

Problem Set 3

Applied Stats/Quant Methods 1

Due: November 19, 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 19, 2023. No late assignments will be accepted.

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`.

```
1 # Run regression model in R
2 Regression_1 <- lm(voteshare ~ difflog, data = incumbents_subset)
3
4 # Get summary of model with coefficient estimates
5 summary(Regression_1)
6
```

Listing 1: Regression Model 1 in R

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

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

```
1 # Create a scatterplot with regression line
2 ScatterplotRegression1<-ggplot(incumbents_subset ,
3   aes(x = difflog , y = voteshare)) +
4   geom_point() +
5   geom_smooth(method = "lm", se = FALSE, color = "blue") +
6   labs(title = "Scatterplot of Regression 1",
7    x = "Difference in Campaign Spending (difflog)",
8    y = "Incumbent Vote Share")
9
10 # Save Scatterplot as an image
11 ggsave("Scatterplot_of_Regression_1.pdf",
12   plot = ScatterplotRegression1,
13   width = 6, height = 4, units = "in")
14
```

Listing 2: Scatterplot 1 code in R

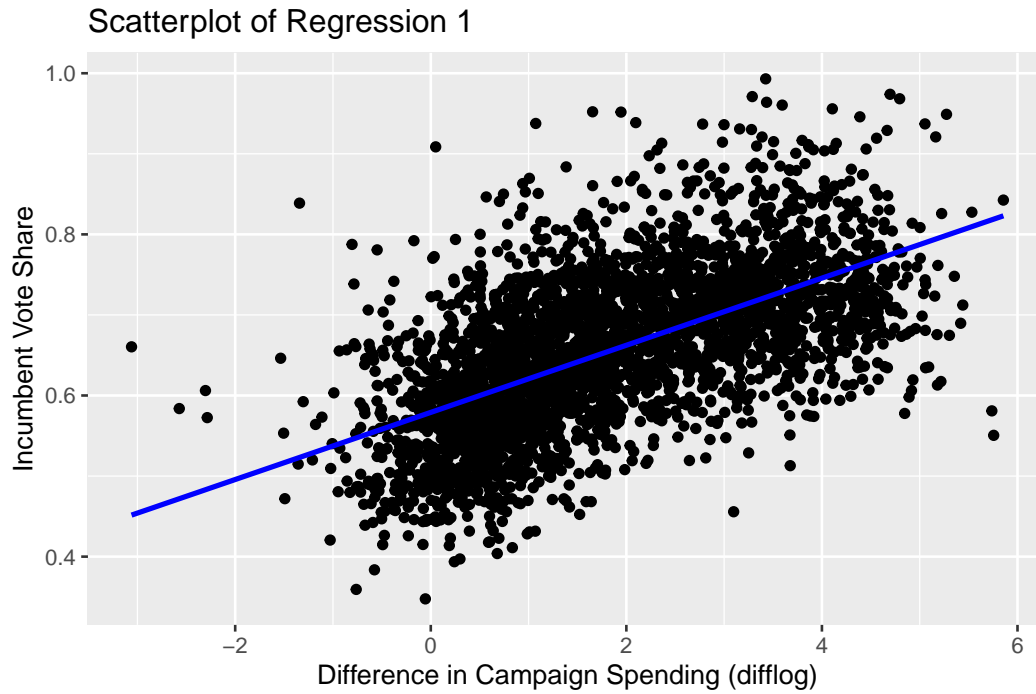


Figure 1: Scatterplot of Regression 1

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

```
1 Save the residuals of the model in a separate object (R).  
2 residuals_1 <- resid(Regression_1)  
3
```

4. Write the prediction equation.

$$Y = B_0 + B_1 \quad (\text{Linear Regression Model})$$

$$\text{Predicted Incumbent voteshare} = 0.579031 + 0.041666 \times \text{difflog} \quad (\text{Specific coefficients})$$

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 #Run a regression model where the outcome variable is presvote and
2 the explanatory variable is difflog.
3 Regression_2 <- lm(presvote ~ difflog, data = incumbents_subset)
4
5 # Get summary of model with coefficient estimates
6 summary(Regression_2)
7
```

Listing 3: Regression Model 2 in R

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

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

