Stata Homework 2 (Homework 10)

You will be using the Ischemic Heart Disease Dataset for this homework. The dataset and a codebook (with description of the dataset) are on the course website. Please paste all relevant figures asked in the questions below into the homework document and answer all questions in words as appropriate. Paste your “.do file” at the end.

1. Open a “.do” file
2. Read the ischemic heart disease dataset into Stata.
3. Create univariate one or more graphical displays of totalcost to determine if a transformation may be needed to symmetrize totalcost. Show the graphical display(s). Describe how you determined a transformation was or was not needed and what the transformation should be if one is needed. (Hint: you should not take logs or raise costs to a negative power if there are zeros. You can create a new variable costs+1 which will solve the problem of taking logs, etc.). Call the ‘transformed’ variable transcost.
4. Create a boxplot display of transcost. Intead of using the defaults, change capstyle, how the median is indicated, and make the boxplot horizontal instead of vertical. Also indicate on the axis what the axis represents (e.g., square root of costs in dollars).
5. Create boxplots of transcosts for patients who have no complications and those who have one or more complications. Be sure to include some way to indicate which boxplot is for which complication level.
6. Make a scatterplot of totalcost and duration of treatment, and a scatterplot of transcost and duration of treatment. Label axes appropriately. Which pair would be a more appropriate choice for performing a linear regression and why?
7. Make a figure of the raw data of variables selected in 6. Add the regression line and the 95% confidence interval for the fit. Explain your choice of a quadratic versus linear fit.
8. Create a diagnostic plot to evaluate residuals and comment on the appropriateness of the assumption of constant variance of the residuals.
9. Close your do file.