Assignment 2: Due on Thursday Nov 11st 11:00pm

Note: Your answer should be in a pdf file, which includes SAS code and output

Q1) (15 marks) Create a SAS dataset using *sleep.dat* dataset (attached on Canvas) and do the followings.

a) Convert the values of -999.0 to a missing value(.) for each variable (Hint: you may use *if* condition).
b) Create a new variable of ratio of brain weight to body weight.
c) Summarize the data displaying the following quantities for each variable in the dataset: Number of observations, number of missing observations, minimum, median, maximum, mean, and standard deviation.
d) Generate a PDF showing the first 10 observations with all variables of the dataset and summary values in part (c).

Note that the data in the file *sleep.dat* consists of the various measures regarding sleep and other factors for a variety of mammals. The variables are:

1. Species of animal
2. Body weight (kg)
3. Brain weight (g)
4. Hours/day of non-dreaming sleep
5. Hours/day of dreaming sleep
6. Total hours/day of sleep
7. Maximum life span in years
8. Gestation time in days
9. Predator Index, from 1 to 5 with 5 be most likely to be preyed upon
10. Sleep exposure index, from 1 to 5 with 5 being the most exposed
11. Overall danger index, from 1 to 5 with 5 being in the most danger

Source: http://lib.stat.cmu.edu/datasets/sleep
*originally presented in Allison, Truett and Cicchetti, Domenic V. (1976), "Sleep in Mammals: Ecological and Constitutional Correlates", **Science**, November 12, vol. 194, pp. 732-734.*

Q2) (15 marks) Note that the data in the file *earthquake.dat* is a record of earthquakes recorded by the USGS forthe period from the end of 1996 to the middle of 1998. The following variables are included:

1. The date of the earthquake in the form yy/mm/dd.
2. The time of the earthquake (Greenich Mean Time) in the form hh:mm:ss
3. The latitude, in degrees, of the center of the quake.
4. The longitude, in degrees, of the center of the quake.
5. The depth, in kilometers, of the earthquake activity.
6. The magnitude, reported as a number, followed by two letters which indicate the scale used to measure the quake.
7. The quality of the location information, ranging from A(best) to D(worst). This field may also contain a blank.
8. A comments field, truncated to 28 characters, generally containing a string (with embedded blanks) describing the location of the quake.

Source: http://geology.usgs.gov/quake.shtml

a) Use set and where statement to subset the *earthquake.dat* dataset and create a permanent SAS dataset which only contains earthquake information happened in Colorado (Comments = ”COLORADO”). Show your output dataset.
b) Load the permanent SAS dataset which you got from part a). Compute the week intervals between consecutive earthquakes happened in Colorado. Show your output dataset