```
1 # Create a scatterplot with regression line
2 ScatterplotRegression2<-ggplot(incumbents_subset ,
3   aes(x = difflog , y = presvote)) +
4   geom_point() +
5   geom_smooth(method = "lm", se = FALSE, color = "green") +
6   labs(title = "Scatterplot of Regression 2",
7     x = "Difference in Campaign Spending (difflog)",
8     y = "Presidential Vote Share")
9
10 # Save Scatterplot as an image
11 ggsave("Scatterplot_of_Regression_2.pdf",
12   plot = ScatterplotRegression1 ,
13   width = 6, height = 4, units = "in")
14
```

Listing 4: Scatterplot 1 code in R

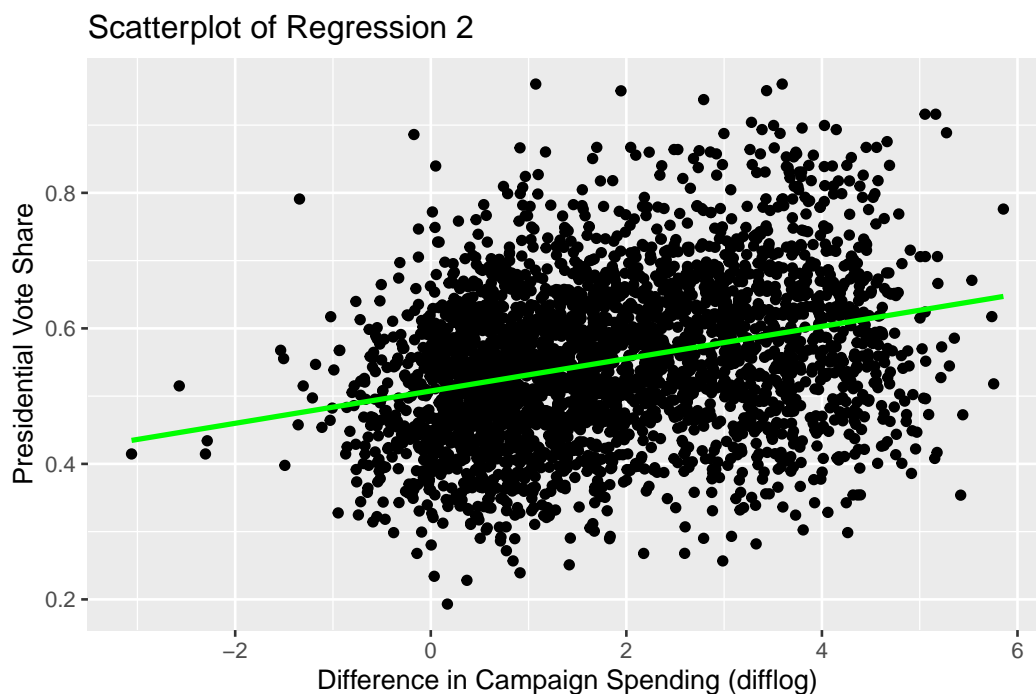


Figure 2: Scatterplot of Regression 2

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

```
1   Save the residuals of the model in a separate object (R) .  
2   residuals_2 <- resid(Regression_2)  
3
```

4. Write the prediction equation.

$$Y = B_0 + B_1 \quad (\text{Linear Regression Model})$$

$$\text{Predicted Presidential voteshare} = 0.507583 + 0.023837 \times \text{difflog} \quad (\text{Specific coefficients})$$

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 #Run a regression model where the outcome variable is voteshare and
2 the explanatory variable is presvote.
3 Regression_3 <- lm(voteshare ~ presvote, data = incumbents_subset)
4
5 # Get summary of model with coefficient estimates
6 summary(Regression_3)
7
```

Listing 5: Regression Model 3 in R

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

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

```
1 # Create a scatterplot with regression line
2 ScatterplotRegression3<-ggplot(incumbents_subset,
3 aes(x = presvote, y = voteshare)) +
4 geom_point() +
5 geom_smooth(method = "lm", se = FALSE, color = "red") +
6 labs(title = "Scatterplot of Regression 3",
7 x = "Presidential Vote Share",
8 y = "Incumbent Vote Share")
9
10 # Save Scatterplot as an image
11 ggsave("Scatterplot_of_Regression_3.pdf",
12 plot = ScatterplotRegression3,
13 width = 6, height = 4, units = "in")
14
```

