Problem Set 3

Applied Stats/Quant Methods 1

Charis Rogers (19333698)

Question 1

We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

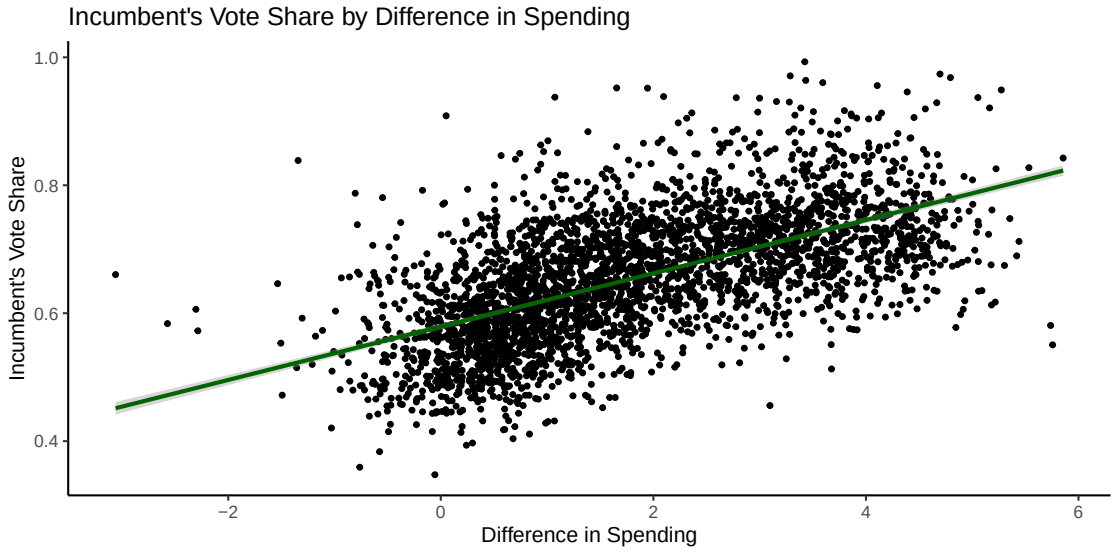
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
q1 <- lm(voteshare~difflog, data=inc.sub)
summary(q1)
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.579031   0.002251  257.19   <2e-16 ***
#difflog     0.041666   0.000968   43.04   <2e-16 ***
```

   Running this regression we can see that there is a significant linear relationship between the difference in spending between the parties and the vote share. For every unit increase in `difflog` there is about a 4% increase in the incumbent's vote share.

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

```
scatterq1<-
  ggplot(data = inc.sub,
         mapping = aes(x = difflog,
                       y = voteshare)) +
  geom_point(size = 1) +
  geom_smooth(method='lm',col="darkgreen") + #Regression line
  labs(x = "Difference in Spending", #Labels
       y = "Incumbent's Vote Share",
       title =  "Incumbent's Vote Share by Difference in Spending") +
  theme_classic()  #Theme
```



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

```
residualsq1 <- lm(voteshare~difflog, data=inc.sub)$residuals
residualsq1
```

4. Write the prediction equation.

```
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.579031   0.002251  257.19   <2e-16 ***
#difflog     0.041666   0.000968   43.04   <2e-16 ***
```

From this regression we have run we can see the y intercept and slope of the line so the prediction equation is Incumbent's Vote Share = 0.579031 + 0.041666*Difference in Spending

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.

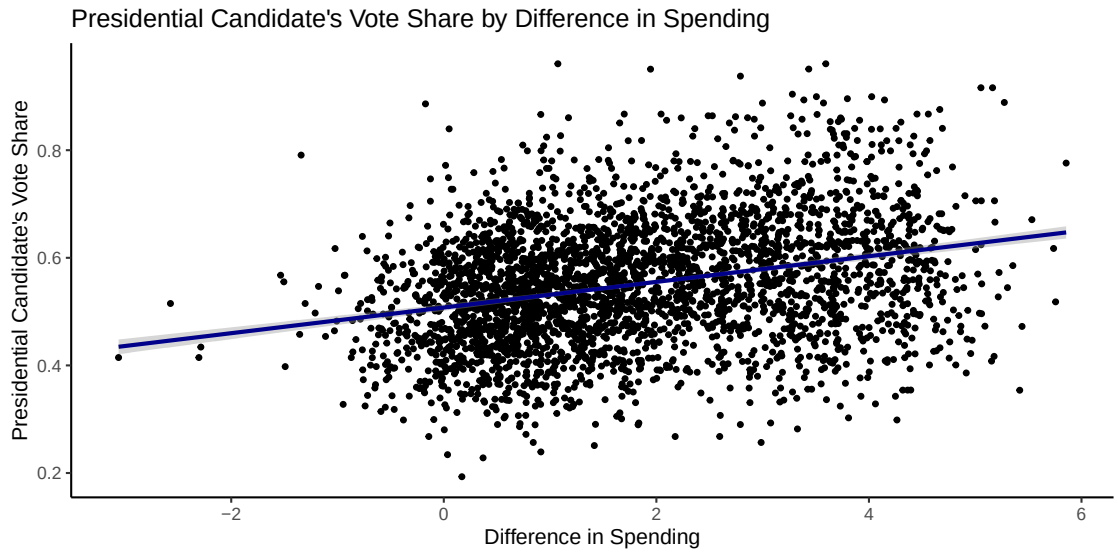
1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```
q2 <- lm(presvote~difflog, data=inc.sub)
summary(q2)
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.507583   0.003161  160.60   <2e-16 ***
#difflog     0.023837   0.001359   17.54   <2e-16 ***
```

