Problem Set 3

Daniel Murray (13303981)

Due: November 20, 2022

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 GitHub.

- This problem set is due before 23:59 on Sunday November 20, 2022. No late assignments will be accepted.

- Total available points for this homework is 80.

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

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`.

    Answer:

    A regression was run and summary statistics produced using the below code:

```
1 # Run regression with outcome variable voteshare and explanatory variable
      difflog
2 mod1 <- lm(voteshare ~ difflog, data = dat)
3
4 # Output summary statistics of regression model as latex code
5 stargazer(mod1, type = 'latex', title = 'Summary Statistics of Regression
      Model')
```

This produced the following output of summary statistics:

Table 1: Summary Statistics of Regression Model

	Dependent variable:
	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

Answer:

A scatterplot was produced using the below code:

```
1 # Output scatterplot of voteshare and difflog with regression line
2 ggplot(dat, aes(x=difflog, y=voteshare)) +
3   geom_point(size=0.75) +
4   geom_abline(intercept = mod1$coefficients[1], slope = mod1$coefficients
      [2], color = "red") +
5   ggtitle("Scatterplot of voteshare ~ difflog") +
```

```
6    theme(plot.title = element_text(hjust = 0.5))
```

This produced the following scatterplot:

Figure 1:

Scatterplot of voteshare ~ difflog
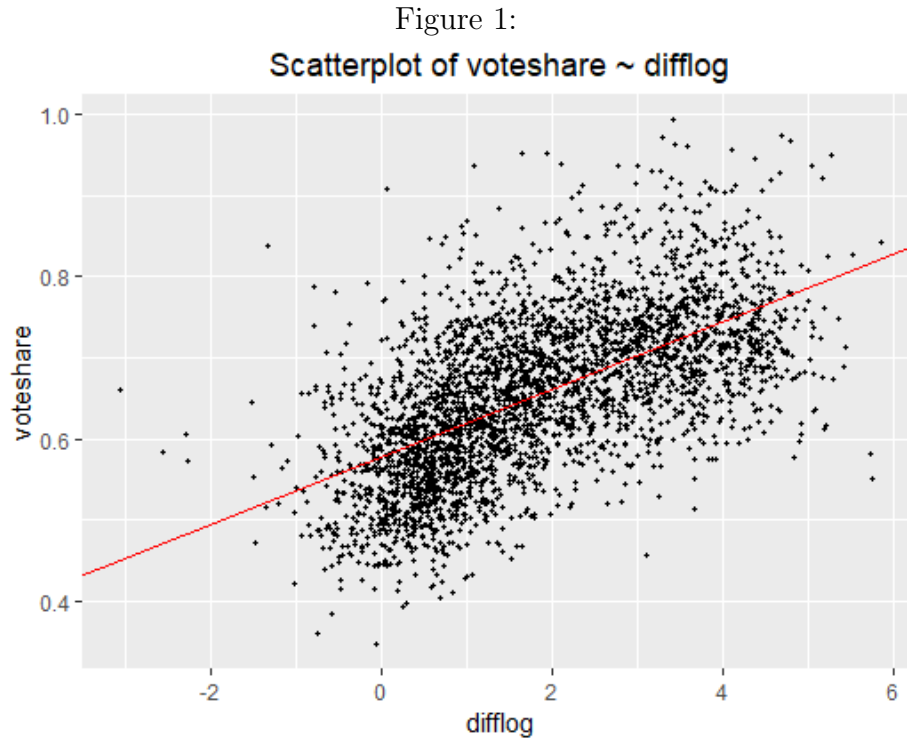
3. Save the residuals of the model in a separate object.

Answer:

The residuals of the model were saved in a separate object using the below code:

```
1 # Save residuals of the model in a separate object
2 mod1_res <- mod1$residuals
```

4. Write the prediction equation.

Answer:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$$
$$voteshare = 0.579 + 0.042 \times difflog$$

3

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`.

