Problem Set 3

Applied Stats/Quant Methods 1

Due: November 11, 2024

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 11, 2024. 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`.

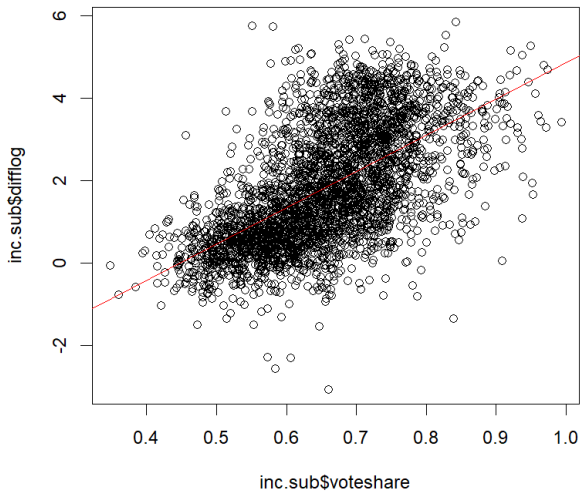
```
reg<-lm(inc.sub$difflog~inc.sub$voteshare)
regtable<-summary(reg)
```

Table 1: Regression Summary: Outcome VoteShare, Explanatory: DiffLog

	Dependent variable:
	difflog
voteshare	8.816^{***}
	(0.205)
Constant	-3.949^{***}
	(0.136)
Observations	3,193
R^2	0.367
Adjusted R^2	0.367
Residual Std. Error	1.144 (df = 3191)
F Statistic	$1{,}852.791^{***}$ (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

```
scpl<-plot(inc.sub$voteshare, inc.sub$difflog)
abline(reg, col = "red", lwd = 1)
```



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

```
res_reg<-residuals(reg)
resreg1sum<-head(res_reg)
```

The residuals of the model have been saved as an object named 'res_reg' in R. As the object is too large to add to the latex pdf, a sample of the head of it is below:

Table 2: Regression Residuals: Outcome Voteshare, Explanatory Difflog

1	2	3	4	5	6
-0.792	-0.326	-0.846	-0.259	-1.053	-1.221

4. Write the prediction equation. Prediction equation for this regression is:

$$\text{voteshare} = \beta_0 + \beta_1 \cdot \text{difflog} + \epsilon$$

where voteshare is the outcome variable, and difflog is the explanatory variable.

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`.

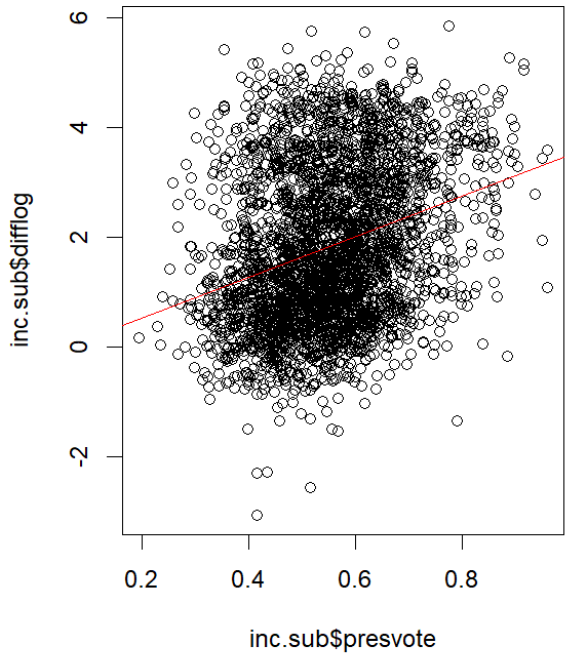
```
reg2<-lm(inc.sub$difflog~inc.sub$presvote)
regtable2<-summary(reg2)
```

Table 3: Regression Summary: Outcome Presvote, Explanatory: DiffLog

	Dependent variable:
	difflog
presvote	3.690^{***}
	(0.210)
Constant	-0.206^{*}
	(0.118)
Observations	3,193
R^2	0.088
Adjusted R^2	0.088
Residual Std. Error	1.374 (df = 3191)
F Statistic	307.715^{***} (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

```
scpl2<-plot(inc.sub$presvote, inc.sub$difflog)
abline(reg2, col = "red", lwd = 1)
```



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

```
res_reg2<-residuals(reg2)
resreg2sum<-head(res_reg2)
```

Similar to the previous question, the object which now holds the residuals (here, 'reg_res2') is too large to display in it's entirety in this pdf, and so the head of it is displayed below:

Table 4: Regression Residuals: Outcome Presvote, Explanatory Difflog

1	2	3	4	5	6
-1.168	-1.011	-1.082	0.313	-0.898	-1.625

4. Write the prediction equation. The prediction equation for this regression is:

$$\text{Presvote} = \beta_0 + \beta_1 \cdot \text{DiffLog} + \epsilon$$

with the explanatory variable being difflog and the outcome variable being presvote.

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`.

