Submission for Problem Set 3

Applied Stats/Quant Methods 1

Duc Minh, VU
TCD StudentID: 22996761 / UCD StudentID: 19211157

Question 1

The incumbent dataset will first be imported and the relevant library will be loaded in R for analysis.

```r
incumb_data <- read.csv("incumbents_subset.csv")
library(ggplot2)
library(texreg)
```

1. Regression modelling with ***voteshare*** as the outcome variable and the ***difflog*** as explanatory variable

```r
vote.lm <- lm(data = incumb_data, voteshare ~ difflog)
summary(vote.lm)
texreg(vote.lm,
       caption = "Vote share (voteshare) and log differences in campaign
   spending (difflog)",
       custom.model.names = "Model 1",
       float.pos = "H", digits = 4)
```

```
Call:
lm(formula = voteshare ~ difflog, data = incumb_data)

Residuals:
Min        1Q   Median       3Q      Max
-0.26832 -0.05345 -0.00377  0.04780  0.32749

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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.07867 on 3191 degrees of freedom
Multiple R-squared:  0.3673,Adjusted R-squared:  0.3671
F-statistic:  1853 on 1 and 3191 DF,  p-value: < 2.2e-16
```

	Model 1
(Intercept)	0.5790^{***}
	(0.0023)
difflog	0.0417^{***}
	(0.0010)
R^2	0.3673
Adj. R^2	0.3671
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

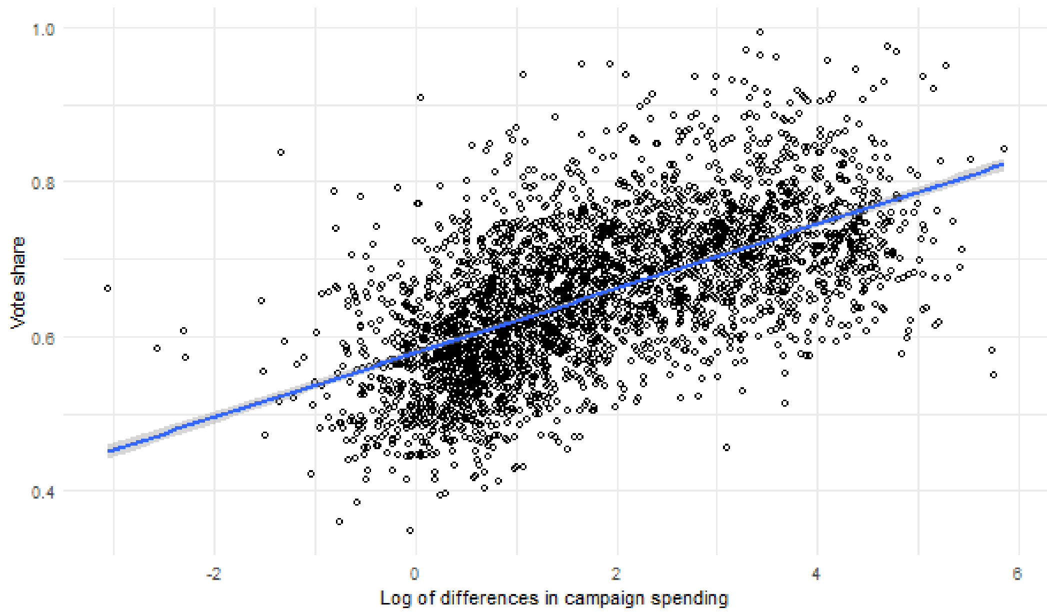
Table 1: Vote share (voteshare) and log differences in campaign spending (difflog)

There is statistical evidence that there is a positive relationship between incumbent's voteshare and the difference in campaign spending between incumbent and challenger. For a one unit increase in the logged difference in spending, the incumbent's voteshare is predicted to increase, on average, by 0.04.

