

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`.

```
difflog_Voteshare <- lm(voteshare ~ difflog, data = incumbents_subset)
summary(difflog_Voteshare)
```

```
Call:
lm(formula = voteshare ~ difflog, data = incumbents_subset)
```

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 ***
difflog	0.041666	0.000968	43.04	<2e-16 ***

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

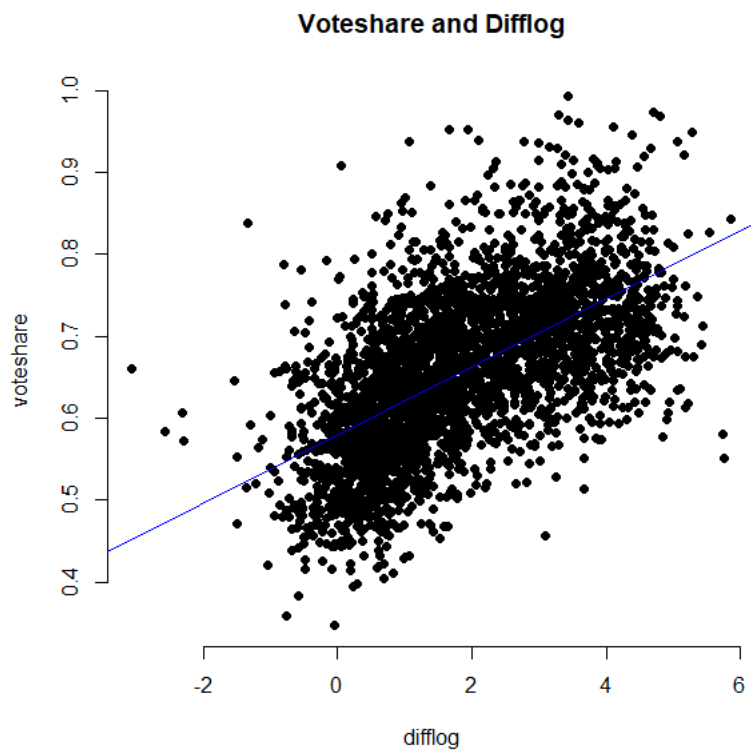
Residual standard error: 0.07867 on 3191 degrees of freedom

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.

```
plot(incumbents_subset$difflog, incumbents_subset$voteshare,
main = "Voteshare and Difflog",
xlab = "difflog", ylab = "voteshare",
pch = 19, frame = FALSE)
abline(lm(voteshare ~ difflog, data = incumbents_subset), col = "blue")
```



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

```
difflog_VoteshareRes <- difflog_Voteshare["residuals"]
```

4. Write the prediction equation.

$$y = B_0 + B_1X + e$$

$$y = .579 + .042(X) + .001$$

In this case $X = \text{Difflog score}$

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`.

```
presvote_Diffloglm <- lm(presvote ~ difflog, data = incumbents_subset)
summary(presvote_Diffloglm)
```

Call:

```
lm(formula = presvote ~ difflog, data = incumbents_subset)
```

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 ***
difflog	0.023837	0.001359	17.54	<2e-16 ***

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

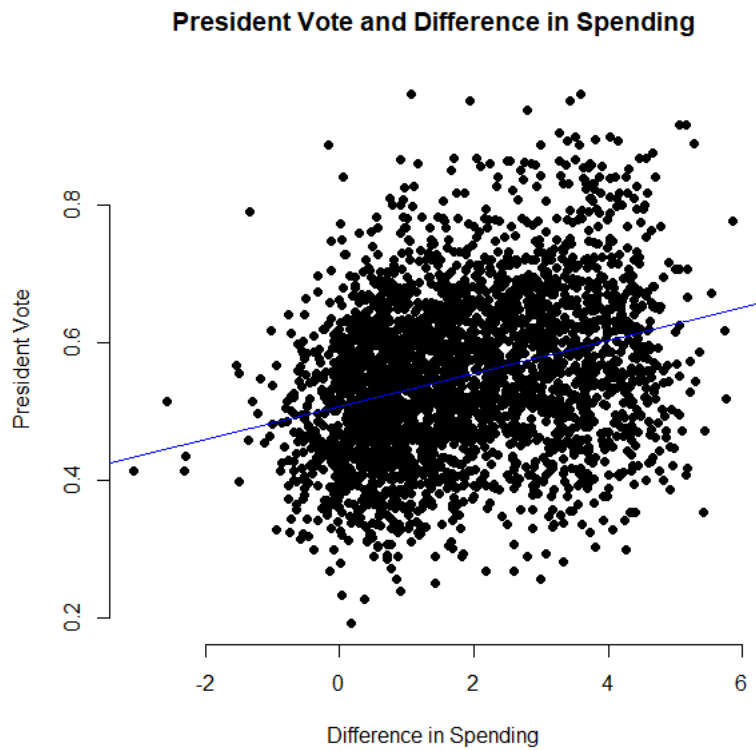
Residual standard error: 0.1104 on 3191 degrees of freedom

