**Stata EDA homework**

For this homework, use the vitamin D data that is on the course website and use Stata to perform all analyses. The purpose of the study is described in the following abstract:

Background: African-Americans suffer disproportionately from diabetes and cardiovascular disease, and are significantly more likely to have sub-optimal levels of circulating vitamin D [25(OH)D].The results of epidemiologic and observational studies suggest that there is a link between vitamin D deficiency and the risk of cardiometabolic disorders, underscoring the importance of maintaining healthy levels of 25(OH)D.

Objective: To investigate whether daily supplementation with 4,000 international units (IU) of vitamin D3 for one year eliminates any disparity in circulating levels of 25(OH)D between African-American and Caucasian men.

Design: Serum levels of 25(OH)Dwere measured in 28 subjects who received 4,000 IU of vitamin D3 for one year.

The variables are as follows:

Age: age in years

Baseline: Baseline levels of Vitamin D,

Exit: Exit levels of Vitamin D

AA: =1 if African-American, =0 if Caucasian

Answer the following questions and show your Stata code.

1. Create graphical displays of pre and post vitamin D levels. Denote race in the graphical display.
2. Compare the mean age in the Caucasian and African-American subjects. Report the statistical significance based on a parametric and a non-parametric approach.
3. Does vitamin D change significantly from baseline to follow-up? Test this in the whole sample, and also within each race.
4. Does the change in vitamin D from baseline to exit differ by race?
5. Create a binary variable for age (>65 vs. <=65 years old). Is age (young vs. old) associated with race in this dataset?
6. Are vitamin D levels normally distributed at baseline? Is there a transformation that would provide a more normal distribution? What about at Exit? What about the difference from baseline to exit? Given these answers, should you transform vitamin D and if so, how?