2. Create a scatterplot for the two variables

```r
ggplot(data = incumb_data, aes(x = difflog, y = voteshare)) +
  geom_point(shape=1) +
  geom_smooth(method = "lm") +
  ylab("Vote share") + xlab("Log of differences in campaign spending") +
  theme_minimal()
```

Figure 1: Scatterplot between voteshare and difflog



3. Save the residuals

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

4. Prediction equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x$$

$$\boldsymbol{\mathtt{voteshare}} = 0.579 + 0.042 \times \boldsymbol{\mathtt{difflog}}$$

Question 2

1. Regression modelling with ***presvote*** as the outcome variable and the ***difflog*** as explanatory variable

```
presvote.lm <- lm(data = incumb_data, presvote ~ difflog)
summary(presvote.lm)
texreg(presvote.lm,
       caption = "Vote share of presidental candidate (presvote) and log
    differences in campaign spending (difflog)",
       custom.model.names = "Model 2",
       float.pos = "H", digits = 4)
```

```
lm(formula = presvote ~ difflog, data = incumb_data)

Residuals:
Min       1Q   Median      3Q      Max
```

```
-0.32196 -0.07407 -0.00102  0.07151  0.42743

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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1104 on 3191 degrees of freedom
Multiple R-squared:  0.08795,Adjusted R-squared:  0.08767
F-statistic: 307.7 on 1 and 3191 DF,  p-value: < 2.2e-16
```

	Model 2
(Intercept)	0.5076***
	(0.0032)
difflog	0.0238***
	(0.0014)
R^2	0.0880
Adj. R^2	0.0877
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

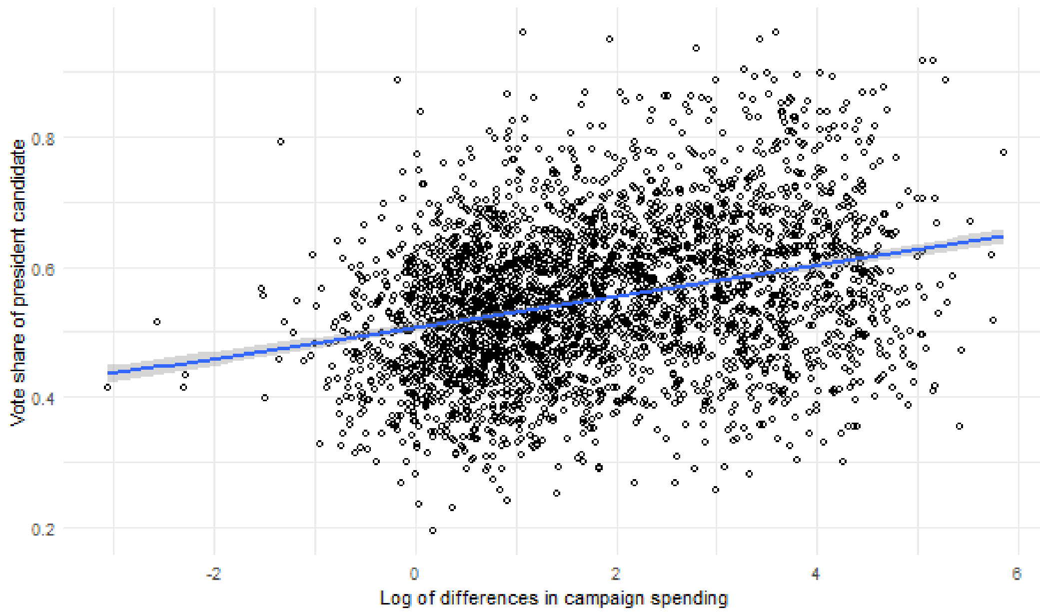
Table 2: Vote share of presidental candidate (presvote) and log differences in campaign spending (difflog)

There is statistical evidence that there is a postive relationship between vote share of the presidential candidate and the difference in spending between incumbent and challengers. For a one unit increase in the logged difference in spending, the incumbent's voteshare of the presidential candidate is predicted to increase, on average, by 0.0238.

2. Create a scatterplot for the two variables

```r
ggplot(data = incumb_data, aes(x = difflog, y = presvote)) +
  geom_point(shape=1) +
  geom_smooth(method = "lm") +
  ylab("Vote share of president candidate") + xlab("Log of differences in campaign spending") +
  theme_minimal()
```

Figure 2: Scatterplot between presvote and difflog



3. Save the residuals

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

4. Prediction equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x$$

$$\boldsymbol{\mathit{presvote}} = 0.5076 + 0.0238 \times \boldsymbol{\mathit{difflog}}$$