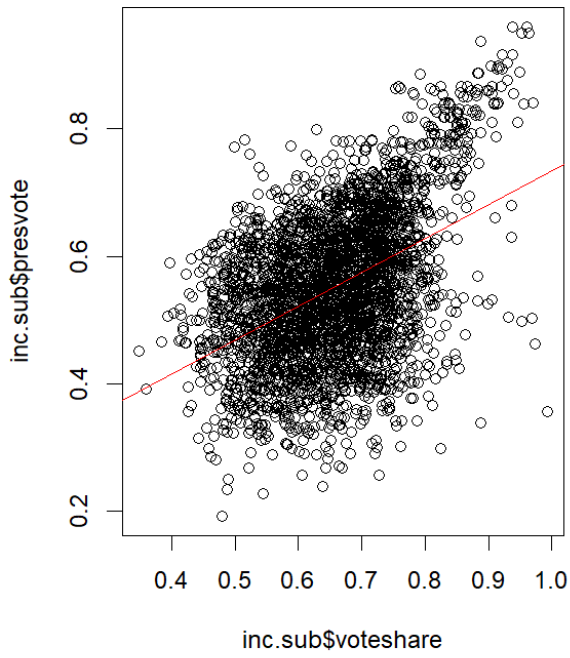
```
reg3<-lm(inc.sub$presvote~inc.sub$voteshare)
regtable3<-summary(reg3)
stargazer(reg3, type="latex",title="Regression Summary: Outcome VoteShare
    , Explanatory Presvote", out = file.path(fp, "pso3q3regsum.tex"))
```

Table 5: Regression Summary: Outcome VoteShare, Explanatory Presvote

	Dependent variable:
	presvote
voteshare	0.530***
	(0.018)
Constant	0.204***
	(0.012)
Observations	3,193
R^2	0.206
Adjusted R^2	0.206
Residual Std. Error	0.103 (df = 3191)
F Statistic	826.950*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

```r
scpl3<-plot(inc.sub$voteshare, inc.sub$presvote)
abline(reg3, col = "red", lwd = 1)
```



3. Write the prediction equation. Prediction equation for this regression is:

$$\text{voteshare} = \beta_0 + \beta_1 \cdot \text{presvote} + \epsilon$$

where voteshare is the outcome variable, and presvote is the explanatory variable.

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.

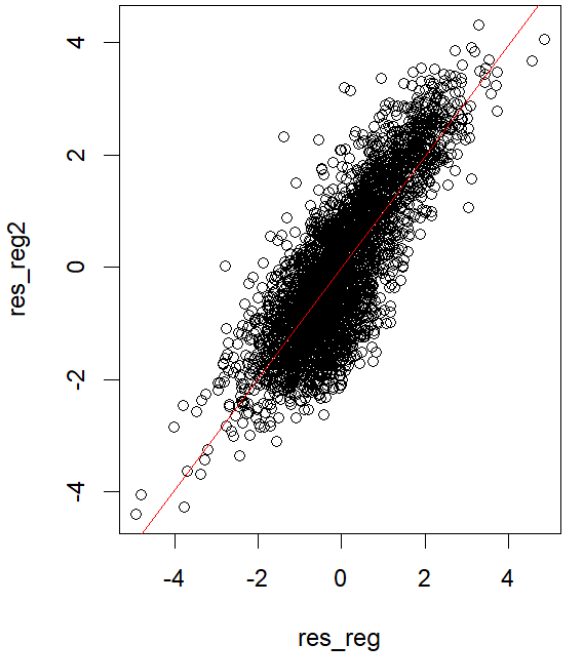
```
reg4<-lm(res_reg2~res_reg)
regtable4<-summary(reg4)
```

Table 6: Regression Summary: Outcome Q1 Residuals, Explanatory Q2 Residuals

	Dependent variable:
	res_reg2
res_reg	0.990***
	(0.012)
Constant	−0.000
	(0.014)
Observations	3,193
R^2	0.680
Adjusted R^2	0.680
Residual Std. Error	0.778 (df = 3191)
F Statistic	6,771.208*** (df = 1; 3191)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

```
scpl4<-plot(res_reg, res_reg2)
abline(reg4, col = "red", lwd = 1)
```



3. Write the prediction equation. Prediction equation for this regression is:

$$\text{res\_reg} = \beta_0 + \beta_1 \cdot \text{res\_reg2} + \epsilon$$

```
where the object 'res_reg' (the residuals from the first regression) is the
outcome variable, and object 'res_reg2' (the residuals from the second
regression) is the explanatory variable.}
```

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`.

```
reg5 <- lm(inc.sub$voteshare ~ inc.sub$difflog + inc.sub$presvote)
```

Table 7: Regression Summary: Outcome Voteshare, Explanatory difflog and presvote

	Dependent variable:
	voteshare
difflog	0.036***
	(0.001)
presvote	0.257***
	(0.012)
Constant	0.449***
	(0.006)
Observations	3,193
R^2	0.450
Adjusted R^2	0.449
Residual Std. Error	0.073 (df = 3190)
F Statistic	1,302.947*** (df = 2; 3190)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Write the prediction equation. Prediction equation for this regression is:

$$\text{voteshare} = \beta_0 + \beta_1 \cdot \text{difflog} + \beta_2 \cdot \text{presvote} + \epsilon$$

where voteshare is the outcome variable, and difflog and presvote are the explanatory

variables.

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 only things which both regressions have in common are
the number of observations - as both utilise the same dataset.
Despite the overlap of variables across both, they are modelling
different relationships between these variables and so the coefficients,
standard errors, etc. are not maintained across both.
```