**Workshop Exercise 5**

**Exercise V: Statistical Procedures**

**Dataset = pulse.xls**

In this exercise we will use the Excel file Pulse.xls, which is located in the labdata folder. The codebook for this dataset can be found in the workbook Appendix.

1. Import the Pulse.xls Excel file using Proc Import Commands. Save it as a temporary data set called work.pulse.
2. Generate formats for the variables SEX, RAN, and ACTIVITY. Apply these formats to the appropriate variables in the pulse data set.
3. Generate formats for the variables SEX, RAN, and ACTIVITY. Apply these formats to the appropriate variables in the pulse data set.
4. Carry out an independent samples t-test to compare the mean of PULSE1 (resting pulse) for those who Ran vs. those who did not run (varuable = RAN). What do you conclude? Carry out this same test for PULSE2 (second pulse measurement). What do you conclude?
5. How does the mean of PULSE1 compare to the mean of PULSE2 for all participants (consider using a paired samples t-test)? What do you conclude about the means of these two variables? What happens if you carry out the same test separately for those who ran and for those who did not run? What do you conclude about each of these subgroups?
6. Compare the proportion of smokers among males and females. What percent of the males smoke? What percent of the females smoke? What do you conclude about this comparison?
7. Fit a linear regression model to predict PULSE2, based on PULSE1, and whether the person ran or not, and whether they smoked or not. (You will need to generate dummy variables for RAN, and SMOKES).
8. Develop a logistic regression model to predict whether the person has a high pulse value at time 1: High pulse1 is defined as having a Pulse1 value >=90. You can use any predictors in the data set that you wish for this question.