Question 3

1. Regression modelling with ***voteshare*** as the outcome variable and the ***presvote*** as explanatory variable

```
vote.lm.2 <- lm(data = incumb_data, voteshare ~ presvote)
summary(vote.lm.2)
texreg(vote.lm.2,
       caption = "Incumbent's electoral success (voteshare) and Vote
    share of presidental candidate (presvote)",
       custom.model.names = "Model 3",
       float.pos = "H", digits = 4)
```

```
lm(formula = voteshare ~ presvote, data = incumb_data)

Residuals:
Min        1Q   Median       3Q      Max
```

```
-0.27330 -0.05888  0.00394  0.06148  0.41365

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 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08815 on 3191 degrees of freedom
Multiple R-squared:  0.2058,Adjusted R-squared:  0.2056
F-statistic:   827 on 1 and 3191 DF,  p-value: < 2.2e-16
```

	Model 3
(Intercept)	0.4413***
	(0.0076)
presvote	0.3880***
	(0.0135)
R^2	0.2058
Adj. R^2	0.2056
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

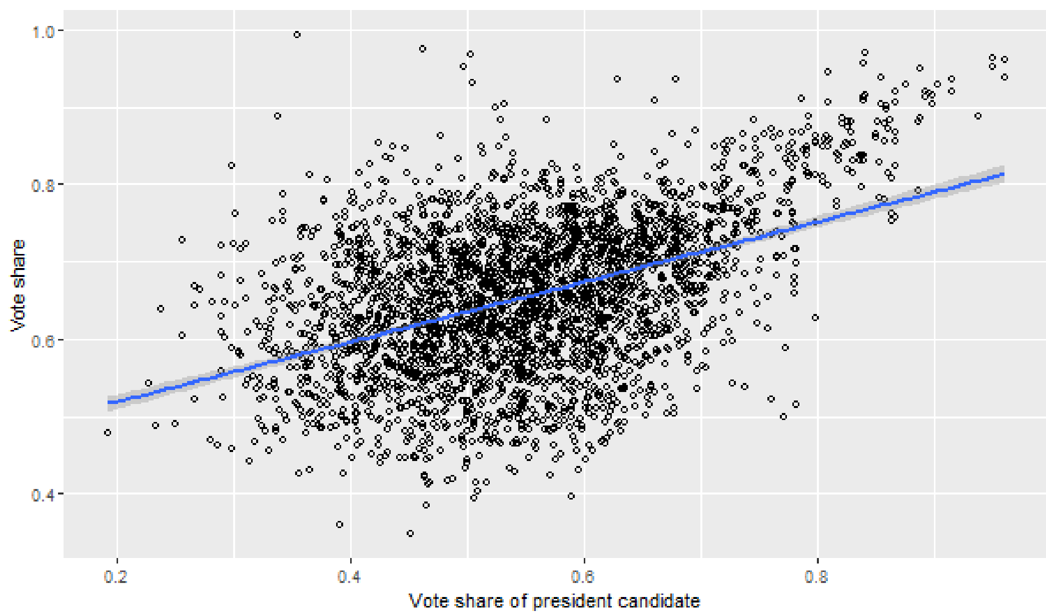
Table 3: Incumbent's electoral success (voteshare) and Vote share of presidental candidate (presvote)

There is statistical evidence that there is a positive relationship between vote share of the presidential candidate and vote share. For a one unit increase the vote share of the presidential candidate, the incumbent's voteshare is predicted to increase, on average, by 0.388

2. Create a scatterplot for the two variables

```
ggplot(data = incumb_data, aes(x = presvote, y = voteshare)) +
  geom_point(shape=1) +
  geom_smooth(method = "lm") +
  ylab("Vote share") + xlab("Vote share of president candidate") +
  theme_minimal()
```

Figure 3: Scatterplot between voteshare and presvote



3. Prediction equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x$$

$$\texttt{voteshare} = 0.4413 + 0.388 \times \texttt{presvote}$$

Question 4

1. Regression modelling with `Q1's residuals` as the outcome variable and the `Q2's residuals` as explanatory variable

```
resid.lm <- lm(vote.lm.resid ~ presvote.lm.resid)
summary(resid.lm)
texreg(resid.lm,
       caption = "Question 1 residuals (vote.lm.resid) and Question 2
   residuals (presvote.lm.resid)",
       custom.model.names = "Model 4",
       float.pos = "H", digits = 4)
```

```
lm(formula = vote.lm.resid ~ presvote.lm.resid)

Residuals:
Min        1Q   Median       3Q      Max
-0.25928 -0.04737 -0.00121  0.04618  0.33126

Coefficients:
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)      -4.860e-18  1.299e-03    0.00        1
presvote.lm.resid  2.569e-01  1.176e-02   21.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07338 on 3191 degrees of freedom
Multiple R-squared:   0.13,Adjusted R-squared:  0.1298
F-statistic:   477 on 1 and 3191 DF,  p-value: < 2.2e-16
```