Listing 6: Scatterplot 3 code in R

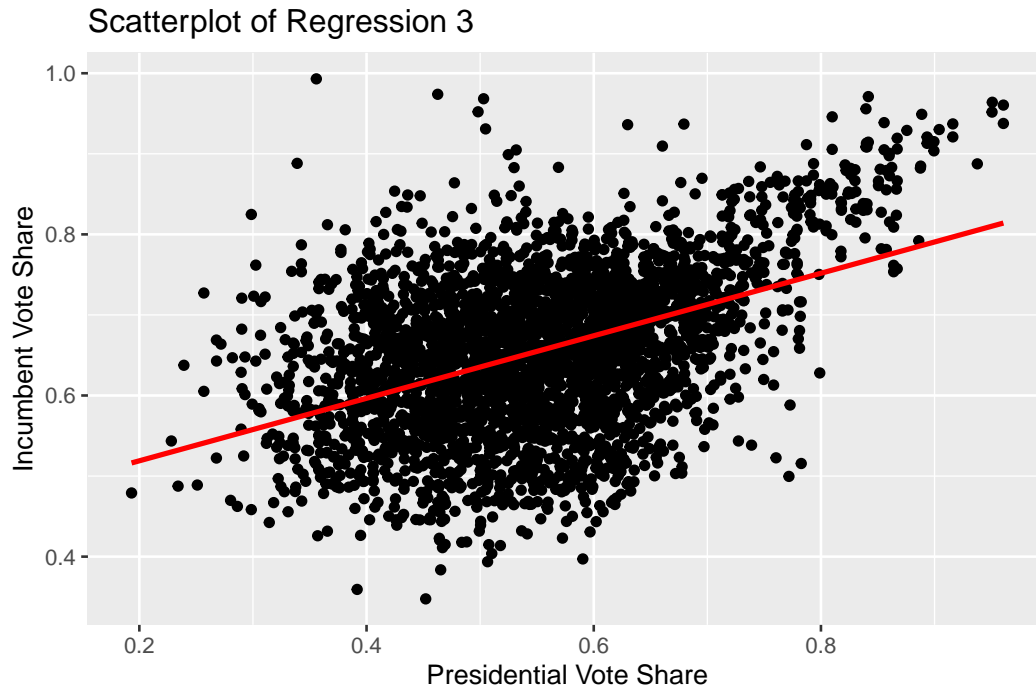


Figure 3: Scatterplot of Regression 3

3. Write the prediction equation.

$$Y = B_0 + B_1 \quad (\text{Linear Regression Model})$$

$$\text{Pred Incumbent VoteShare} = 0.441330 + 0.388018 \times \text{Presidential VoteShare} \quad (\text{Specific coefficients})$$

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 #Run a regression with residuals_1 as the outcome and
2 residuals_2 as the explanatory variable
3 Regression_4_resid <- lm((residuals_1 ~ residuals_2)
4
5 # Get summary of model with coefficient estimates
6 summary(Regression_4_resid)
7
```

Listing 7: Regression Model 4 in R

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.520e-18	1.299e-03	0.00	1
residuals_2	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

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

```
1 ScatterplotRegression4<-ggplot(incumbents_subset, aes(x = residuals_2,
2 y = residuals_1)) +
3 geom_point() +
4 geom_smooth(method = "lm", se = FALSE, color = "purple") +
5 labs(title = "Scatterplot of Regression 4",
6 x = "Residuals from Regression 2",
7 y = "Residuals from Regression 1")
8
9 # Save Scatterplot as an image
10 ggsave("Scatterplot_of_Regression_4.pdf",
11 plot = ScatterplotRegression4,
12 width = 6, height = 4, units = "in")
13
```

Listing 8: Scatterplot 3 code in R



Figure 4: Scatterplot of Regression 4

3. Write the prediction equation.

$$Y = B_0 + B_1 \quad (\text{Linear Regression Model})$$

$$\text{Resid from Regression 1} = -5.520 + 2.569 \times \text{Resid from Regression 2} \quad (\text{Specific coefficients})$$

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 #Run a regression where the outcome variable is the incumbent's
  voteshare and the explanatory variables are difflog and presvote.
2 Regression_5 <- lm(voteshare ~ difflog + presvote, data = incumbents_
  subset)
3
4 # Get summary of model with coefficient estimates
5 summary(Regression_5)
6
```

Listing 9: Regression Model 5 in R

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

Residual standard error: 0.07339 on 3190 degrees of freedom

Multiple R-squared: 0.4496, Adjusted R-squared: 0.4493

F-statistic: 1303 on 2 and 3190 DF, p-value: < 2.2e-16

2. Write the prediction equation.

$$Y = B_0 + B_1X_1 + B_2X_2 \quad (\text{Multiple Linear Regression Model})$$

$$\text{Incumbent's Voteshare} = 0.4486 + 0.0355 \cdot \text{difflog} + 0.2568 \cdot \text{presvote}$$

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

The residual standard errors (RSE) are almost identical 0.07338 for Regression 4 and 0.07339 for Regression 5, this could occur because...