Assignment 2: Due on Thursday 22nd by 3.00 pm (PDT)

Note: For each question, produce separate PDF files for SAS code and output. Those PDF files should be uploaded to the Crowdmark under each question.

Q1) You have a text file called *stocks.txt* containing stock symbol, a price, and the number of shares.

a) Using this raw data file, create a temporary SAS dataset. In addition, create a new variable (Value) equal to the stock price times the number of shares.

b) Write the appropriate statement to compute the average price and the average number of shares of your stock.

c) Produce a PDF document to present the dataset without any observation numbers.

Q2) 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

Missing values for all of the variables are represented by -999.0.

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.*

Q3) Create a SAS data set using the *earthquake.dat* dataset. Make sure to use an appropriate *informat* for the date and time of the earthquake. For each location specified in the comments field, calculate the time between successive earthquakes. Generate a PDF showing the first 10 observations with all variables of the dataset.

Note that the data in the file *earthquake.dat* is a record of earthquakes recorded by the USGS for the 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*

Q4) Read the data in file *President.txt* using FILENAME statement. The variables in order are FullName, City, State, ASP, DOB, Term, Votes, PercVote. ASP means “age starting presidency" and PercVote is the “percent of the vote". For this data set do the following:

a) Assign an appropriate format to DOB, votes, and PercVote.
b) Create a character variable (DOBC) that contains DOB in the form 01JAN1960.
c) Use the SUBSTR() function to create a new character variables to hold the day, month, and year from DOBC (Hint: Read the document on SUBSTR() available at https://documentation.sas.com or any other related SAS online document).
d) Create another character version of DOB and name it DOBC2, these dates should be in the form 31/01/1960.
e) Using the SCAN() function to create three numeric variables that hold the day, month, and year from DOBC2. (Hint: Read the document on SCAN() available at https://documentation.sas.com or any other related SAS online document).
f) What is the difference between SUBSTR() and SCAN(). Include a brief comment in your code.
g) Produce a PDF showing the first 10 observations with all variables and the descriptor portion of the data set with an appropriate title.

Q5) You have a text file called *geocaching.txt* with data values arranged as follows: (To learn about geocaching (treasure hunting with a hand-held GPS), go to www.geocaching.com.)

Variable	Description	Starting Column	Ending Column	Data Type
Name	Cache name	1	20	Char
LongDeg	Longitude degrees	21	22	Num
LongMin	Longitude minutes	23	28	Num
LatDeg	Latitude degrees	29	30	Num
LatMin	Latitude minutes	31	36	Num

a) Create a temporary SAS data set called *Cache* using this data file. Use column input to read the data values and produce a PDF to present dataset.
b) Repeat the problem using formatted input to read the data values instead of column input. Produce a simple report using PROC REPORT and produce a PDF document of that report. In your report, add a title called “Simple Report of Geocaching”. When printing the report, replace each variable name with its description. For example, ‘Name’ variable should be displayed as ‘Cache name’ and so on.