	Model 4
(Intercept)	-0.0000
	(0.0013)
presvote.lm.resid	0.2569^{***}
	(0.0118)
R^2	0.1300
Adj. R^2	0.1298
Num. obs.	3193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

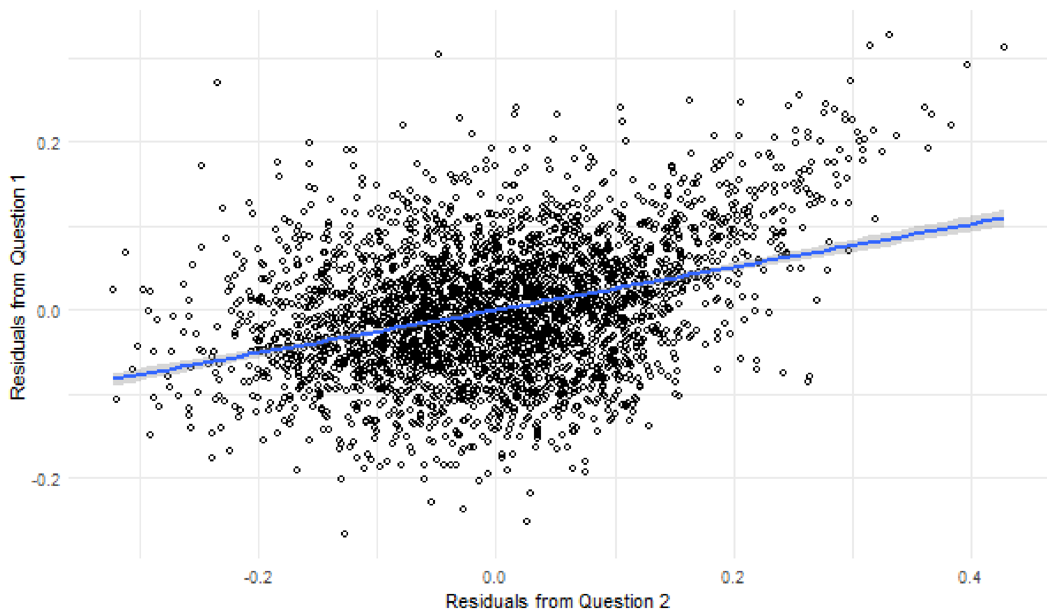
Table 4: Question 1 residuals (vote.lm.resid) and Question 2 residuals (presvote.lm.resid)

There is statistical evidence that there is a positive relationship between residuals from Question 1 and residuals from Question 2. For a one unit increase the residuals from question 1, the residuals from question 2 is predicted to increase, on average, by 0.2569.

2. Create a scatterplot for the two variables

```
ggplot(data = incumb_data, aes(x = presvote.lm.resid, y = vote.lm.resid))
    +
  geom_point(shape=1) +
  geom_smooth(method = "lm") +
```

Figure 4: Scatterplot between residuals from Question 1 and 2



3. Prediction equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x$$

$$\text{Q1's residuals} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{Q2's residuals}$$

$$\texttt{vote.lm.resid} = 0 + 0.2569 \times \texttt{presvote.lm.resid}$$

Question 5

1. Regression modelling with `voteshare` as the outcome variable and, `difflog` and `presvote` as explanatory variables

```
vote.lm.3 <- lm(data = incumb_data, voteshare ~ difflog + presvote)
summary(vote.lm.3)
texreg(vote.lm.3,
       caption = "Vote shares (voteshare) with difference in spending (
   difflog) and president's popularity (presvote)",
       custom.model.names = "Model 5",
       float.pos = "H", digits = 4)
```

```
lm(formula = voteshare ~ difflog + presvote, data = incumb_data)

Residuals:
Min        1Q   Median       3Q      Max
-0.25928 -0.04737 -0.00121  0.04618  0.33126
```

```
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 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

	Model 5
(Intercept)	0.4486***
	(0.0063)
difflog	0.0355***
	(0.0009)
presvote	0.2569***
	(0.0118)
R^2	0.4496
Adj. R^2	0.4493
Num. obs.	3193

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 5: Vote shares (voteshare) with difference in spending (difflog) and president's popularity (presvote)

2. Prediction equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times x_1 + \hat{\beta}_2 \times x_2$$

$$\texttt{voteshare} = 0.4486 + 0.0355 \times \texttt{difflog} + 0.2569 \times \texttt{presvote}$$

3. Comparison

