Introductory Statistical Analysis Session 4 Task

ASSIGNED ON: Wednesday 2/27/2013
DUE Tuesday 3/12/2013 9pm

1. Please specify ONE hypothesis to work with for the rest of the sessions. Identify the outcome variable and one MAIN covariate (independent variable) from your hypothesis. "Explore the association between X and Y": let either X or Y be the outcome and the other be the main independent variable "XYZ are risk factors of W": let W be the outcome and choose ONE from XYZ to be the main independent variable.

2. Choose 3 other variables from the data you have organized, so that you have a total of 2 categorical and 2 quantitative covariates, and the outcome variable, to work with in ALL subsequent tasks.

3. Univariate graphic display 5 variables using ALL of the following techniques:
Bar plot Table Density plot Stem-and-leaf plot

4. Perform 3 transformations to serially reduce skewness for 1 of the 2 continuous independent variables. Comment on which transformation works better, numerically justify your argument. You are invited but not required to calculate the descriptive statistics and generate displays of the transformed variable.
!Save the transformed data points. You may append columns to the current spreadsheet!

5*. Categorical variables give rise to subgroups. Identify all the subgroups formed as the result of one (of the two) categorical variables and calculate descriptive statistics for each of the subgroups. If the categorical variable if dichotomous, there are two subgroups, etc.

How to turn in this assignment:
IN xlsx FORMAT. E-mail me (li.xie@nemours.org) back the Excel Workbook, name it like this:
LastnameFirstinitial_session4hw_mmddyyyy

For example, my name is Li Xie, so I would name my file "xiel_session4hw_3012013"
Thank you and see you all on Wednesday!

For any questions regarding either the concepts or the programming, please contact me via e-mail first, then we could set up a time to meet. I am here to help.