**POL 212: Midterm Exam Part 2**: Working alone, please answer the following questions showing all work/calculations. You may use R to estimate the models and do the plots. You can use any software package you want to manage the data. This portion of the exam is due Thursday morning by 10 AM.

On the website, there is a data set called california.dta, a Stata data set that can be imported into R using the foreign package and the read.dta function. For the dependent variable, use the variable called voteshare. This gives the vote share for the incumbent U.S. House member from the congres-sional district. The district identifier is is the variable district. For the two independent variables, use the variable partyid (coded 1 if a Democrat and 0 if a Republican) and termsserved which is the number of terms served in the House by the MC. For this exam, do the following:

1. Estimate an OLS regression model and give a substantive interpretation of the model including giving hypothesis tests for the two independent variables. In this part of the answer, put your results in a publication quality table. (25 points)

2. Diagnostics: provide the full complement of diagnostic tests for this model. Your analysis should include the following: 1) a plot of the hat diagonal elements; 2) an outlier test giving both the standard \( p \) and the Bonferroni \( p \) value for the maximal student’s \( r \); a plot of the hat diagonal elements by the corresponding student’s \( r \); and plots of \( DFFITS \), \( DFBETAS \), and \( Cook’sD \). All plots should be interpreted and any observations that merit “flagging” should be discussed/pointed out. Please label all plots neatly. (25 points)

3. Reestimate the regression coefficients’ standard errors using a standard White HC estimator as well as the Long and Ervin HC estimator. Do your inferences/hypothesis test conclusions change? (10 points)

4. Create two dummy variables of the following type. The variable LA should be scored 1 for all districts in the Los Angeles area (these are districts 29-40, 46-48) and 0 otherwise; the variable SF should be scored 1 for all districts in the SF Bay Area (these are districts 1, 6-17). Reestimate your model in question 1 with these two indicators. Provide a full interpretation of the model including hypothesis tests. Put results in a publication quality table. Additionally, plot the response function for Democratically held districts in the Bay Area, LA, and non-Bay Area, non-LA districts. How does this plot correspond to your estimates? (40 points)