Biostat 510
Homework 1
Due Tuesday, January 20, 2009

1. **Download the Excel files** for the 10 groups from the first day of class. Save them in a folder that you will be able to access later. They are in the Biostat 510 section of my web page at http://www-personal.umich.edu/~kwelch/510/2009/homework_2009.html

   Check the data in the Excel files, **fix any problems**, and **resave** the fixed files.
   a) Open each file in Excel and make sure the **first row** of the Excel file contains the variable names.
   b) Make sure there are **no empty rows** between row 1 and the rows containing the data.
   c) Note the **Sheet name** on the tab at the bottom of the worksheet.
   d) Make sure that **GROUP is filled down** for all cases in each spreadsheet.
   e) Make sure that each file is saved as **.xls, and not .xlsx** .
   f) **Make sure all variable names are spelled the same way in all Excel files**, and that there are no blanks in the variable names.

   You should have the following variables in each Excel file:

Variable Name	Description	Type
GROUP	Group Number	Numeric
ID	ID Number within Group	Numeric
RAN	1= Ran, 2=Did not run	Numeric
AGEYR	Age in years	Numeric
AGEMO	Age in months	Numeric
SEX	M or F	Character
HR1	Heartrate at time 1	Numeric
HR2	Heartrate at time 2	Numeric

   (NB: Upper and lower case for the Variable Names do not matter, but they should all be spelled the same way!!).

   g) **Resave** each of the Excel files after you have made any corrections.
   h) **Close** all the Excel files before importing them into SAS.

2. Use the Import Wizard to import the Excel files into SAS.
   a) **Import** Group1.xls into SAS, and save it as a **temporary data set** called WORK.Group1.
   b) **Save the commands** that you used to import the data in a command file called homework1.sas.

c) **Open the homework1.sas commands** in the **SAS Enhanced Program Editor** window, by going to File…Open program… and browsing to where you saved the commands. The commands that were used to import the group1.xls file should look something like those shown below:

```
PROC IMPORT OUT= WORK.Group1
            DATAFILE= "group1.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;
```

Note: you do not have to type the commands as shown above if you use the Import Wizard in SAS. The import Wizard will automatically create the commands and save them in your command file. See the handout on my web page for instructions on how to use the Import Wizard to import an Excel file.

Note: If the import Wizard will not import an Excel file, save each data file as CSV (comma separated values) file from Excel and import them using the Import Wizard. Sample Commands for importing a CSV file are shown below:

```
PROC IMPORT OUT= WORK.Group1
     DATAFILE= "group1.csv"
     DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
```

Notice that the commands above do not include a sheet name, as when importing an Excel spreadsheet.

d) **Browse the Group1 data** set in the VIEWTABLE Window in SAS to be sure it has been imported correctly.
e) **Close the viewtable view of the data** set in SAS, being careful not to close all of SAS in the process, and return to the Enhanced Program Editor window.
f) In the SAS Enhanced Program Editor window, **copy and paste the SAS commands for group1**, and modify them to create the data sets for the remaining groups, they will be called WORK.group2, WORK.group3, etc.
    i. Make sure the data set name is specified correctly for each group.
    ii. Make sure the sheet name is specified correctly for each group.
g) **Submit the PROC IMPORT commands** for the remaining data sets.
h) **Check the SAS log** to be sure there are no errors.

✓ Include the portion of your SAS log showing the successful importation of the Group 2 data in your homework write-up. You don't need to include the log for each group's data.

3. **Combine the data sets** for all groups by stacking them into one temporary data set, using a **SET** statement in SAS. Your SAS commands will look something like those shown below (e.g. for 3 groups):

```
data work.Allgroups;
   set group1 group2 group3;
run;
```

Note: the libname, WORK, is optional for any temporary data set name.

a) Browse your Allgroups data set.
b) If there is a problem in creating your data set, go back and fix it, and recreate the data set by highlighting and resubmitting your commands.

✓ Include the portion of your SAS log showing the successful creation of the Allgroups data set.

4. Get a Proc Contents for the allgroups data set, with the variables in ***creation order***, by using the **varnum** option. Include the output from Proc Contents in your homework write-up.
5. Get descriptive statistics for all numeric variables in your combined set using Proc Means. Include this output in your homework write-up.

   Use a **class statement** to get descriptive statistics for each GROUP. Include this output in your homework write-up.

6. Get frequencies for the categorical variables GROUP, SEX, and RAN, using Proc Freq. Get a two-way cross-tabulation with GROUP as the row variable and RAN as the column variable. Include the output from this question in your homework write-up.
7. Get a **histogram** for HR2, paneled by RAN. Include this paneled graph in your write-up.
8. Get a **boxplot** of HR1 for those who ran and didn't run. Get a **boxplot** of HR2 for those who ran and didn't run. Include both boxplots in your write-up.
9. Save your data set as a permanent SAS data set, using commands similar to those shown below:

```
libname b510 "c:\510\2009";
data b510.allgroups;
   set work.allgroups;
run;
```

10. Save your SAS commands as homework1.sas. Re-run all of the commands in this command file and make sure there are no errors in your log.

Answer the following questions, based on the results of your descriptive analysis.