   From this regression we can see that there is a significant linear relationship between the difference in spending and the vote share of the incumbent party's candidate. With a 1 unit increase in `difflog` the presidential candidate's vote share increases by about 2%.

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

```
scatterq2<-
  ggplot(data = inc.sub,
         mapping = aes(x = difflog,
                       y = presvote)) +
  geom_point(size = 1) +
  geom_smooth(method='lm',col="darkblue") + #Regression line
  labs(x = "Difference in Spending", #Labels
       y = "Presidential Candidate's Vote Share",
       title =  "Presidential Candidate's Vote Share by Difference in
    Spending") +
  theme_classic()  #Theme
```



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

```
residualsq2 <- lm(presvote~difflog, data=inc.sub)$residuals
residualsq2
```

4. Write the prediction equation.

```
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.507583   0.003161  160.60   <2e-16 ***
#difflog     0.023837   0.001359   17.54   <2e-16 ***
```

After running the regression on these variables we can see that the prediction equation is Presidential Candidate's Vote Share = 0.507583 + 0.023837*Difference in Spending

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.

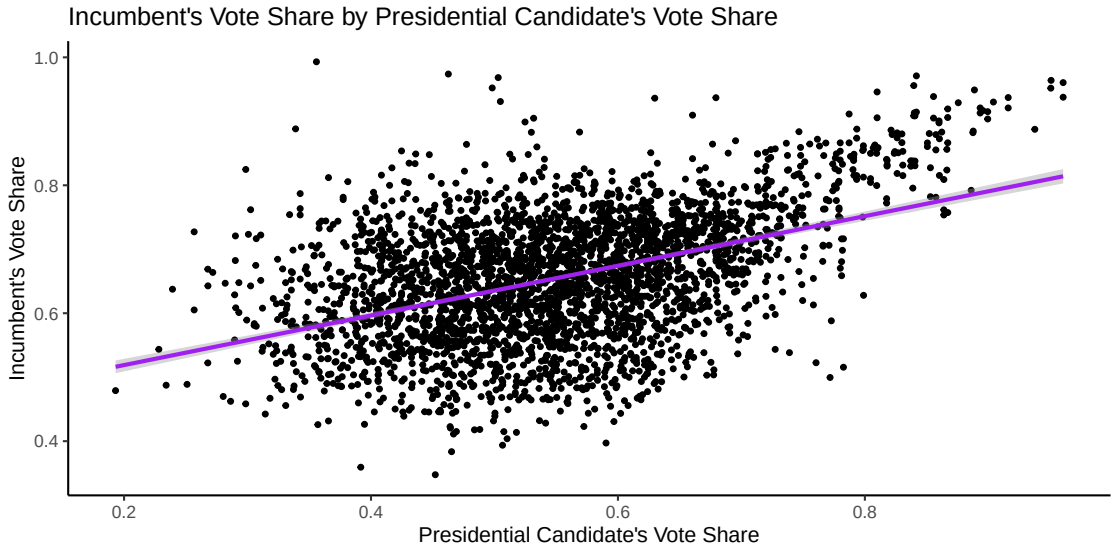
1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`.

```
q3 <- lm(voteshare~presvote, data=inc.sub)
summary(q3)
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.441330   0.007599   58.08   <2e-16 ***
#presvote    0.388018   0.013493   28.76   <2e-16 ***
```

   Running this regression we can see that there is a significant linear relationship between the vote share of the incumbent party and the presidential candidate's vote share. When the presidential candidate's vote share increases by 1 unit the incumbent's vote share increases by about 39%.

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

```
scatterq3<-
  ggplot(data = inc.sub,
         mapping = aes(x = presvote,
                       y = voteshare)) +
  geom_point(size = 1) +
  geom_smooth(method='lm',col="purple") + #Regression line
  labs(x = "Presidential Candidate's Vote Share", #Labels
       y = "Incumbent's Vote Share",
       title =  "Incumbent's Vote Share by Presidential Candidate's Vote
    Share") +
  theme_classic()  #Theme
```



3. Write the prediction equation.

```
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.441330   0.007599   58.08   <2e-16 ***
#presvote    0.388018   0.013493   28.76   <2e-16 ***
```

From this regression we can see that the prediction equation for these two variables is Incumbent's Vote Share = 0.441330 + 0.388018 * Presidential Candidate's Vote Share

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.

