

# Problem Set 3

## Applied Stats/Quant Methods 1

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**.

**R code:** `difflog_voteshare_regression <- lm(voteshare ~ difflog, data = Incumbent_data)`

	Min	1Q	Median	3Q	Max
<b>Residuals:</b>	-0.26832	-0.05345	-0.00377	0.04780	0.32749

		Estimate	Standard Error	t-value	Pr(> t )
<b>Summary Table:</b>	(Intercept)	0.579031	0.002251	257.19	<2e-16***
	difflog	0.041666	0.000968	43.04	<2e-16***

Significance Codes: 0'\*\*\*' 0.001'\*\*\*' 0.01'\*' 0.05'.' 0.2 '' 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

The p-value is significant at 99.9%

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

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

**R code:** `Q1residuals<-resid(difflog_voteshare_regression)`

4. Write the prediction equation.

$$\hat{y} = 0.579 + 0.042(\text{difflog})$$

## 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`.

**R code:** `presvote_difflog_regression <- lm(presvote ~ difflog, data = Incumbent_data)`

	Min	1Q	Median	3Q	Max
<b>Residuals:</b>	-0.32196	-0.07407	-0.00102	0.07151	0.42743

		Estimate	Standard Error	t-value	Pr(> t )
<b>Summary Table:</b>	(Intercept)	0.507583	0.003161	160.60	<2e-16***
	difflog	0.023837	0.001359	17.54	<2e-16***

Significance Codes: 0'\*\*\*' 0.001'\*\*\*' 0.01'\*' 0.05'.' 0.2 " 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

The p-value is significant at 99.9%

2. Make a scatterplot of the two variables and add the regression line.
3. Save the residuals of the model in a separate object.

**R code:** `Q2residuals<-resid(difflog_presvote_regression)`

4. Write the prediction equation.

$$\hat{y} = 0.508 + 0.024(\text{difflog})$$

## 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**.

**R code:** `voteshare_presvote_regression <- lm(voteshare ~ presvote, data = Incumbent_data)`

	Min	1Q	Median	3Q	Max
<b>Residuals:</b>	-0.27330	-0.05888	0.00394	0.06148	0.41365

	Estimate	Standard Error	t-value	Pr(> t )
<b>Summary Table:</b> (Intercept)	0.441330	0.007599	58.08	<2e-16***
presvote	0.388018	0.013493	28.76	<2e-16***

Significance Codes: 0'\*\*\*' 0.001'\*\*\*' 0.01'\*' 0.05'.' 0.2 " 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

The p-value is significant at 99.9%

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

3. Write the prediction equation.

$$\hat{y} = 0.441 + 0.388(\text{presvote})$$

## 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.

**R code:** `Q1residuals_regressed_on_Q2residuals <- lm(Q1residuals ~ Q2residuals)`

	Min	1Q	Median	3Q	Max
<b>Residuals:</b>	-0.25928	-0.04737	-0.00121	0.04618	0.33126

		Estimate	Standard Error	t-value	Pr(> t )
<b>Summary Table:</b>	(Intercept)	-4.860e-18	1.299e-03	0.00	1
	Q2 residuals	2.569e-01	1.176e-02	21.84	<2e-16***

Significance Codes: 0'\*\*\*' 0.001'\*\*\*' 0.01'\*\*' 0.05'.' 0.2 " 1

Residual standard error: 0.07338 on 3191 degrees of freedom

Multiple R-squared: 0.13, Adjusted R-squared: 0.12

F-statistic: 477 on 1 and 3191 DF

p-value: < 2.2e-16

The p-value is significant at 99.9%

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

3. Write the prediction equation.

$$\hat{y} = -4.860e - 18 + 2.59e - 01(\text{Q2residuals})$$

## 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`.

**R code:** `voteshare_regressed_on_difflog_and_presvote <- lm(voteshare ~ difflog + presvote, data = Incumbent_data)`

	Min	1Q	Median	3Q	Max
<b>Residuals:</b>	-0.25928	-0.04737	-0.00121	0.04618	0.33126

		Estimate	Standard Error	t-value	Pr(> t )
<b>Summary Table:</b>	(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***

Significance Codes: 0'\*\*\*' 0.001'\*\*\*' 0.01'\*' 0.05'.' 0.2 " 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

The p-value is significant at 99.9%

2. Write the prediction equation.

$$\hat{y} = 0.449 + 0.388(\text{difflog}) + 0.257(\text{presvote})$$

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 residuals in both Question 4 and Question 5 are identical. I believe this to be so because both are examining different sides of the same coin; i.e., the same variables: `presvote`, `difflog`, and `voteshare`.