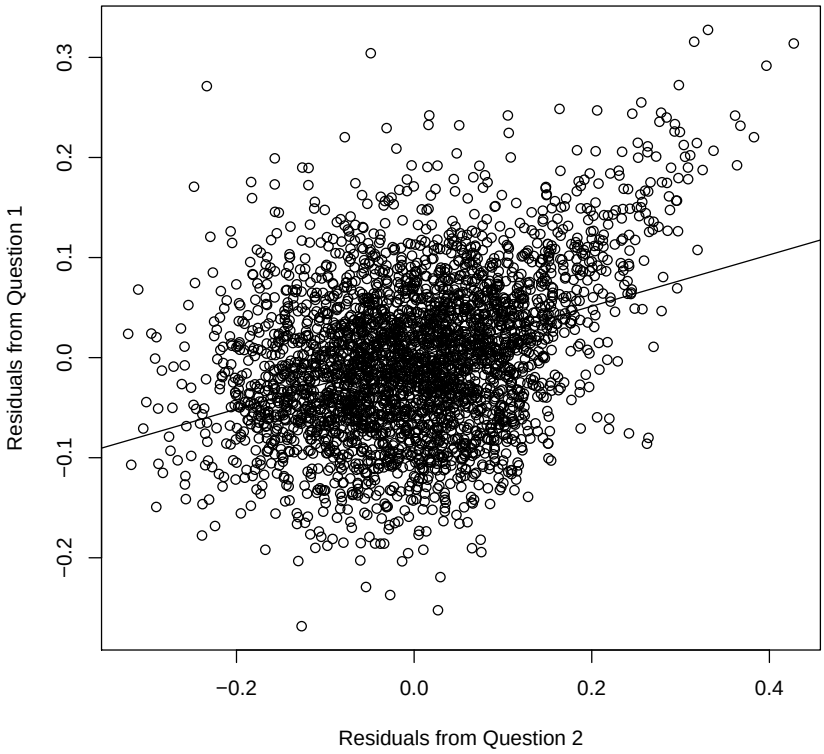


Figure 4: Scatterplot of Q1 Residuals vs. Q2 Residuals

There seems to be a positive, moderate linear relationship between the two variables, which can be confirmed by the correlation coefficient which is approximately 0.361.

3. Write the prediction equation.

   $\hat{y}$ = -4.86e-18 + 0.26x

# 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`.

```
# Running linear regression of voteshare on difflog and presvote
reg_model_q5 <- lm(incumbents$voteshare ~ incumbents$difflog + incumbents
    $presvote)
reg_model_q5
summary(reg_model_q5) # Summary statistics of the model
```

   Using the estimated coefficients and the summary statistics from the linear regression model, we can calculate the fitted model as $\hat{y} = 0.449 + 0.036x_1 + 0.257x_2$. The first slope of the fitted model is 0.036, and we can interpret the slope as such: when the difference in campaign spending ncreases by 1%, the incumbent's vote share increases by 3.6%. The second slope of the fitted model is 0.257 and it can be interpreted as such: when the vote share of the presidential candidate of the incumbent's party increases by 1%, the incumbent's vote share increases by 25.7%. The intercept of the fitted model is 0.449 and we can interpret it as such: when the differences in campaign spending is 0 and the vote share of the presidential candidate is 0, the incumbent's vote share will be approximately 44.9%.

2. Write the prediction equation.

   $\hat{y} = 0.449 + 0.036x_1 + 0.257x_2$

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 outputs are similar in that they have the same residuals.