Answer Key: Problem Set 3

Applied Stats/Quant Methods 1

Jeffrey Ziegler

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 in* `.pdf` *form.*
- *This problem set is due before 8:00 on Friday October November 12, 2021. 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`*.*

```
# read in data
inc.sub <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2021/main/datasets/incumbents_subset.csv")

# (a) run regression where the outcome variable is \texttt{voteshare}
# and the explanatory variable is \texttt{difflog}
model_a <- lm(voteshare ~ difflog, data= inc.sub)
```

Table 1: Outcome variable is `voteshare` and the explanatory variable is `difflog`.

	Model 1
(Intercept)	0.579***
	(0.002)
difflog	0.042***
	(0.001)
R^2	0.367
Adj. R^2	0.367
Num. obs.	3193
RMSE	0.079

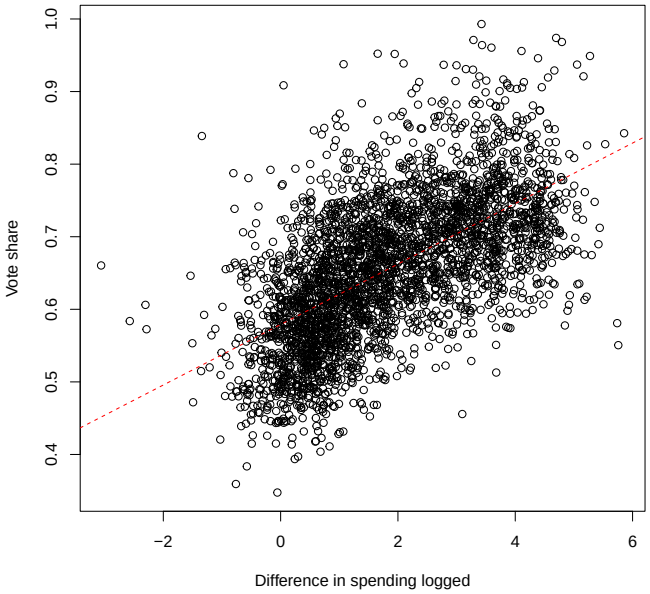
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

There is a positive and statistically reliable relationship between the amount of spending between the incumbent and challenger and the incumbent's vote share, such that a one unit increase in the logged difference in spending is associated with an average increase of 0.04 in the incumbent's vote share (4%)

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

```
plot(inc.sub$difflog, inc.sub$voteshare,
     ylab="Vote share", xlab="Difference in spending logged")
abline(model_a, col="red", lty=2)
```

Figure 1: Scatter plot of `difflog` and `voteshare`.



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

```
model_a.resid <- resid(model_a)
```

4. *Write the prediction equation.*

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

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`**.

```
# (a) run regression where the outcome variable is \texttt{presvote}
# and the explanatory variable is \texttt{difflog}
model_b <- lm(presvote ~ difflog, data= inc.sub)
```

Table 2: Outcome variable is `presvote` and the explanatory variable is `difflog`.

	Model 1
(Intercept)	0.508***
	(0.003)
difflog	0.024***
	(0.001)
R^2	0.088
Adj. R^2	0.088
Num. obs.	3193
RMSE	0.110

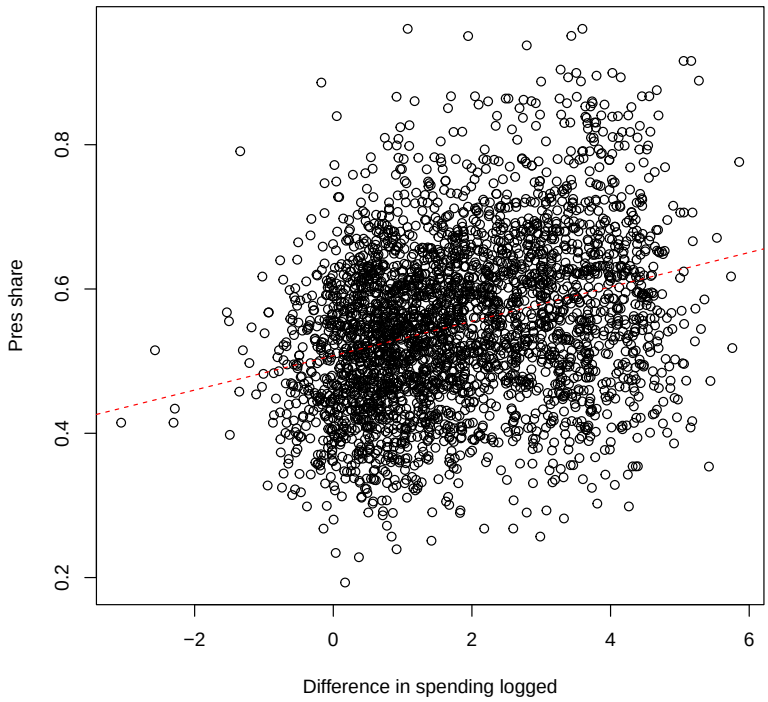
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

