

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`.

Answer

After we load our dataset into our working environment, we execute our regression model in which the incumbent's vote share is explained by the difference in campaign spending between incumbent and challenger.

Step 1: We begin by outlining our hypotheses.

Null hypothesis: A difference in campaign spending between incumbents and challengers has no impact on incumbent vote share

Alternative hypothesis: A difference in incumbent campaign spending either increases (or decreases) their vote share

$$H_0 : \beta = 0$$

$$H_A : \beta \neq 0$$

Step 2: We run our regression in R

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

We then run a summary to check the coefficients

```
summary(model)
```

Call:

```
lm(formula = inc.sub$voteshare ~ inc.sub$difflog, data = inc.sub)
```

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 ***
inc.sub\$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

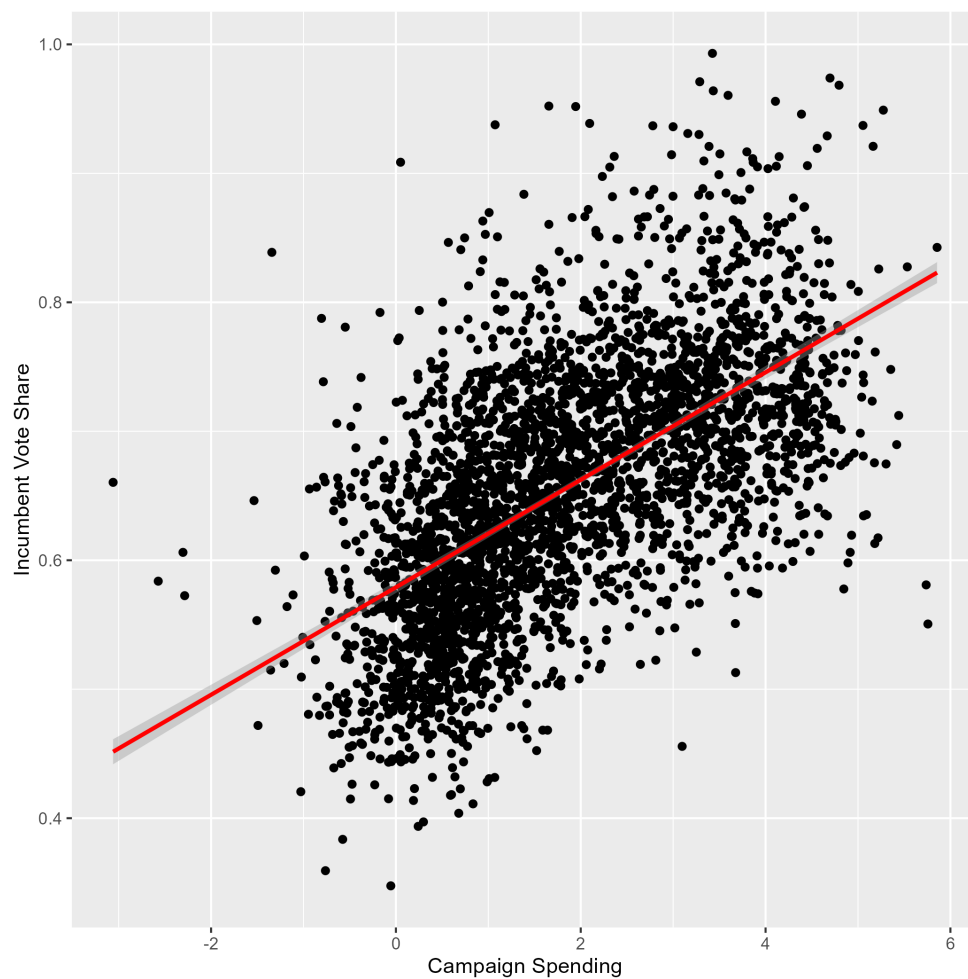
Step 3: Conclusions:

We have evidence to support the view that a one unit increase in spending leads to a 0.04 unit increase in vote share for the incumbent party. The estimated coefficient is statistically differentiable from zero at the $\alpha = 0.05$ level because the p-value < 0.05 ($\approx 2e-16$).

