

Problem Set 3

Ella Karagulyan

Applied Stats/Quant Methods 1

We read our data and explore the dataset.

```
1 # set wd for current folder
2 setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
3
4 # read in data
5 inc.sub <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/incumbents_subset.csv")
6
7 # Exploring the data frame and selected variables
8 head(inc.sub, n=10)
9 summary(inc.sub)
10 summary(inc.sub)
11 sum(is.na(inc.sub))
12 sum(is.na(inc.sub))
```

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**.

```
1 # Regressing the selected variables and saving the model as an object
2 modell <- lm(voteshare ~ difflog, data = inc.sub)
3 summary(modell)
4
5 # Creating a table of the results
6 #install.packages("stargazer")
7 library(stargazer)
8 stargazer(modell, type = "latex", title = "Linear Regression Results:
  Difference in campaign spending and vote share",
9           out = "regression_table_modell.tex")
```

The table below shows the results of bivariate regression of difference in campaign spending on incumbent and challenger and the incumbent's vote share. According to the results, there is a positive significant effect of campaign spending on incumbent's vote share, with a one unit increase in spending difference, the vote share increases by 0.04.

Table 1: Linear Regression Results: Difference in campaign spending and vote share

	<i>Dependent variable:</i>
	voteshare
difflog	0.042*** (0.001)
Constant	0.579*** (0.002)
Observations	3,193
R ²	0.367
Adjusted R ²	0.367
Residual Std. Error	0.079 (df = 3191)
F Statistic	1,852.791*** (df = 1; 3191)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

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

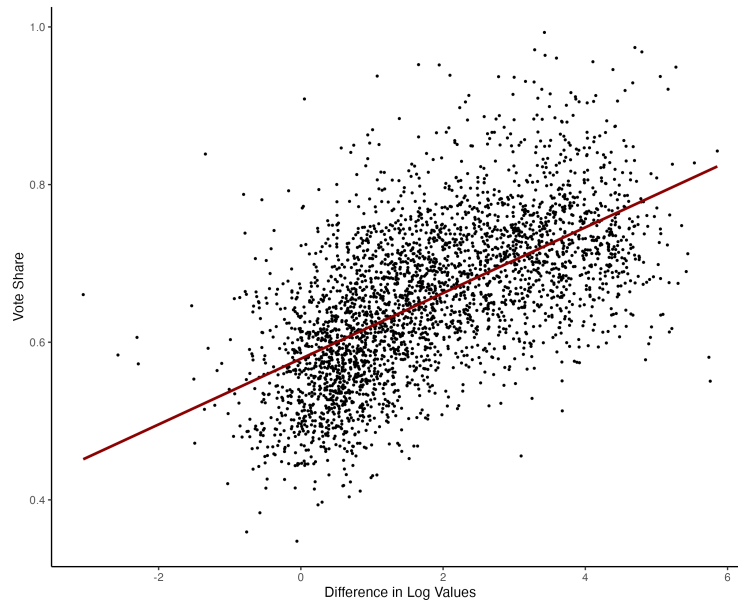
```

1 # Creating a scatterplot
2 library(ggplot2)
3 scatter1 <- ggplot(data = inc.sub,
4                     mapping = aes(x = difflog,
5                                   y = voteshare)) +
6   geom_point(size=0.5) +
7   geom_smooth(method = "lm", se = FALSE, color = "darkred") + # Add
8   regression line
9   labs(x = "Difference in Log Values", # Add labels
10        y = "Vote Share") +
11   theme_classic() + # Change theme
12   theme(legend.box.background = element_rect(size = 0.1), # Change
13         legend.position = c(0.85, 0.85)) # Change position of legend
14 # Printing and saving the scatterplot
15 scatter1
16 ggsave(scatter1, file = "scatter_vs_dl.png")

```

The figure below depicts the relationship between the outcome and input variables and the bivariate model regression line.

Figure 1: Difference in campaign spending and incumbent vote share



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

```
1 # Saving the residuals as an object and summarizing
2 residuals_voteshare_difflog <- residuals(modell)
3 summary(residuals_voteshare_difflog)
```

4. Write the prediction equation.

The bivariate regression tests how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

$$Y = \beta_0 + \beta_1 \cdot X$$

where:

Y (voteshare) - incumbent's vote share

X (difflog) - the difference in campaign spending between incumbent and challenger

$$voteshare = \beta_0 + \beta_1 \cdot (\text{difflog})$$

$$voteshare = 0.579 + 0.042 \cdot (\text{difflog})$$

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**.

```

1 # Regressing the selected variables and saving the model as an object
2 model2 <- lm(presvote ~ difflog, data = inc.sub)
3 summary(model2)
4
5 # Creating a table of the results
6 stargazer(model2, type = "latex", title = "Linear Regression Results:
  Difference in campaign spending
  and vote share of the presidential candidate",
7           out = "regression_table_model2.tex")
8

```

The table below shows the results of bivariate regression between the difference of incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party. According to the results, there is a positive significant effect of campaign spending on presidential candidate's vote share, with a one unit increase in spending, the vote share of the presidential candidate increases by 0.257.

