Problem Set 3

Applied Stats/Quant Methods 1

Due: November 12, 2021

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 class on Friday 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.

Create object votesharedifflog from regression on voteshre and difflog

Use lm function. Code is as follows....

- votesharedifflog ¡- lm(incumbents dollar sign voteshare   incumbents dollar sign difflog)

- votesharedifflog

Intercept = 0.58, slope = 0.04

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

Use plot() and abline functions in R

- plot(votesharedifflog)

- abline(votesharedifflog)

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

Use residuals() function in R on votesharedifflog

- residuals1 ¡- residuals(votesharedifflog)

Print residuals1

- residuals1

4. Write the prediction equation.

Equation = intercept + slope multiplyed by random x value

E.g. use 15 as random x value

- 0.58 + 0.04*15

Predicted value is 1.18

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`.

   Create object difflogpresvote from regression of presvote and difflog

   Use lm function for regression

   - difflogpresvote ¡- lm(incumbents*presvote incumbents*difflog)

   Print difflogpresvote

   Intercept = 0.51, slope = 0.02

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

   Make scatterplot of difflogpresvote and add regression line using plot and abline functions in R

   - plot(difflogpresvote)

   - abline(difflogpresvote)

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

Save residuals of model in sperate object residuals2

- residuals2 ¡- residuals(difflogpresvote)

Print residuals2 - residuals2

4. Write the prediction equation.

   Equation = intercept + slope multiplyed by random x value

   E.g. use 5 as random x value

   - 0.51 + 0.02*5

   Predicted y value is 0.61

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`.

   Create object presvotevoteshare from regression on presvote and voteshare

   Use lm function for regression in R

   - presvotevoteshare ¡- lm(incumbents$presvote$ incumbents$voteshare)

   Print presvotevoteshare

   - presvotevoteshare

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

   Use plot and abline functions in R

   - plot(presvotevoteshare)

   - abline(presvotevoteshare)

3. Write the prediction equation.

   Equation is y = 0.2036 (intercept) + 0.5304 (slope) * random x

   E.g. X value = 9

- 0.2036 + 0.5304*9

Predicted y value = 4.9772

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.

   Run regression between outcome var residuals1 and explanatory var residuals2

   Create object combresiduals from regression

   Code is as follows...

   - combresiduals ¡- lm(residuals1   residuals2)

   Print combresiduals

   - combresiduals

   Intercept = 4.498, slope = 6.866

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

Use plot and abline functions on combresiduals

- plot(combresiduals)

- abline(combresiduals)

3. Write the prediction equation.

Add intercept and slope and multipy slope by random x value (i.e. residual2 value)

E.g. X value = 20

- 4.498 + 6.866*20

Predicted y value when x is 20 = 141.818

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`.

   Multi-Variate Regression

   Create object votesharediffpresvote from regression.

   Use lm function. Use + sign to add on a second predictor (or x) variable

   - votesharediffpresvote ¡- lm(voteshare   difflog+presvote, data=incumbents)

   Print votesharediffpresvote

   - votesharediffpresvote

   Intercept = 0.44864, slopes = 0.03554 and 0.25688

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?