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

```
1
2 # Scatter plot
3 scatter <-
4   ggplot(data = inc.sub,
5         mapping = aes(x = difflog,
6                       y = voteshare)) +
7   geom_point() +
8   labs(x = "Campaign Spending",
9        y = "Incumbent Vote Share") +
10
11   geom_smooth(method='lm', col="red")
12
13 ggsave(scatter, file = "vote_share_incumbent_scatter.png")
14
15 # Print plot object
```

Figure 1: Incumbent Vote Share as compared to Campaign Spending. Using ggplot.



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

```
vote_residuals <- model$residuals  
vote_residuals
```

4. Write the prediction equation.

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

$$\hat{Y}_i = 0.579031 + 0.041666 * (\text{Campaign Spending})$$

On average, with every additional dollar spent on the campaign, we can expect the vote share for the incumbent Party to increase by 0.041666 scale points.

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

After we load our dataset into our working environment, we execute our regression model in which the presidential candidate of the incumbent's party vote share is explained by the difference in campaign spending between incumbent and challenger.

Step 1: We begin by outlining our hypotheses.

Null hypothesis: A difference in campaign spending between incumbents and challengers has no impact on the presidential candidate of the incumbent's party vote share

Alternative hypothesis: A difference in campaign spending between incumbents and challengers either increases (or decreases) the vote share for the presidential candidate of the incumbent's party.

$$H_0 : \beta = 0$$

$$H_A : \beta \neq 0$$

Step 2: We run our regression in R

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

We then run a summary to check the coefficients

```
summary(model_pres)
```

Call:

```
lm(formula = inc.sub$presvote ~ inc.sub$difflog, data = inc.sub)
```

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 ***
inc.sub\$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

Step 3: Conclusions:

We have evidence to support the view that a one unit difference in spending leads to a 0.023837 unit change in vote share for the the presidential candidate of the incumbent's party. The estimated coefficient is statistically differentiable from zero at the $\alpha = 0.05$ level because the p-value < 0.05 ($\approx 2e-16$).

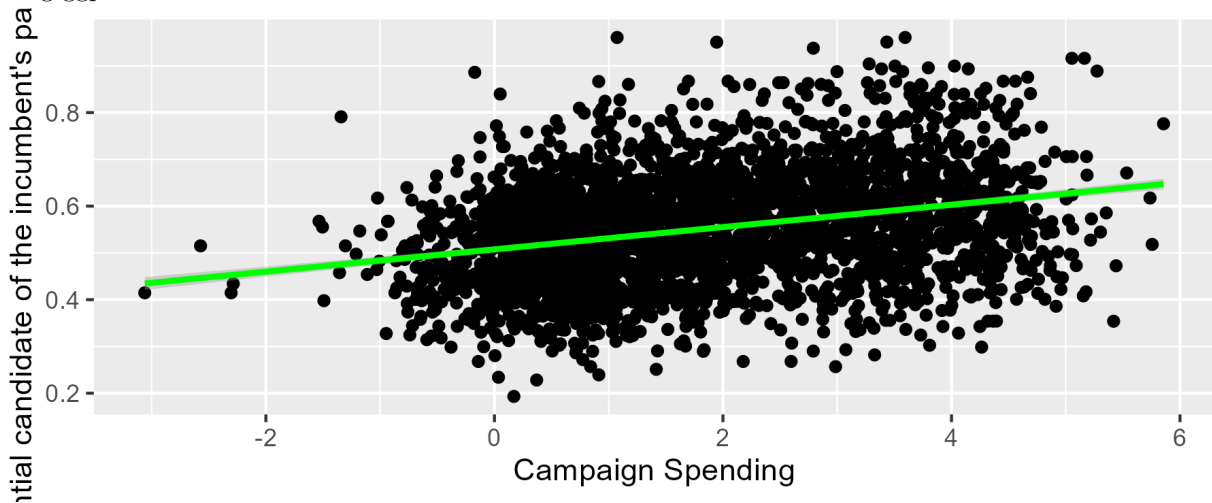
2. Make a scatterplot of the two variables and add the regression line.

```

1  geom_smooth(method='lm', col="green")
2
3
4  ggsave(scatter_pres, file = "vote_share_pres_scatter.png")
5
6  # Print plot object
7  scatter_pres
8
9  # Now we add the regression line
10
11 # (2.c.) Residuals -----
12
13 # Save the residuals of the model in a separate object
14
15 pres_residals <- model_pres$residuals

```

Figure 2: Presidential candidate of the incumbent's party vote share as compared to Campaign Spending. Using ggplot.