There is a positive and statistically reliable relationship between the amount of spending between the incumbent and challenger and the president vote share, such that a one unit increase in the logged difference in spending is associated with an average increase of 0.024 in the incumbent's vote share (2.4%)

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

```
plot(inc.sub$difflog, inc.sub$presvote,
     ylab="Pres share", xlab="Difference in spending logged")
abline(model_b, col="red", lty=2)
```

Figure 2: Scatter plot of `difflog` and `presshare`.



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

```
model_b.resid <- resid(model_b)
```

4. *Write the prediction equation.*

$$presvote = 0.508 + 0.024 \times \text{difflog}$$

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****.*

```
# (a) run regression where the outcome variable is \texttt{voteshare}
# and the explanatory variable is \texttt{presvote}
model_c <- lm(voteshare ~ presvote, data= inc.sub)
```

Table 3: Outcome variable is `voteshare` and the explanatory variable is `presvote`.

	Model 1
(Intercept)	0.441***
	(0.008)
presvote	0.388***
	(0.013)
R^2	0.206
Adj. R^2	0.206
Num. obs.	3193
RMSE	0.088

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

There is a positive and statistically reliable relationship between president vote share and voteshare, such that a one unit increase in the president's vote share is associated with an average increase of 0.388 in the incumbent's vote share (38.8%)

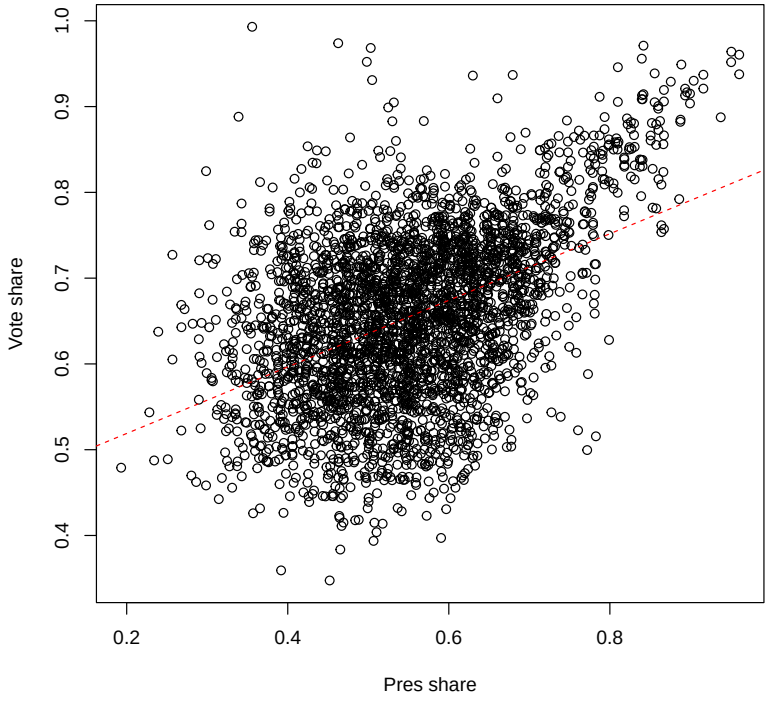
2. *Make a scatterplot of the two variables and add the regression line.*

```
plot(inc.sub$presvote, inc.sub$voteshare,
     ylab="Vote share", xlab="Pres share")
abline(model_c, col="red", lty=2)
```

3. *Write the prediction equation.*

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

Figure 3: Scatter plot of `presshare` and `voteshare`.