Table 2: Linear Regression Results: Difference in campaign spending and vote share of the presidential candidate

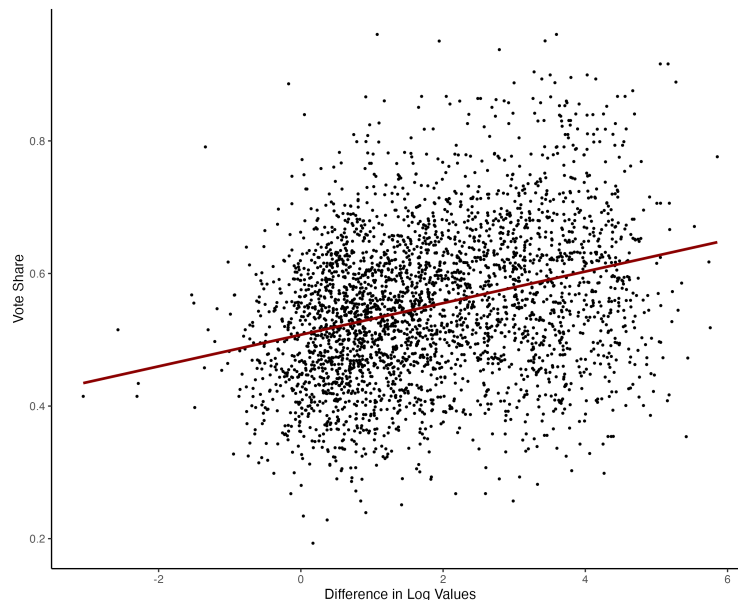
	<i>Dependent variable:</i>
	presvote
difflog	0.024*** (0.001)
Constant	0.508*** (0.003)
Observations	3,193
R ²	0.088
Adjusted R ²	0.088
Residual Std. Error	0.110 (df = 3191)
F Statistic	307.715*** (df = 1; 3191)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

```
1 # Creating a scatterplot
2 library(ggplot2)
3 scatter2 <- ggplot(data = inc.sub,
4                   mapping = aes(x = difflog,
5                                 y = presvote)) +
6   geom_point(size=0.5) +
7   geom_smooth(method = "lm", se = FALSE, color = "darkred") + # Add
8     regression line
9   labs(x = "Difference in Log Values", # Add labels
10        y = "Vote Share") +
11   theme_classic() + # Change theme
12   theme(legend.box.background = element_rect(size = 0.1), # Change
13         background
14         legend.position = c(0.85, 0.85)) # Change position of legend
15
14 # Printing and saving the scatterplot
15 scatter2
16 ggsave(scatter2, file = "scatter_pv_dl.png")
```

The figure below depicts the relationship between the outcome and input variables and the bivariate model regression line.

Figure 2: Difference in campaign spending and vote share of the presidential candidate



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

```
1 # Saving the residuals as an object and summarizing
2 residuals_presvote_difflog <- residuals(model2)
```

```
3 summary(residuals_presvote_difflog)
```

4. Write the prediction equation.

The bivariate regression tests how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

$$Y = \beta_0 + \beta_1 \cdot X$$

where:

Y (presvote) - vote share of the presidential candidate of the incumbent's party

X (difflog) - the difference in campaign spending between incumbent and challenger

$$\text{presvote} = \beta_0 + \beta_1 \cdot (\text{difflog})$$

$$\text{presvote} = 0.508 + 0.024 \cdot (\text{difflog})$$

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**.

```
1 # Regressing the selected variables and saving the model as an object
2 model3 <- lm(voteshare ~ presvote, data = inc.sub)
3 summary(model3)
4
5 # Creating a table of the results
6 stargazer(model3, type = "latex", title = "Linear Regression Results:
  Vote share of the presidential candidate and the incumbent",
7           out = "regression_table_model3.tex")
```

The table below shows the results of bivariate regression testing whether the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success. According to the results, there is a positive significant effect of the vote share of the presidential candidate of the incumbent's party on incumbent's

success, with a one unit increase in presidential candidate's vote share the incumbent's vote share increases by 0.39.