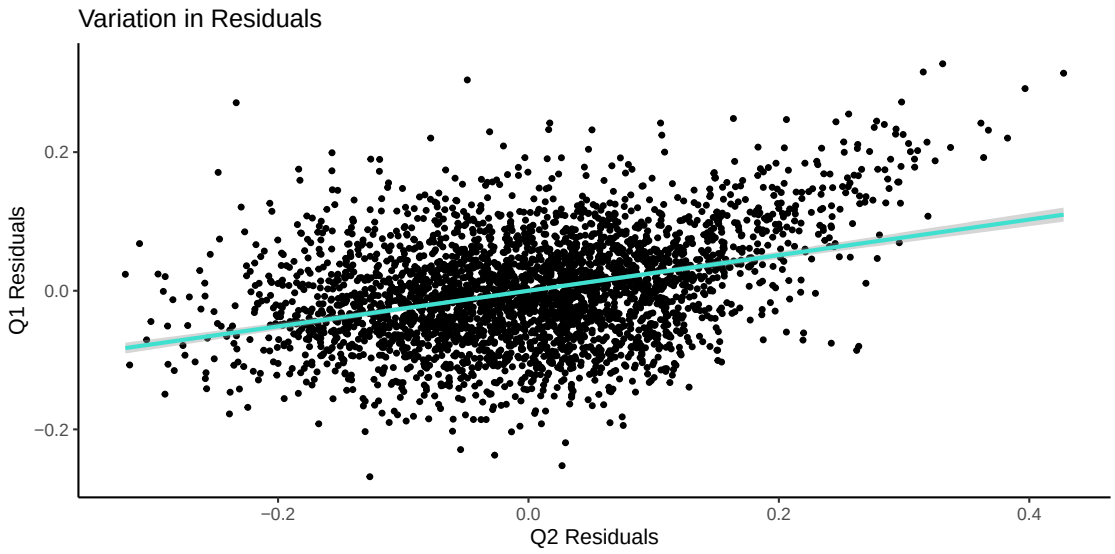
1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
q4 <- lm(residualsq1~residualsq2)
summary(q4)
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept) -5.934e-18  1.299e-03    0.00        1
#residualsq2  2.569e-01  1.176e-02   21.84   <2e-16 ***
```

   From this regression we can see that the residuals from the two regressions above are statistically associated. This means that there is a relationship between the variation in vote share (`voteshare`) not explained by difference in spending and the presidential candidate's vote share (`presvote`) not explained by difference in spending. As the residuals from Question 2 increase by 1 unit, the residuals from Question 2 increase by 26%.

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

```
scatterq4<-
  ggplot(data = inc.sub,
         mapping = aes(x = residualsq2,
                       y = residualsq1)) +
  geom_point(size = 1) +
  geom_smooth(method='lm',col="turquoise") + #Regression line
  labs(x = "Q2 Residuals", #Labels
       y = "Q1 Residuals",
       title =  "Variation in Residuals") +
  theme_classic()  #Theme
```



3. Write the prediction equation.

```
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept) -5.934e-18  1.299e-03    0.00        1
#residualsq2  2.569e-01  1.176e-02   21.84   <2e-16 ***
```

As calculated above, the prediction equation for the relationship between the residuals is Residuals Q1 = -5.934e-18 + 2.569e-01 * Residuals Q2 which can be approximated to Residuals Q1 = 0.2569 * Residuals Q2

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`.

```
q5 <- lm(voteshare~difflog+presvote, data=inc.sub)
summary(q5)
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.4486442  0.0063297   70.88   <2e-16 ***
#difflog     0.0355431  0.0009455   37.59   <2e-16 ***
#presvote    0.2568770  0.0117637   21.84   <2e-16 ***
```

   This regression shows that both of these explanatory variables are significant, even when the other is controlled for. When vote share increases by one unit the difference in spending increases by 3.6%, while controlling for the president's popularity. When vote share increases by one unit the president's popularity increases by 25.7%, while controlling for vote share.

2. Write the prediction equation.

   Incumbent's Vote Share = 0.4486442 + 0.0355431 * Difference in Spending + 0.2568770 * Presidential Candidate's Vote Share

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 slope for the presidential candidate's vote share, controlling for difference in spending, is identical to the slope for the residuals in Q2, the variation in `presvote` *not* explained by the difference in spending. This is because both of these equations relate to the slope of `presvote` which explains vote share. The slope is identical as it concerns (1) when `difflog` is controlled or (2) looking at the variation *not* explained by `difflog`, which is the same thing.