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.*

```
# (a) run regression where the outcome variable is \texttt{model_a.resid}
# and the explanatory variable is \texttt{model_b.resid}
model_d <- lm(model_a.resid ~ model_b.resid)
```

   There is a positive and statistically reliable relationship between `model_a.resid` and `model_b.resid`, such that a one unit increase in the residual error from model A is associated with an average increase of 0.257 in residual error from model B (25.7%).

Table 4: Outcome variable is `model_a.resid` and the explanatory variable is `model_b.resid`.

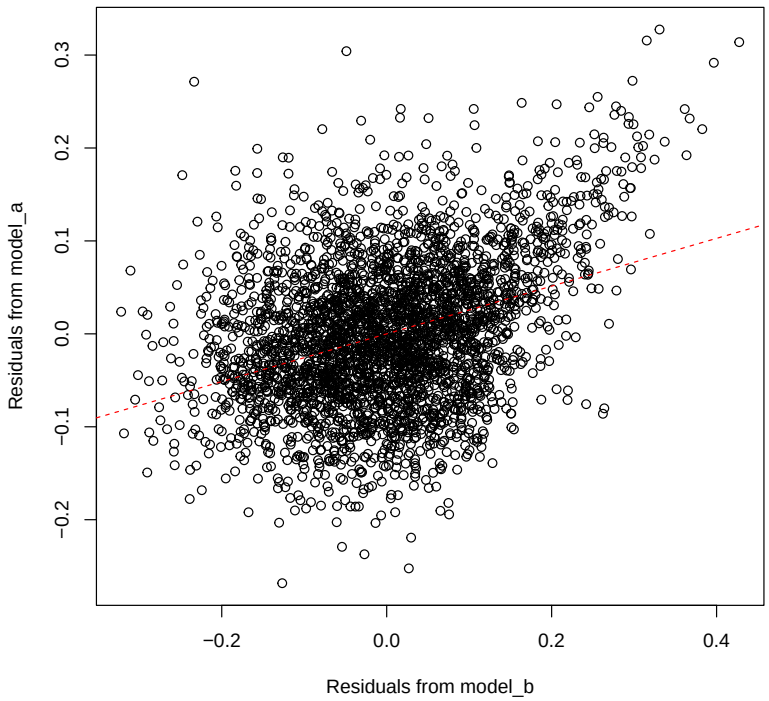
	Model 1
(Intercept)	−0.000
	(0.001)
model_b.resid	0.257***
	(0.012)
R^2	0.130
Adj. R^2	0.130
Num. obs.	3193
RMSE	0.073

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

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

```
plot(model_b.resid, model_a.resid,
     ylab="Residuals from model_a", xlab="Residuals from model_b")
abline(model_d, col="red", lty=2)
```

Figure 4: Scatter plot of `model_a.resid` and `model_b.resid`.



3. *Write the prediction equation.*

$$\texttt{model\_a.resid} = 0 + 0.257 \times \texttt{model\_b.resid}$$

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****.*

```
# (a) run regression where the outcome variable is \texttt{voteshare}
# and the explanatory variable is \texttt{difflog} and presvote
model_e <- lm(voteshare ~ difflog + presvote, data= inc.sub)
```

Table 5: Outcome variable is `voteshare` and the explanatory variables are `difflog` and `presvote`.

	Model 1
(Intercept)	0.449***
	(0.006)
difflog	0.036***
	(0.001)
presvote	0.257***
	(0.012)
R^2	0.450
Adj. R^2	0.449
Num. obs.	3193
RMSE	0.073

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

2. *Write the prediction equation.*

$$voteshare = 0.449 + 0.036 \times difflog + 0.257 \times presvote$$

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

   The coefficient for `presvote` in `model_e` is the same as the coefficient for `model_b.resid` in `model_d` because it represents the partial effect of `presvote` (the amount of co-variation between `presvote` and `voteshare` that is unexplained by `difflog`).