Table 3: Linear Regression Results: Vote share of the presidential candidate and the incumbent

<i>Dependent variable:</i>	
voteshare	
presvote	0.388*** (0.013)
Constant	0.441*** (0.008)
Observations	3,193
R ²	0.206
Adjusted R ²	0.206
Residual Std. Error	0.088 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

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

```

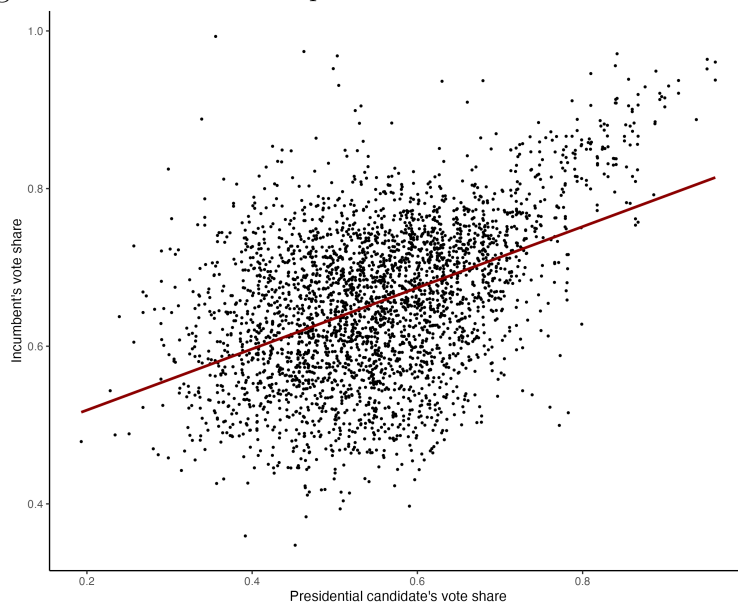
1 # Creating a scatterplot
2 library(ggplot2)
3 scatter3 <- ggplot(data = inc.sub,
4                     mapping = aes(x = presvote,
5                                   y = voteshare)) +
6   geom_point(size=0.5) +
7   geom_smooth(method = "lm", se = FALSE, color = "darkred") + # Add
8     regression line
9   labs(x = "Presidential candidate's vote share", # Add labels
10        y = "Incumbent's vote share") +
11   theme_classic() + # Change theme
12   theme(legend.box.background = element_rect(size = 0.1), # Change
13         legend.position = c(0.85, 0.85)) # Change position of legend
14 # Printing and saving the scatterplot
15 scatter3
16 ggsave(scatter3, file = "scatter_vs_pv.png")

```

The figure below depicts the relationship between the outcome and input variables and

the bivariate model regression line.

Figure 3: Vote share of the presidential candidate and the incumbent



3. Write the prediction equation.

$$Y = \beta_0 + \beta_1 \cdot X$$

where:

Y (voteshare) - incumbent's vote share

X (presvote) - vote share of the presidential candidate of the incumbent's party

$$voteshare = \beta_0 + \beta_1 \cdot (\text{presvote})$$

$$voteshare = 0.441 + 0.388 \cdot (\text{presvote})$$

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.

```
1 # Regressing the selected variables and saving the model as an object
2 model4 <- lm(residuals_voteshare_difflog ~ residuals_presvote_difflog,
3             data = inc.sub)
4
5 # Creating a table of the results
6 stargazer(model4, type = "latex", title = "Linear Regression Results:
7           Model residuals",
8           out = "regression_table_model4.tex")
```

The table below shows the results of bivariate regression between the unexplained variation /residuals/ of the effect of difference in spending on incumbent's vote share /model 1/ and the unexplained variation of the effect of difference in spending on the presidential candidate's vote share /model 2/. There is a positive significant effect of the residuals from model 1 on that of model 2, with a one unit increase in one associated with an increase of 0.257 of the other. This means that there are confounding factors other than the difference in spending that affect both vote shares, explaining some of the variation.

Table 4: Linear Regression Results: Model residuals

	<i>Dependent variable:</i>
	residuals_voteshare_difflog
residuals_presvote_difflog	0.257*** (0.012)
Constant	-0.000 (0.001)
Observations	3,193
R ²	0.130
Adjusted R ²	0.130
Residual Std. Error	0.073 (df = 3191)
F Statistic	476.975*** (df = 1; 3191)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

