# ST 307 Project 1

For this activity you will create a SAS program and upload that program to wolfware. Each member of your group must submit their own code. Be sure that your SAS file adheres to the SAS file submission guidelines (available on wolfware). **You must put your group number in the header. You must assign your group members a score for participation on the assignment. Give them 100% for full participation and 0% for no participation.**

The data set for this activity is available on wolfware. You should download this and put it in a folder on the VCL that you will use as a SAS library.

In this project we are going to work on a blood data set from Ron Cody's book *Learning SAS by Example*: *A Programmer's Guide*. The sample consists of 1000 subjects. From each subject, value of the following variables are collected:

Subject: subject number
Gender: gender of the subject (Male or Female)
BloodType: blood type of the subject (A, B, AB, or O)
AgeGroup: for simplicity, subjects are divided into two age groups (Young or Old)
WBC: white blood corpuscle
RBC: red blood corpuscle
Chol: cholesterol level

These variables appear in the same order in the data set as they are here.

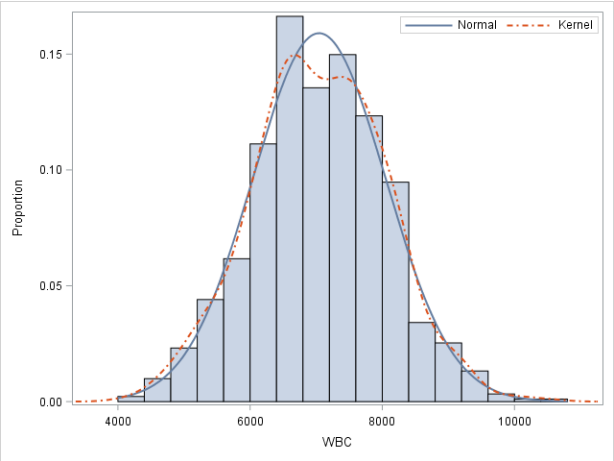
1. Create a library called Project1 where you are going to read the blood.txt from and save the output data sets to.

2. Read in the file blood.txt and save it as a SAS data set called Blood in the Project1 library.

3. You are interested in WBC count, since it is an important indicator of the health of the immune system. Write a SAS step and answer the following questions in the comment:
   a. What is the largest WBC count in the sample? What about the smallest? Which subjects have them?
   b. What is the standard deviation of the sample WBC count?

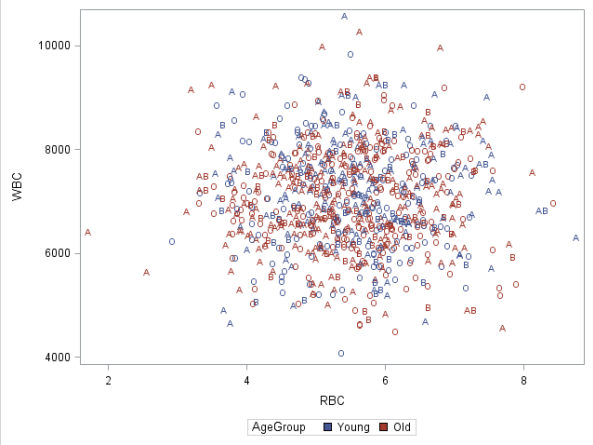
4. Create a histogram for the WBC variable using PROC SGPLOT. Add some statements and/or options to meet the following requirements:
   a. Instead of percentages, make the vertical axis show the values as proportions (0.0 to 1.0) of the total.
   b. Add two density curves to the histogram, a normal one and a kernel one.
   c. Change the line type of the kernel density curve to dash dot line.
   d. Move the legend of the density curves inside the plot on the top right corner.(**Hint**: you may find the KEYLEGEND statement useful)

   After all these settings, you see a plot like the one on the next page:




5. Create a scatter plot for WBC against RBC. (WBC on the y axis and RBC one the x axis).
   a. Instead of small circles as markers, make the markers show the blood type of the subject.
   b. Give different color to subjects from different age groups.

   After these settings, you would like to see plot like this:




6. You are interested in the distribution of blood type, so you want to:
   a. Create a one-way frequency table for the blood type and save this table as a permanent data set file called BloodTypeFreq in your Project1 folder. (**Hint**: make sure you figure out the difference between an OUTPUT statement under PROC FREQ and an OUT = option on the TABLES statement before you choose the right one to use.)
   b. Create a two-way frequency table for blood type and age group, but suppress the display of frequencies.
   c. Answer in the comment: what would we do to save the data set in part a as temporary?

7. Create a horizontal box plot for RBC for each blood type. Add some options to meet the following requirements:
   a. Use blood type as a category variable, so that a box plot is created for each distinct blood type.
   b. Make the marker for the mean a red filled diamond.
   c. Make the marker for the median a black line and a little thicker than default.
   d. Make the box outline and the whiskers of the same thickness as the median line.

   After these settings, you would like to see plot like this: