**HOMEWORK #6**

**Part 1. Ordinary Least Squares (OLS) Regression**

Use the data <affairs.dta>. You will be using these data to examine the correlates of extra-marital affairs (naffairs).

The predictors will include **male** (=1 if male), **age**, **yrsmarr** (years married) **educ** (years of education), **kids** (=1 if have any children, 0 if no children), **vryhap** (1 if very happy, 0 otherwise), and **relig** (=5 if a very religious, 1 for anti)

1. Explore the distribution of naffairs. Does this variable appear to be normally distributed? Why or why not? (Provide evidence from Stata)
2. Estimate an **OLS (least squares regression) model** predicting naffairs using all of the predictors listed above. Paste your regression output below.
3. Explain the overall fit of the model and note which indicator(s) you used (for model fit).
4. Based on your regression output, which coefficients are statistically significant? Provide substantive interpretations of the coefficients that are statistically significant.

**Part 2. Linear Probability**

1. Generate a dichotomous indicator called “affairs” using the variable naffairs. The new variable (affairs) should be equal to one if the individual reported any affairs (naffairs > 0), and zero for no affairs. Paste a tabulation of your new variable.
2. Estimate a **linear probability model (LPM)** using the new variable (affairs) as the dichotomous outcome and the same predictor variables from part 1. Paste your regression output below.
3. Explain the overall fit of the model and note which indicator(s) you used (for model fit).
4. Based on your regression output, which coefficients are statistically significant? Provide substantive interpretations of the coefficients that are statistically significant.

**Part 3. Logistic Regression**

1. Estimate a **logistic regression** using the new variable (affair) as your dichotomous outcome and the same predictor variables. (*Important tip: Use the factor notation for dummy variables in order to estimate the predicted probabilities below).* Paste your output below.
2. Explain the overall fit of the model and note which indicator(s) you used (for model fit).
3. Provide a statistical and substantive interpretation of the **odds ratios** for **age** and for **relig**.
4. Provide an interpretation based on the **percent change in the odds** for **age**, and for **relig**.
5. Following the logistic regression, use the post-estimation command “margins” to produce the predicted probability of an extra-marital affair for men and women holding everything else at the mean. Describe the results.

**Part 4. Comparing the Models**

Look at the predictor coefficients across all 3 models. What do we learn in each model? Does it make sense to dichotomize the number of affairs (any affairs / no affairs) to use it in a linear probability or logistic regression model, or does it make more sense to use naffairs as continuous in OLS? Provide justification to support your conclusions.