   Answer:

   A regression was run and summary statistics produced using the below code:

```
1 # Run regression with outcome variable presvote and explanatory variable
      difflog
2 mod2 <- lm(presvote ~ difflog, data = dat)
3
4 # Output summary statistics of regression model as latex code
5 stargazer(mod2, type = 'latex', title = 'Summary Statistics of Regression
      Model')
```

   This produced the following output of summary statistics:

Table 2: Summary Statistics of Regression Model

|	*Dependent variable:*
	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

   Answer:

   A scatterplot was produced using the below code:

```
1  # Output scatterplot of presvote and difflog with regression line
2  ggplot(dat, aes(x=difflog, y=presvote)) +
3    geom_point(size=0.75) +
4    geom_abline(intercept = mod2$coefficients[1], slope = mod2$coefficients
       [2], color = "red") +
5    ggtitle("Scatterplot of presvote ~ difflog") +
6    theme(plot.title = element_text(hjust = 0.5))
```

   This produced the following scatterplot:

Figure 2:
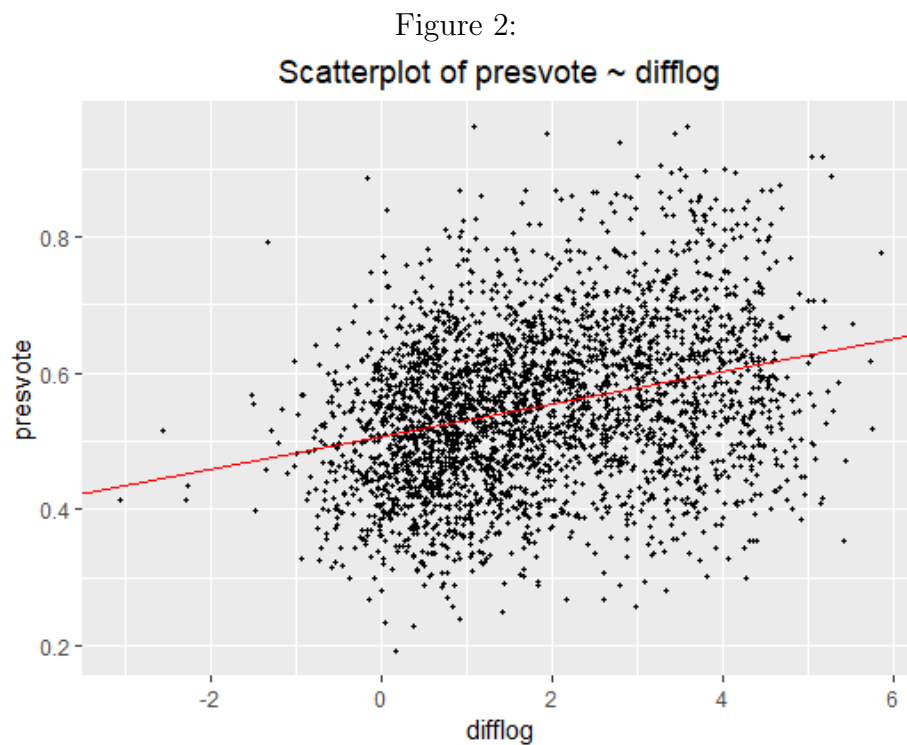
3. Save the residuals of the model in a separate object.

   Answer:

The residuals of the model were saved in a separate object using the below code:

```
1  # Save residuals of the model in a separate object
2  mod2_res <- mod2$residuals
```

4. Write the prediction equation.

   Answer:

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

$$presvote = 0.508 + 0.024 \times difflog$$

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`.

