Introductory Statistical Analysis Session 7 Task

ASSIGNED ON: Wednesday 3/20/2013
DUE Tuesday 3/26/2013 9pm
You are to find out how to do the hypothesis tests on your own for this Task. I suggest you at least finish steps 1-4 by Sunday 3/24/2013. This Task is due in 1 week.

Out of the 5 variables you have chosen since Session 4 Task, if there are 2 continuous variables, then work with both. If there are 3 continuous variables AND 1 of the 3 is the outcome of interest, then work with the outcome AND 1 of the other 2 continuous variables.
Example: If you are interested in finding risk factors of high BMI from family history of diabetes, age, race/ethnicity and weight, outcome of interest is BMI, a continuous variable, age and weight are continuous variables. For this Task, pick BMI and age or BMI and weight.

If the risk factors you picked are family history of diabetes, age, race/ethnicity and gender, only BMI and age are continuous. For this Task, work with these 2 continuous variables.

1. **For both continuous variables, calculate GENDER-SPECIFIC average, median, stdev, var, skewness, min, max. You should obtain 14 statistics. Make a table for them.**
2. **Make a box plot for each of these 2 continuous variables BY GENDER**
3. **Scatter-plot these 2 continuous variable against each other, calculate the CORRELATION (excel command "=CORREL") and R-SQUARED (plot then "add trendline" then "display R-squared value on chart").**
4. **For EACH of the 2 continuous variables, select the appropriate hypothesis test to test gender-specific differences. State the null hypotheses and ALL alternative hypotheses for the tests you chose, Justify for your choices using information from steps 1-3.**
   If you choose any non-parametric tests, indicate the parametric assumptions potentially violated and explain the reasons of your suspicion. You are NOT required to run F or Shapiro-Wilk tests.
   There should be 4 hypotheses associated with each hypothesis test: 1 null hypothesis, 2 one-sided hypotheses, 1 two-sided hypothesis.
5. **Run BOTH the parametric tests and their non-parametric analogues on EACH of the 2 continuous variables. For each of the 2 continuous variables, you should obtain 1 two-sided p-value from each of the 2 hypothesis tests. Interpret the 4 p-value's verbatim from the lecture slide in the context of the variables involved AND the null hypotheses.**
   For example, for continuous variable #1, if you choose two-sample unpaired t test, then run both this test and the Mann Whitney U test. Each of these 2 tests is associated with a p-value.
6. **Provide a 2-sentence comment on each of the 4 test results. THEN Compare and contrast the parametric and non-parametric test results for the same continuous variable.**

SPSS Resource: http://math.csuci.edu/ocspss/hypothesistesting.pdf
Excel Resource: Open Excel then hit F1; Google; Youtube.
Online calculators: for t test: http://www.graphpad.com/quickcalcs/ttest1.cfm
for Wilcoxon signed-rank test: www.vassarstats.net/wilcoxon.html
for Mann-Whitney U test: www.vassarstats.net/utest.html

How to turn in this assignment:
E-mail me (li.xie@nemours.org) your file (any format) and name it:
LastnameFirstinitial_session5hw_mmddyyyy
For example, my name is Li Xie, so I would name my file "xiel_session7hw_03252013"

For any questions regarding either the concepts or the programming, please feel free to contact me.