Assignment 3

1.   Consider the data file `senate_expanded.dta`. The variable `inc_2p_share` is the two party vote share of the incumbent candidate. The variable `inc_tenure` is the number of years the incumbent has served in the Senate, the variable `ch_qual` is a variable indicating the quality of the challenger, based on their previous office holding experience, with 4 indicating former House member, 3 indicating governors, lesser statewide elected officials, and former Senators, 2 indicating local elected officials, 1 indicating state legislators, and 0 indicating no prior office-holding experience. The variable `inc_pos` is the incumbents Common Space score, a measure of voting ideology with higher values indicating more conservative voting records. The variable `st_uemp` is the state unemployment rate.

(i)   Run a regression with `inc_2p_share` as the dependent variable and `inc_tenure`, `ch_qual`, `inc_pos`, and `st_uemp` as independent variables and report the results.

(ii)   Interpret each of the coefficients in the regression including the constant term.

(iii)   Comment on the statistical significance of the independent variables.

(iv)   Report the results of a hypothesis test that the coefficients on all four independent variables are zero in the population, using the R-squared to compute this value.

(v)   Test the null hypothesis that the coefficients on `inc_pos` and `st_uemp` are zero in the population using the test command in stata and verify this calculation using the formula based on the restricted and unrestricted R-squares.

(vi) Test the null hypothesis that the coefficient on `inc_pos` is 0.5 in the population. Intepret the results of this test.

(vii) Calculate the AIC and BIC for this model. Calculate the AIC and BIC for the same model with `st_uemp` omitted. What do the different values of AIC and BIC tell us about the relative fit of the two models. (hint: to ensure that the second regression has the same sample size, use the command `reg inc_2p_share inc_tenure ch_qual inc_pos if st_uemp != .`)

(viii) Consider a freshman incumbent running for re-election, whose position is 0.5 running against a challenger with no prior office-holding experience, in a state with an unemployment rate of 5%. Predict this incumbent's vote share and form a 95% confidence interval for this prediction.

(ix) Check for heteroskedasticity by plotting the squared residuals against each of the independent variables. Interpret the results.

(x) Check for heteroskedasticity using the white test. Interpret the results.

(xi) Apply weighted least squares to the regression using the same independent variables to predict the variance of the error term.

(xii) Estimate a linear model with heteroskedasticity using maximum likelihood estimation.

(xiii) Check for normality of the error terms using a kernel density plot. Interpret the results.

(xiv) Check for normality of the error terms using Jarque-Berra and Kolmogorov Smirnov tests. Interpret the results.

(xv) The p-value for the Kolmogorov Smirnov test is larger than the p-value from the Jarque-Berra test. Comment on why this is likely to be the case.

(xvi) What are the various consequences of violating homoskedasticity and normality?

(xvii) Given the results of (ix) through (xv), is there anything you would change about how you would analyze the data? If so, reanalyze the data given your findings.