   Answer:

   A regression was run and summary statistics produced using the below code:

   ```
   # Run regression with outcome variable voteshare and explanatory variable
   # presvote
   mod3 <- lm(voteshare ~ presvote, data = dat)

   # Output summary statistics of regression model as latex code
   stargazer(mod3, type = 'latex', title = 'Summary Statistics of Regression
       Model')
   ```

   This produced the following output of summary statistics:

   Table 3: Summary Statistics of Regression Model

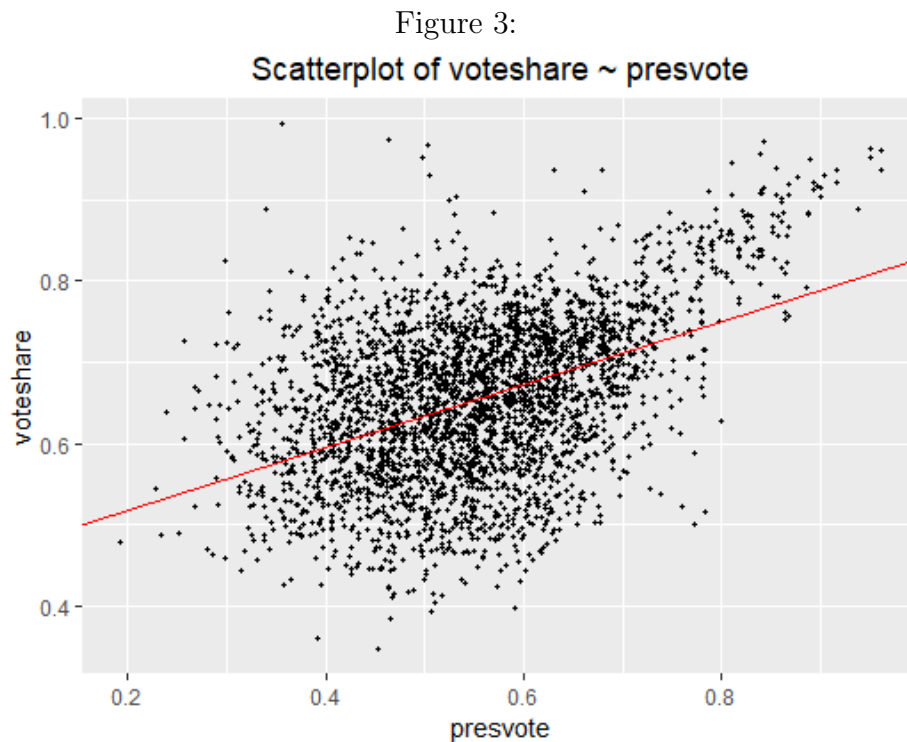 | | *Dependent variable:* |
   |---|---|
   | | voteshare |
   | presvote | 0.388*** |
   | | (0.013) |
   | Constant | 0.441*** |
   | | (0.008) |
   | Observations | 3,193 |
   | R^2 | 0.206 |
   | Adjusted R^2 | 0.206 |
   | Residual Std. Error | 0.088 (df = 3191) |
   | F Statistic | 826.950*** (df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

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

   Answer:

   A scatterplot was produced using the below code:

```
1  # Output scatterplot of voteshare and presvote with regression line
2  ggplot(dat, aes(x=presvote, y=voteshare)) +
3    geom_point(size=0.75) +
4    geom_abline(intercept = mod3$coefficients[1], slope = mod3$coefficients
       [2], color = "red") +
5    ggtitle("Scatterplot of voteshare ~ presvote") +
6    theme(plot.title = element_text(hjust = 0.5))
```

   This produced the following scatterplot:

Figure 3:

Scatterplot of voteshare ~ presvote

3. Write the prediction equation.

   Answer:

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

8

$$voteshare = 0.441 + 0.388 \times presvote$$

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.

   Answer:

   A regression was run and summary statistics produced using the below code:

```
# Run regression with outcome variable mod1_res and explanatory variable
# mod2_res
mod4 <- lm(mod1_res ~ mod2_res)

# Output summary statistics of regression model as latex code
stargazer(mod4, type = 'latex', title = 'Summary Statistics of Regression
    Model')
```

   This produced the following output of summary statistics:

Table 4: Summary Statistics of Regression Model

|	*Dependent variable:*
	mod1_res
mod2_res	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

   Answer:

