Survival Analysis with Base SAS Course

Assignment 1

**Due Date: 06/11/2017 Sunday 11:59 PM Eastern Time**

**No credits will be given to overdue assignments.**

**Please Read Carefully:**

* **This assignment covers Chapter 1 through Chapter 4.**
* **Datasets “Crayon.dat” and “Hotel.dat” are provided with the assignment. You are expected to write SAS code to carry out tasks requested in the questions.**
* **You are expected to submit one SAS code file. Please name submission files to include your own names.**
* **Filename format is <firstname>.<lastname>.assignment1.sas**
* **Do not share any information on how to solve the assignment on the Discussion Forums. However, you can seek clarifications for the questions.**
* **Grading:** 
  + **Q1 – 40 points**
  + **Q2 – 60 points**
* **Good luck! 😊**

1. Crayola crayons were introduced in 1903, and since then numerous standard colors have been released. Each crayon has a unique name, which corresponds to a hexadecimal code and RGB triplet. The raw data file Crayons.dat contains information on these standard crayon colors with variables corresponding to crayon number, color name, hexadecimal code, RGB triplet, pack size, year issued, and year retired.
2. *(3 points)* Open the raw data file Crayons.dat in a simple editor such as WordPad. In a comment in your program, state which variables must be read in as character and which variables should be read in as numeric.
3. *(10 points)* Read the raw data file into a permanent SAS data set.
4. *(7 points)* Create a variable “popularity” which shows the popularity of the colors depending on the pack size:
   * If the color is found in packs sized 16 or less, it is popular.
   * If size is 48 or less, it’s uncommon.
   * If the size is over 48, the popularity is rare.
5. *(5 points)* Create a table to show how many colors fall into which category.
6. *(5 points)* Create a new dataset named rare\_colors which includes only the colors with rare popularity.
7. *(10 points)* Print the first 5 observations of the “rare\_colors” dataset for the colors with a name comprised of more than one word **sorted by ascending issued year (the smallest issue year is at the top) and then by name**. Title the report.
8. The new management of a local hotel decided to update their recently acquired (and very outdated) property by installing wireless Internet service for their guests. They are also considering updating their billing system because the method used by the previous owner seems faulty. In order to conduct a billing analysis, they would like some calculations about the guests who stayed with them during the first part of February (this was the first month after the change of ownership). The raw data file Hotel.dat contains variables with information on room number, number of guests, check-in month, day, year, check-out month, day, year, use of wireless Internet service, number of days of Internet use, room type, and room rate.
   1. *(10 points)* Examine the raw data file Hotel.dat and read it into SAS.
   2. *(10 points)* Create date variables for the check-in and check-out dates, and format them to display as readable dates, use a format to show the weekdays also.
   3. *(10 points)* Create a variable that calculates the subtotal as the room rate times the number of days in the stay, plus a per person rate ($10 per day for each person beyond one guest), plus an Internet service fee ($9.95 for a one-time activation and $4.95 per day of use).
   4. *(5 points)* Create a variable that calculates the grand total as the subtotal plus sales tax at 7.75%. The result should be rounded to two decimal places and should be formatted as currency.
   5. *(5 points)* Sort the resulting data set with respect to **the grand totals descending** (the highest is at the top). Print the top 5 entries of the resulting data set.
   6. *(10 points - Use a procedure to complete this task*) Create a report for descriptive statistics of “number of days in the stay” and “number of guests” for each level of internet usage (YES/NO). Choose an appropriate title for your table.
   7. *(10 points - Use a procedure to complete this task*) Create a frequency table with titles which shows a cross-tabulation of the room type and check-in weekday. (Hint: You must find the weekday value (Sunday, Monday, etc.) of check-in date to complete this analysis.)