Problem Set 6

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.

<div align="center">Pages attached:_____</div>

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, November 15, 2017**. No late assignments will be accepted.

5. Total available points for this homework: 35.

Question 1   (Total: 20 points)

In this question, we use the `prestige` dataset in the `car` library. First run the following commands:

```
install.packages(car)
library(car)
data(Prestige)
help(Prestige)
```

We would like to study whether individuals with higher levels of income have more prestigious jobs. Moreover, we would like to study whether professionals have more prestigious jobs than blue and white collar workers.

(a) (2 points) Create a new variable `professional` by recoding the variable `type` so that professionals are coded as 1, and blue and white collar workers are coded as 0. (Hint: `ifelse`.)

(b) (4 points) Run a linear model with prestige as an outcome and `income, professional`, and the interaction of the two as predictors. (Note: this is a continuous $\times$ dummy interaction.)

(c) (2 points) Write the prediction equation based on the result.

(d) (3 points) Interpret the coefficient for `income`.

(e) (3 points) Interpret the coefficient for `professional`.

(f) (3 points) What is the effect of a \$1,000 increase in income on prestige score for professional occupations? In other words, we are interested in the marginal effect of income when the variable `professional` takes the value of 1. Calculate the change in $\hat{y}$ associated with a \$1,000 increase in income based on your answer for (c).

(g) (3 points) What is the effect of changing one's occupations from non-professional to professional when her income is \$6,000? We are interested in the marginal effect of professional jobs when the variable `income` takes the value of $6,000$. Calculate the change in $\hat{y}$ based on your answer for (c).

Question 2   (Total: 15 points)

In this question, we analyze the `newhamp` dataset in the `faraway` library again:

```
library(faraway)
data(newhamp)
help(newhamp)
```

In the 2008 Democratic Party Primary in New Hampshire, Hillary Clinton defeated Barack Obama by surprise, contrary to the expectations of pre-election opinion polls. As we know, some districts in New Hampshire used paper ballots, counted by hand, while others used optically scanned ballots.

Analyzing the primary results, one of your smartest friends, who has never been wrong, noticed that there is a weird pattern in the outcomes: among the paper ballots, Obama had more votes than Clinton, while Clinton defeated Obama on just the machine-counted ballots. His finding seems to be grand-breaking because we believe that the ways in which votes are counted should not make any difference in election outcomes. Let's evaluate his claim.

Table 1. Determinants of Obama Votes

| | DV = Proportion Obama ||
	(1)	(2)
Hand-Count		
	( )	( )
Proportion Dean		
		( )
Proportion of Non-Hispanic		
		( )
Income per capita		
		( )
Constant		
	( )	( )
N	276	276
R^2	0.083	0.501
Adjusted R^2	0.080	0.494
Residual Std. Error	0.068 (df = 274)	0.051 (df = 271)
F Statistic	24.932* (df = 1; 274)	68.030* (df = 4; 271)
Note:		*p<0.05

(a) (3 points) Run a linear regression with proportion voting for Obama (pObama) as an outcome and the voting system used (votesys) as a predictor. Based on the result, fill in the coefficients and standard errors in the first column of the table above. Provide a "star" if the p-value of the coefficient is smaller than 0.05.

(b) (2 points) Interpret the coefficient for the hand-counted system. According to your answer for (a), is your friend correct?

(c) (3 points) Re-run a linear regression including three additional variables: (1) the proportion of voters for Howard Dean in the 2004 Democratic primary (Dean), (2) the proportion of non-Hispanic whites (white), and (3) per capita annual income (pci). Fill in the second column of the table above.

(d) (2 points) Interpret the coefficient for the hond-counted system in Model 2. According to your answer for (c), is your friend correct?

(e) (3 points) Interpret the coefficient for the proportion of votes for Dean in Model 2. How does a one percentage point increase in the proportion of votes for Dean in 2004 affect the proportion of votes for Obama in 2008?

(f) (2 points) According to Model 2, what is the predicted voteshare of Obama when $\texttt{votesysH} = 1$, $\texttt{Dean} = 0.3$, $\texttt{white} = 0.9$, and $\texttt{pci} = 0$?