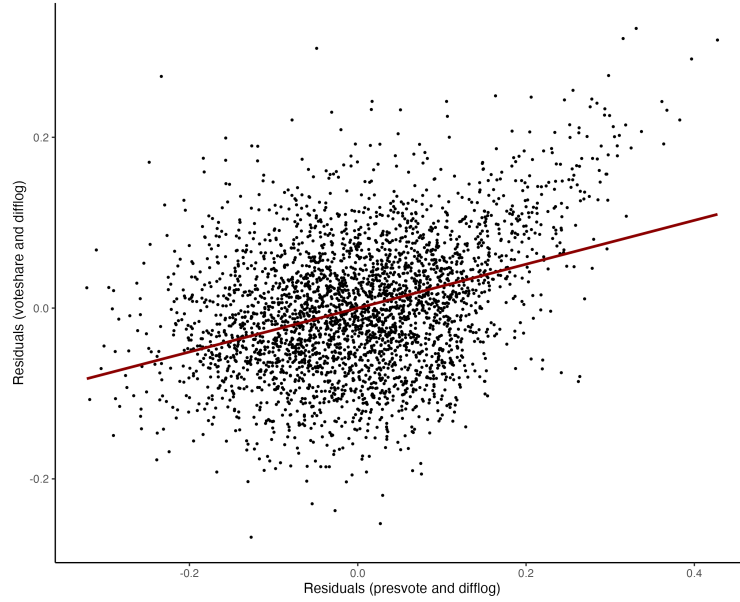
```

1 # Creating a scatterplot
2 library(ggplot2)
3 scatter4 <- ggplot(data = inc.sub,
4                     mapping = aes(x = residuals_presvote_difflog,
5                                   y = residuals_voteshare_difflog)) +
6   geom_point(size=0.5) +
7   geom_smooth(method = "lm", se = FALSE, color = "darkred") + # Add
8     regression line
9   labs(x = "Residuals (presvote and difflog)", # Add labels
10        y = "Residuals (voteshare and difflog)") +
11   theme_classic() + # Change theme
12   theme(legend.box.background = element_rect(size = 0.1), # Change
13         legend.position = c(0.85, 0.85)) # Change position of legend
14 # Printing and saving the scatterplot
15 scatter4
16 ggsave(scatter4, file = "scatter_res1_res2.png")

```

The figure below depicts the relationship between the outcome and input variables and the bivariate model regression line.

Figure 4: Association between model residuals



3. Write the prediction equation.

$$Y = \beta_0 + \beta_1 \cdot X$$

where:

Y - Residuals of difflog regressed against voteshare

X - Residuals of difflog regressed against presvote

$$\text{Residualsofdifflogregressedagainstvoteshare} = \beta_0 + \beta_1 \cdot (\text{residuals of difflog regressed against presvote})$$

$$\text{Residualsofdifflogregressedagainstvoteshare} = 0.257 \cdot (\text{residuals of difflog regressed against presvote})$$

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`.

```
1 # Creating a table of the results
2 stargazer(model4, type = "latex", title = "Multivariate Linear Regression
  Results",
3           out = "regression_table_model5.tex")
```

The table below shows the results of multivariate regression between the president's popularity and the difference in spending between incumbent and challenger and the incumbent's vote share. Both factors, although to a different extent, have a significant positive effect of the incumbent vote share. For every one unit increase in the presidential candidate vote share, the incumbent's vote share increases by 0.257 units, controlling for the difference in spending. While one unit increase in the difference in spending increases the vote share by 0.036.

Table 5: Multivariate Linear Regression Results

	<i>Dependent variable:</i>
	voteshare
presvote	0.257*** (0.012)
difflog	0.036*** (0.001)
Constant	0.449*** (0.006)
Observations	3,193
R ²	0.450
Adjusted R ²	0.449
Residual Std. Error	0.073 (df = 3190)
F Statistic	1,302.947*** (df = 2; 3190)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

2. Write the prediction equation.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2$$

where:

Y - Incumbent vote share

X_1 - Vote share of the presidential candidate

X_2 - Difference in spending between incumbent and challenger

$$Incumbentvoteshare = \beta_0 + \beta_1 \cdot (\text{vote share of the presidential candidate}) + \beta_2 \cdot (\text{difference in spending})$$

$$Incumbentvoteshare = \beta_0 + \beta_1 \cdot (\text{vote share of the presidential candidate}) + \beta_2 \cdot (\text{difference in spending})$$

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 coefficient of the effect of presidential candidate vote share (presvote) is the same as the coefficient of the effect of residuals of spending differences regressed against presidential vote share. Once we control for the difference in spending, the effect of presidential candidates on the incumbent vote share decreases from 0.388 to 0.257, which is natural given the spending difference significantly affects both variables. Additionally, there are some other confounding factors influencing both variables, apart from possible other unrelated factors. The fact that the coefficients are exactly the same is a coincidence.