STA 402/502 Homework 1

Due: September 7th(Friday), before class

Please read the homework guidelines before working on the homework. This homework will cover the materials discussed during class on first week, and you will do some reading for specific procedures not discussed in class. Please submit a well-formatted printed copy of the homework.

You are to complete this assignment on your own. Remember to include an intro comment block on all programs written. Each problem should be attempted as its own program.

1. This question is based on exercise 2 of section 1.8 in the textbook. The text file sleep.txt (originally from `http://lib.stat.cmu.edu/datasets/sleep`) are from the Australian Data and Story Library. The file includes the body and brain weight, life span, gestation time,time sleeping and predation and danger indices for 62 species of mammals. Some variables are missing and labeled with -999.

   (a) Read in the text file in SAS using INFILE , making sure to reassign the missing code from -999 to . (Hint: use an IF statement in the DATA block). Then print the first three observations from this dataset.

   (b) Provide a table consisting of mean, standard deviation and the 5-number summary for the gestation time of the 62 animals. (Hint: use PROC MEANS statement, read the SAS help documentation on PROC MEANS to learn how to produce the 5-number summary and the other measures.)

2. This question is based on exercise 3 of section 1.8 in the textbook. Read the description of the data in your textbook. Alternative to the steps outlined in the text, the data can be found by searching "Aerobic Fitness Prediction" and click on the first website in google or from this website `https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect055.htm`.

(a) Read in the data into SAS using the DATALINES statement. (you may just copy from the website and do not modify the data in any way).

(b) Produce a table containing the follow summary statistics for heart rate while resting (Hint: using PROC UNIVARIATE): mean, median, mode, std deviation, variance, range and interquartile range.

(c) Fit a simple linear regression using oxygen intake rate (ml per kg body weight per minute) as the response variable and time to run 1.5 miles (minutes) as predictor variable. What is the estimated intercept and slope of the regression function? (Hint: refer to the above mentioned website, or search google to find a solution.)

3. The data file conference.txt contains information about 113 registered attendees in a regional SAS conference: first name, last name, attendee ID, business phone, home phone, mobile phone, OK to contact attendee at business (Yes/No), OK to contact at home (Yes/No), OK to contact at mobile (Yes/No), registration rate, will attend Wednesday mixer (Yes/No), will attend Thursday lunch (Yes/No), whether willing to volunteer at the conference(Yes/No), and if there's eating restrictions (Yes/No).

   (a) Examine the data file conference.txt and read it into SAS (Name the variables by yourself and make sure data is read in correctly. Note that all are character variables, except the variable "rate"). Print the 2, 3, and 4th observations from this dataset.

   (b) Count the number of people who will attend Thursday lunch and who will not attend Thursday lunch. (Hint: use PROC FREQ)

   (c) Using the following registration rates, create a new variable that groups attendees as "Regular", "Early", or "On-Site". Then count the number of attendees within each category.

Academic Regular	$350
Student Regular	$200
Regular	$450
Academic Early	$295
Student Early	$150
Early	$395
On-Site	$550