

Problem Set 8

Quantitative Political Methodology (L32 PS 363)

Instructions

1. Print out and write your WUSTL ID at the top of **each** page, and complete each exercise in the space allotted. You may attach extra pages if the space provided is not sufficient, but please indicate that you have done so below.

Pages attached: _____

2. Please **show your work** if possible. You may lose points by simply writing in the answer. If the problem requires you to use R, please include the code you used to get your answers. If you are not sure if work needs to be shown for a particular problem, please ask a TA or post a question on Facebook.
3. The various pages of your homework should be **stapled together** (no paper clips please). If pages are lost because of a lack of a staple, no credit will be granted for that portion of the homework.
4. This problem set is **due at the beginning of class on Wednesday, December 6, 2017**. No late assignments will be accepted.
5. Total available points for this homework: 30.

Question 1 (Total: 13 points)

In this question, you will run several regressions and examine the implication of R^2 using the `newhamp` dataset in the `faraway` package.

```
library("faraway")
data("newhamp")
colnames(newhamp)
```

(a) (6 points) Run six regressions using `pObama` as a dependent variable and the following independent variable(s):

1. `votesys` only
2. `votesys` and `povrate`
3. `votesys`, `povrate`, and `pci`
4. `votesys`, `povrate`, `pci`, and `Dean`
5. `votesys`, `povrate`, `pci`, `Dean`, and `white`
6. `Dean` only

(b) (4 points) Get R^2 and adjusted R^2 from each of the six models above and fill in the table below. The first column is for R^2 s, and the second column is for adjusted R^2 s.

(Optional: If you summarize changes in R^2 s and adjusted R^2 s graphically using `plot`, you will receive a 1 extra point, conditional on proper title, labels, and legend. There is no “correct” style!)

	R^2	Adjusted R^2
Model 1		
Model 2		
Model 3		
Model 4		
Model 5		
Model 6		

- (c) (3 points) Which model has the highest R^2 ? What happens to R^2 when you add a lot of variables in your model?

Question 2 (Total: 17 points)

In this question, you will run several regressions and create an add variable plot (see Lecture 16) in R using the `incumbents_subset.csv` dataset. Include the code.

- (a) (4 points) We are interested in knowing how 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`.
 2. Make a scatterplot of the two variables and add the regression line.
 3. Save the residuals of the model in a separate object.
 4. Write the prediction equation.
- (b) (4 points) 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`.
 2. Make a scatterplot of the two variables and add the regression line.
 3. Save the residuals of the model in a separate object.
 4. Write the prediction equation.

- (c) (3 points) 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`.
 2. Make a scatterplot of the two variables and add the regression line.
 3. Write the prediction equation.
- (d) (3 points) 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 part (a) and the explanatory variable is the residuals from part (b).
 2. Make a scatterplot of the two residuals and add the regression line.
 3. Write the prediction equation.

- (e) (3 points) 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`.
 2. Write the prediction equation.
 3. What is it in this output that is identical to the output in part (d)? Answer in one sentence!
 4. Reflect on your finding. Don't write anything. Just think about it.