   A scatterplot was produced using the below code:

```
1  # Output scatterplot of mod1_res and mod2_res with regression line
2  ggplot(dat, aes(x=mod2_res, y=mod1_res)) +
3    geom_point(size=0.75) +
4    geom_abline(intercept = mod4$coefficients[1], slope = mod4$coefficients
        [2], color = "red") +
5    ggtitle("Scatterplot of mod1_res ~ mod2_res ") +
6    theme(plot.title = element_text(hjust = 0.5))
```

   This produced the following scatterplot:

Figure 4:

Scatterplot of mod1_res ~ mod2_res
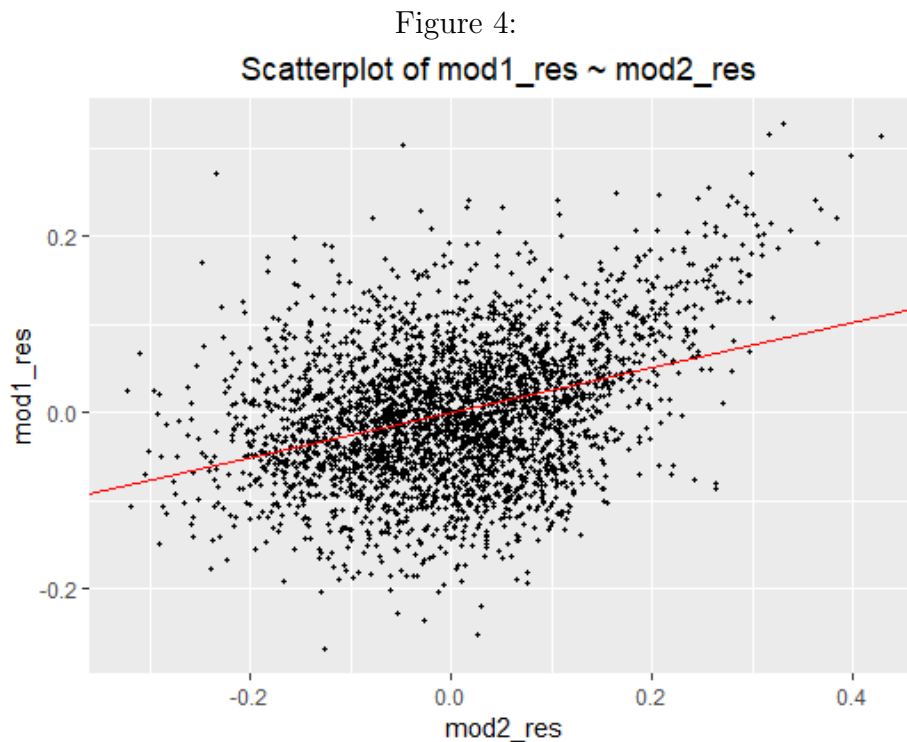
3. Write the prediction equation.

   Answer:

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

11

$$mod1\_res = 0 + 0.257 \times mod2\_res$$

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`.

   Answer:

   A regression was run and summary statistics produced using the below code:

   ```
   # Run regression with outcome variable voteshare and explanatory
       variables mod2_res
   mod5 <- lm(voteshare ~ difflog + presvote, data = dat)

   # Output summary statistics of regression model as latex code
   stargazer(mod5, type = 'latex', title = 'Summary Statistics of Regression
       Model')
   ```

   This produced the following output of summary statistics:

   Table 5: Summary Statistics of Regression Model

	Dependent variable:
	voteshare
difflog	0.036***
	(0.001)
presvote	0.257***
	(0.012)
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)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Write the prediction equation.

   Answer:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$
$$voteshare = 0.449 + 0.036 \times difflog + 0.257 \times presvote$$

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case?

   Answer:

   The coefficient for `presvote` in this question has the same value as the coefficient for `mod2_res` in Question 4. The residuals from Question 1 represent the variation in `voteshare` that is not explained by `difflog`, and the residuals from Question 2 represent the variation in `presvote` that is not explained by `difflog`. Therefore, the coefficient for `mod2_res` in Question 4 represents the amount of co-variation between `presvote` and `voteshare` that is not explained by `difflog`.