```
texreg(list(resid.lm, vote.lm.3),
       custom.model.names = c("Model 4","Model 5"),
       caption = "Comparison between model from Question 4 and 5",
       float.pos = "H", digits = 4)
```

	Model 4	Model 5
(Intercept)	-0.0000	0.4486^{***}
	(0.0013)	(0.0063)
presvote.lm.resid	0.2569^{***}	
	(0.0118)	
difflog		0.0355^{***}
		(0.0009)
presvote		0.2569^{***}
		(0.0118)
R^2	0.1300	0.4496
Adj. R^2	0.1298	0.4493
Num. obs.	3193	3193

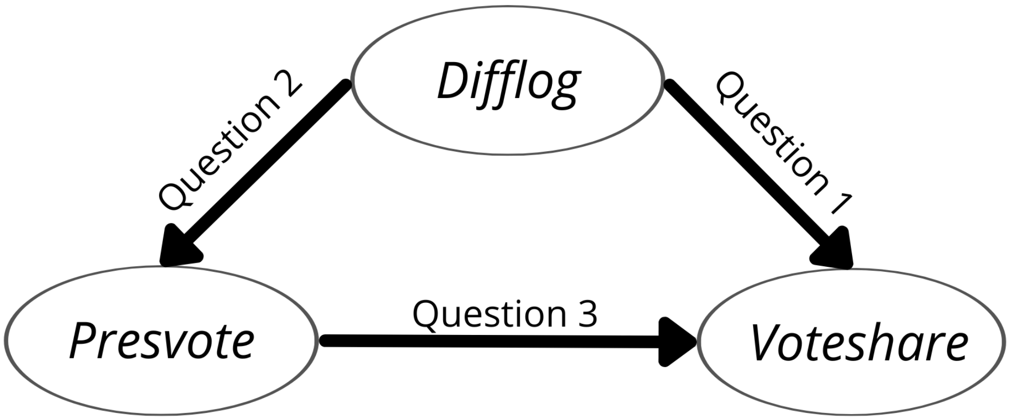
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Comparison between model from Question 4 and 5

From Table 6, we can see that both the coefficients for the residuals ***presvote.lm.resid*** from Model 4 of Question 4 and ***presvote*** from Model 5 of Question 5 have the same value of 0.2569.

First, let's look at the relationships between the three variables ***presvote***, ***voteshare*** and ***difflog***. Question 3 shows us that there is a relationship between ***presvote*** and ***voteshare***. But, at the same time, we also know that ***difflog*** have a relationship with both ***voteshare*** and ***presvote*** from Question 1 and 2 respectively. See Figure 5 below which I have constructed manually.

Figure 5: Graphical depiction of the relationship between *voteshare*, *presvote* and *difflog*



Hence, when using only *presvote* to predict *voteshare*, we cannot estimate its "pure" effect on *voteshare*, because there is a shared variance from both *presvote* and *difflog* (due to their relationship) in explaining *voteshare*. Hence, the residuals from Question 1 and 2 provide us with the variance in *voteshare* and *presvote* **unexplained** by *difflog*. In essence, by obtaining the residuals from these bivariate regressions, we have "clean" both *voteshare* and *presvote* of their correlation with *difflog*. Therefore, when we run the bivariate model between these residuals, we are getting the 'pure' (relatively as in not influenced by *difflog*) covaration between *voteshare* and *presvote*.

This is essentially what multiple regression do as it control for the effects of other variables on the dependent variables. In our case, the influence of *difflog* is controlled for when it has been included into the regression model, so we can get the partial effect of *presvote* only. Hence, this is why we have the same coefficients or to sum up, both the coefficients for the residuals *presvote.lm.resid* from Model 4 of Question 4 and *presvote* from Model 5 of Question 5 show us the partial effect of *presvote*.