The New School for Social Research
Advanced Econometrics 1
Fall 2017
Christian Schoder
Jangho Yang

Assignment 6

Due Nov 13 (Mon) 6:00 pm

1. Hierarchical logistic regression : Load the attached two data files. The first data is Pew Research Center polls taken during the 2008 election campaign and the second data is about 2008 election result.

   (a) Take the first data file. 1) Subset the data so that you have all states but Hawaii, Alaska, and Washington D.C and have only four columns "state," "marital," "heat2," and "heat4." 2) If no data is available in "heat2," replace **na** for the corresponding value in "heat4." If neither of "heat2" and "heat4" has data, erase the corresponding row. 3) Subset the data so that you only have "dem/lean dem" and "rep/lean rep" in the "heat2" column. 4) Change the label of all the variables but 'married' in the "marital" column to 'other.' Name your manipulated data set *data_ before*.

   (b) Using *data_ before*, we will create new data frame. The task is to do some calculations for each state. The final data, therefore, will have 48 states in the rows and the target variables in the columns. Calculate 1) the proportion of the democratic supporters, 2) the proportion of the married people, 3) the ratio of the married people among the democratic supporters to the total married people, 4) the ratio of non-married among the democratic to the total non-married people, 5) the difference of 3) and 4). Multiply all values by 100 to convert to percentage. Show the first 5 lines of the data set. It will be helpful to use the function **ddply()** of the package **"plyr"**.

   (c) Take the second data file. Subset the data so that 1) you have all but three states, Hawaii, Alaska, and Washington D.C, and 2) only two columns "state," and "vote_Obama_pct." Show the first 5 lines of the data set.

   (d) Use a hierarchical logistic regression predicting vote intention given state, using the indicator for being married. Set up a proper link function and come up with a reasonable hyper prior for your coefficients, if possible. Plot your inference for the predicted vote share by state, along with the actual vote intention, plotting them vs. Obama's actual vote share in the second data file. You need to calculate the proportion of the married people in each state from *data_ before* to answer this question.

   (e) The marriage gap is defined as the difference of Obama's vote share among married and non-married people ("other"). Try to figure out how to infer this marriage gap from your model in (d). Plot your inference for the marriage gap, along with the raw marriage gaps from the data, plotting them vs. Obama's vote share.