Survival Analysis with Base SAS Course

Assignment 2

**Due Date: 06/18/2017 Sunday 11:59 PM Eastern Time**

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

**Please Read Carefully:**

* **This assignment covers Chapter 6 and Chapter 7.**
* **Datasets 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>.assignment2.sas**
* **Do not share any information on how to solve the assignment on the Discussion Forums. However, you can seek clarifications for the questions.**
* **The scope of the questions may require that you do some online research to reach the solutions.**
* **Grading:** 
  + **Q1 – 25 points**
  + **Q2 – 30 points**
  + **Q3 – 25 points**
  + **Q4 – 20 points**
* **Good luck! 😊**

**Q1. Merge datasets**

A wildlife rescue center maintains a database consisting of two SAS data sets called FRIENDS and NEWINFO. The FRIENDS data set contains one observation for each supporter. Every time the center receives a donation, the staff add an observation to the NEWINFO data set. They also add an observation to NEWINFO every time there is a new volunteer or a change of address. To create an up-to-date data set of supporters, the staff need to combine the NEWINFO data set with the FRIENDS data set.

1. *(5 points)* Examine these SAS data sets including the variable labels and attributes. Add a comment to your program that notes the sort order of the observations in these data sets. Then state the number of variables that appear in both data sets, and the number of variables that appear in just the FRIENDS data set or just the NEWINFO data set.
2. *(10 points)* Add the information in the NEWINFO data set to the FRIENDS data set. Do not add any new variables to the FRIENDS data set. Be sure to make a backup copy of the FRIENDS data set before modifying it.
3. *(10 points)* The staff would like to see a report showing the total donations received from each friend. Compute the total donations by ID, and combine the totals with the FRIENDS data. Then produce a report listing the ID, first and last name, and total donations for each friend.

**Q2. Visualize data**

The United States Geological Survey provides data on earthquakes of historical interest. The SAS data set called EARTHQUAKES contains data about earthquakes with a magnitude greater than 2.5 in the United States and its territories. The variables are year, month, day, state, and magnitude.

1. *(5 points)* Examine this SAS data set including the variable labels and attributes. Create a scatter plot of year and magnitude for earthquakes that occurred in the year 2000 and beyond.
2. *(10 points)* Overlay a time series plot of the mean magnitude for each year on the same graph as part a). This line should appear in red on your graph.
3. *(5 points)* Include a legend for your graph that labels the time series plot as Mean. The legend should have no border and appear on the bottom right side of the plot.
4. *(5 points)* Earthquakes are classified by their magnitude. Overlay reference lines on your graph for light, moderate, strong, major, and great earthquakes defined at magnitudes of 4.0, 5.0, 6.0, 7.0, and 8.0, respectively. These lines should be labeled, dashed, and have 50% transparency.
5. *(5 points)* Make sure that all years appear on the X axis of your graph.

**Q3. Specific PROC Use**

Suppose that at a local university the study guidelines for the College of Science and Math are to study two to three hours per unit per week. The instructor of the class, Orientation to the Statistics Major, takes these guidelines very seriously. He asks students to record their study time each week, and at the end of the term he compares their average study time per week to their term GPA. The SAS data set called STUDY\_GPA contains student identification information, orientation course-section number, number of units enrolled, average time studied, and term GPA.

1. *(5 points)* Examine this SAS data set including the variable labels and attributes. Create box plots to compare the time studied between the two sections.
2. *(5 points)* Create a simple linear regression plot for time studied and GPA. Turn off the legend.
3. *(5 points)* Create a simple linear regression plot for time studied and GPA with a line for each section. Move the legend to the far-right side of the plot.
4. *(5 points)* Add 95% confidence limits for the mean predicted values to your plot from part c). Adjust the transparency so that bands for both sections are visible on the plot.
5. *(5 points)* Add a comment to your program about any potential relationships that you see between the variables included on these three plots.

**Q4. Specific PROC Use**

The local public transportation company would like to conduct a study of the impact of three possible plans on a certain popular bus route in town. Plan one consists of traffic lights timed in sync via a special remote transmitter on the bus that interacts with the traffic signal. Plan two consists of shutting down a regular traffic lane to become an express lane for buses only. Plan three is the current normal bus route with no changes. To assess the impact of a change, the company implemented plan one for a month, and then implemented plan two for a different month. All three plans were measured during summer months to ensure consistency in weather. The time (in minutes) for a bus to complete its route was measured for randomly selected weekdays and weekends, and these data are stored in the SAS data set called BUS.

a.   *(5 points)* Examine this SAS data set including the variable labels and attributes. Calculate the mean, median, and standard deviation travel time for each plan. Present these descriptive statistics to one decimal place.

b.   *(5 points)* Test to see if there is any difference between the mean times for any of the plans, and compare the resulting p-value to alpha = 0.05.

c.   *(5 points)* Add pairwise comparisons of group means to the analysis from part b).

d.   *(5 points)* In a comment in your program, discuss your findings. Which plan should the city adopt?