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

```
pres_residuals <- model_pres$residuals
pres_residuals
```

4. Write the prediction equation.

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

$$\hat{Y}_i = 0.507583 + 0.02383 * (\text{Campaign Spending})$$

On average, with every additional dollar spent on the campaign, we can expect the vote share for the presidential candidate of the incumbent's party vote share to increase by 0.041666 scale points.

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

After we load our dataset into our working environment, we execute our regression model in which we check for association between the vote share of the presidential candidate of the incumbent's party and the incumbent's electoral success.

Step 1: We begin by outlining our hypotheses.

Null hypothesis: There is no association between the vote share of the presidential candidate of the incumbent's party and the incumbent's electoral success

Alternative hypothesis: There is an association between the vote share of the presidential candidate of the incumbent's party and the incumbent's electoral success

$$H_0 : \beta = 0$$

$$H_A : \beta \neq 0$$

Step 2: We run our regression in R

```
model <- lm(inc.sub$voteshare ~ inc.sub$presvote, data=inc.sub)
```

We then run a summary to check the coefficients

```
summary(model_pres)
```

Call:

```
lm(formula = inc.sub$voteshare ~ inc.sub$presvote, data = inc.sub)
```

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 ***
inc.sub\$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

Step 3: Conclusions:

We have evidence to support the view that a one unit increase in the incumbent party's electoral success leads to a 0.388 increase in vote share for the President's Party. The estimated coefficient is statistically differentiable from zero at the $\alpha = 0.05$ level because the p-value < 0.05 ($\approx 2e-16$).

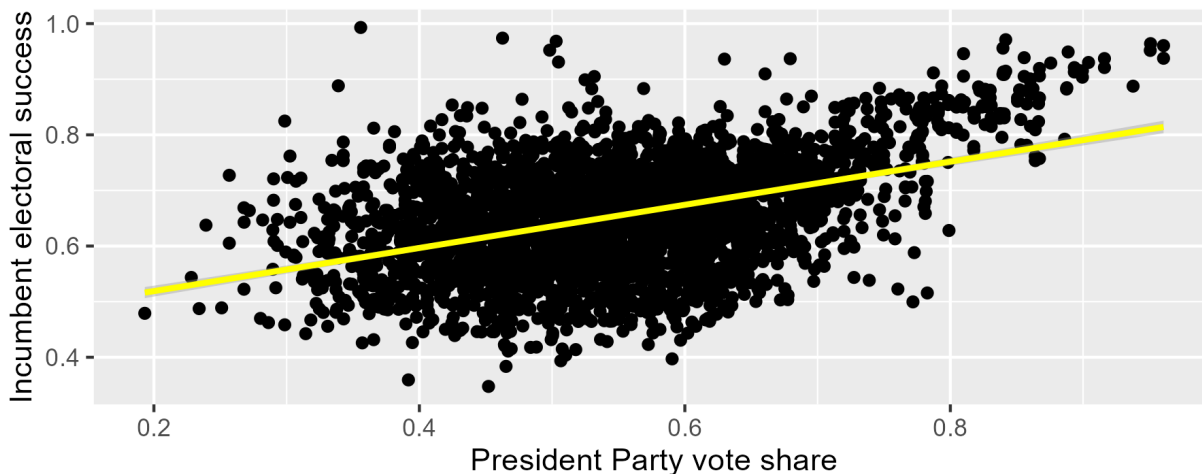
2. Make a scatterplot of the two variables and add the regression line.

```

1  geom_smooth(method='lm', col="yellow")
2
3
4  ggsave(scatter_vote, file = "vote_share2_scatter.png")
5
6  # Print plot object
7  scatter_vote
8
9  # Now we add the regression line
10
11 # (3.c.) Residuals _____

```

Figure 3: President Party's Vote Share as compared to Incumbent's electoral success. Using ggplot.