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.

```
plot(incumbents_subset$difflog, incumbents_subset$presvote,
main = "President Vote and Difference in Spending",
xlab = "Difference in Spending", ylab = "President Vote",
pch = 19, frame = FALSE)
abline(lm(presvote ~ difflog, data = incumbents_subset), col = "blue")
```



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

```
difflog_PresvoteRes <- presvote_Diffloglm["residuals"]  
difflog_PresvoteRes
```

4. Write the prediction equation.

$$y = .508 + .024(X) + .001$$

Where x = difflog score

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`.

```
voteshare_Presvote <- lm(voteshare ~ presvote, data = incumbents_subset)

summary(voteshare_Presvote)

Call:
lm(formula = presvote ~ difflog, data = incumbents_subset)

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 ***
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

> voteshare_Presvote <- lm(voteshare ~ presvote, data = incumbents_subset)
>
> summary(voteshare_Presvote)

Call:
lm(formula = voteshare ~ presvote, data = incumbents_subset)

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 ***
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

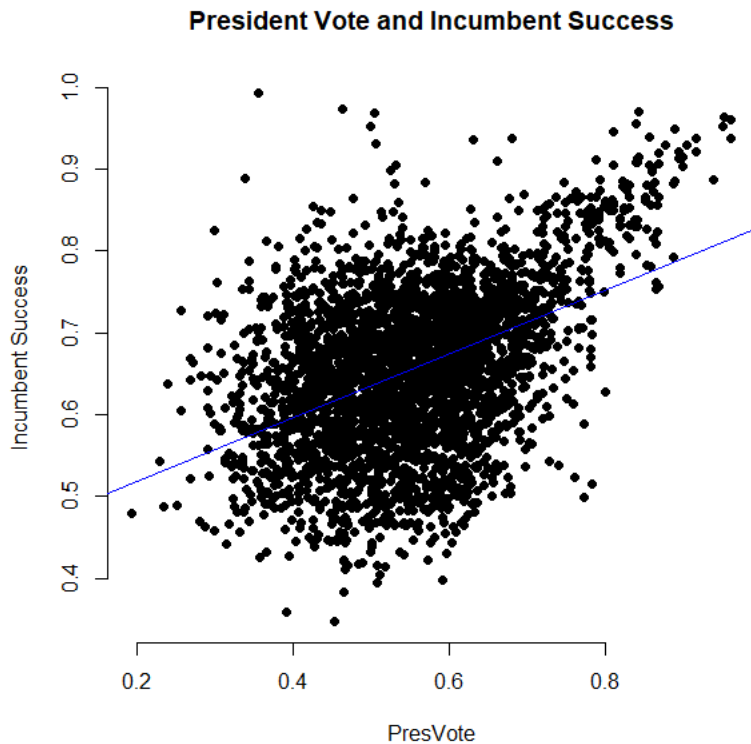
```

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

```

plot(incumbents_subset$presvote, incumbents_subset$voteshare,
     main = "President Vote and Incumbent Success",
     xlab = "PresVote", ylab = "Incumbent Success",
     pch = 19, frame = FALSE)
abline(lm(voteshare ~ presvote, data = incumbents_subset), col = "blue")

```



3. Write the prediction equation.

$$y = .441 + .388(X) + .013$$

x = Presvote Score

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.

```
regressionlm <- lm(unlist(difflog_VoteshareRes) ~ unlist(difflog_PresvoteRes),
data = incumbents_subset)
```

```
summary(regressionlm)
```

```
Call:
```

```
lm(formula = unlist(difflog_VoteshareRes) ~ unlist(difflog_PresvoteRes),
    data = incumbents_subset)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.25928	-0.04737	-0.00121	0.04618	0.33126

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.207e-18	1.299e-03	0.00	1
unlist(difflog_PresvoteRes)	2.569e-01	1.176e-02	21.84	<2e-16

```
(Intercept)
```

```
unlist(difflog_PresvoteRes) ***
```

```
---
```

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

```
Residual standard error: 0.07338 on 3191 degrees of freedom
```

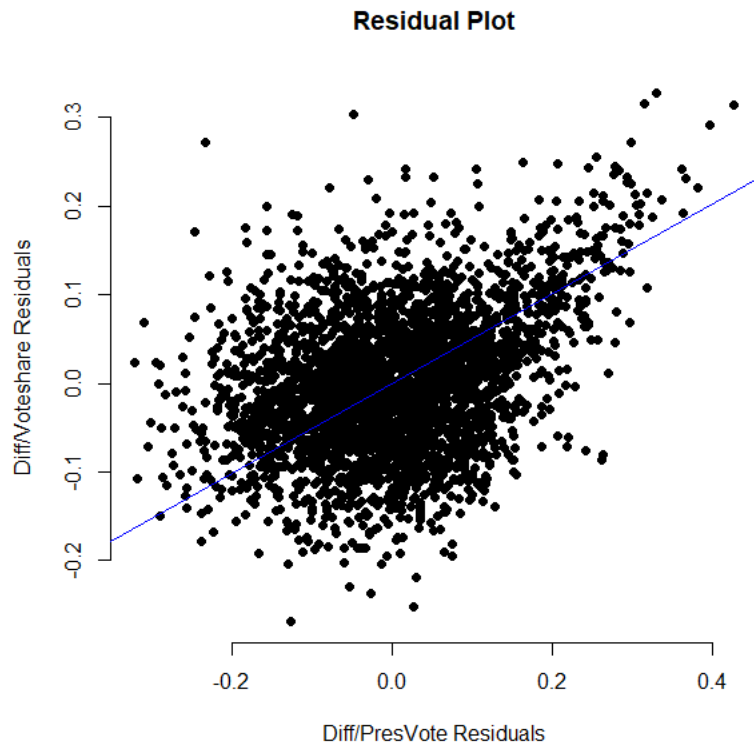
```
Multiple R-squared:  0.13, Adjusted R-squared:  0.1298
```

```
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.

```
plot(unlist(difflog_PresvoteRes), unlist(difflog_VoteshareRes),
main = "Residual Plot",
xlab = "Diff/PresVote Residuals", ylab = "Diff/Voteshare Residuals",
```

```
pch = 19, frame = FALSE)
abline(lm(unlist(difflog_PresvoteRes) ~ unlist(difflog_VoteshareRes),
data = incumbents_subset), col = "blue")
```



3. Write the prediction equation. $y = -5.207 + 2.569(X) + 1.176$
Here X is the point in Difflog/Presvote regression residuals.

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`.

```
model <- lm(voteshare ~ difflog + presvote,
data = incumbents_subset)
summary(model)
Call:
lm(formula = voteshare ~ difflog + presvote, data = incumbents_subset)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.25928	-0.04737	-0.00121	0.04618	0.33126

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.

$$\text{Voteshare} = .449 + .036(\text{difflog}) + .257(\text{presvote})$$

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 of the Difflog and Presvote regression model's t value in Question 4 and the Presvote variable in question 5's model t value are both 21.84.

The residuals of the $\text{difflog} / \text{presvote}$ regression explain how much of the variation in presvote is not explained by the difference in spending between incumbent and challenger in the district.

This explains why the value is the same as in the Question 5 model the Presvote t value is the remaining variation explained by Presvote once the difference in spending between incumbent and challengers is controlled for.