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

```
pres_residuals <- model_pres$residuals  
pres_residuals
```

4. Write the prediction equation.

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

$$\hat{Y}_i = 0.441330 + 0.388 * (\text{incumbent electoral success})$$

On average, with every one unit of electoral success for incumbent party, we can expect the vote share for the President's Party to increase by 0.388 scale points.

With regard to association, the slope is positive, indicating a positive association. The slope does not however tell us the strength of the association.

If we look at the R^2 value, it is 0.2058, indicating a weak positive relationship between the two variables.

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.

Null hypothesis: There is no association between the residuals

Alternative hypothesis: There is an association between the residuals

$$H_0 : \beta = 0$$

$$H_A : \beta \neq 0$$

Step 1: Run the regression ...

```
model_residuals <- lm(pres_residuals ~ vote_residuals, data=inc.sub)

summary(model_residuals)
```

Call:

```
lm(formula = pres_residuals ~ vote_residuals, data = inc.sub)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.32618	-0.07532	-0.00485	0.06698	0.45288

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.874e-18	1.928e-03	0.00	1
vote_residuals	-2.059e-01	2.188e-02	-9.41	<2e-16 ***

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

Residual standard error: 0.1089 on 3191 degrees of freedom

Multiple R-squared: 0.027, Adjusted R-squared: 0.02669

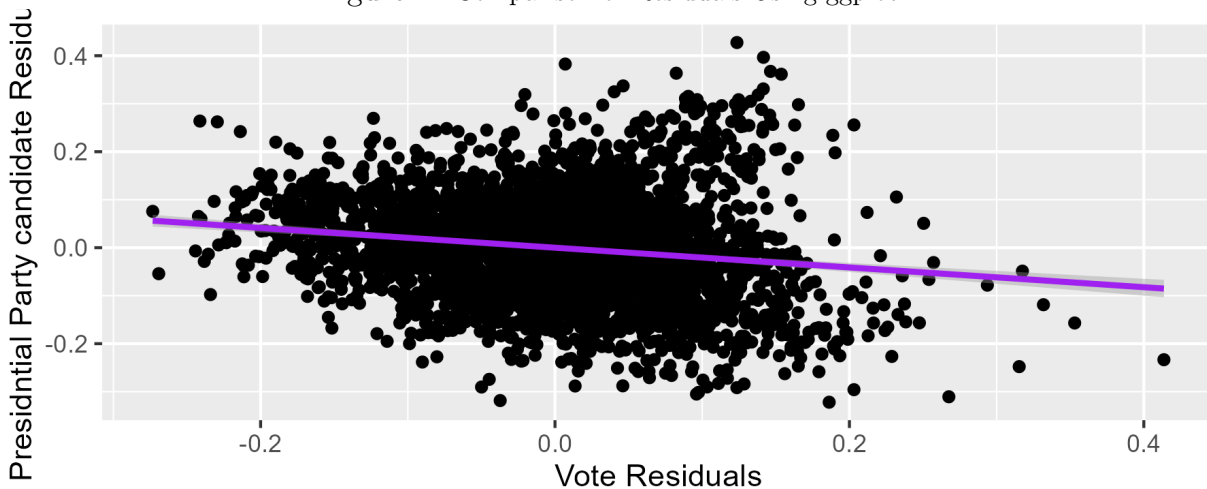
F-statistic: 88.54 on 1 and 3191 DF, p-value: < 2.2e-16

The slope is equal to $-2.059e-01$. So we do not have enough evidence to reject the null hypothesis in this instance. Note that the estimated coefficient is statistically differentiable from zero at the $\alpha = 0.05$ level because the p-value < 0.05 ($\approx 2e-16$).

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

```
1 # Scatter plot
2 scatter_residuals <-
3   ggplot(data = inc.sub,
4     mapping = aes(x = vote_residuals,
5       y = pres_residuals)) +
6   geom_point() +
7   labs(x = "Vote Residuals",
8     y = "Presidential Party candidate Residuals") +
9
10  geom_smooth(method='lm', col="purple")
11
12 ggsave(scatter_residual, file = "residuals_scatter.png")
13
14 # Print plot object
15 scatter_residuals
```

Figure 4: Comparison of Residuals Using ggplot.



3. Write the prediction equation.

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

$$\hat{Y}_i = -2.874e - 18 + -2.059e - 01 * (\text{Residuals from part b})$$

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