

# **Biostat 510: Statistical Computing Packages**

## **SAS Homework 1**

### **Due Thursday, Sept 12, 2002**

#### Topics:

- Importing an Excel file
  - Creating a permanent SAS data set
  - Setting up missing values
  - Recodes
  - Transformations
  - Distributions of continuous variables
  - Histograms
  - Frequencies and Cross-tabs
  - Independent Samples t-test
  - Paired Samples t-test
1. Make an Excel file containing the data collected on the first day of class. Save it as an Excel 4.0 worksheet on your h: drive. Give it the name HRTRATE.XLS. Be sure to use consistent values for such variables as Gender (F,M). Enter the appropriate missing values codes (such as 99 or 999) for missing data.
  2. Import the Excel file into SAS, using the Import Wizard, and save it as a temporary SAS data set called HRTRATE in the Work library. Save your commands from importing and then reopen the file and use it as the basis for your commands for this assignment.
  3. Check the log to be sure your data was imported correctly. If not,
    - a. Be sure the Excel file is closed before you try to open it in SAS.
    - b. Be sure the Excel file was saved as an Excel 4.0 worksheet.
  4. Open the command file that you saved from importing the data set and modify it for the rest of the commands for the assignment.
  5. Create a new permanent data set called B510.HRTRATE2. Set up missing values in this data set, and create a new variable called WTKG that is weight in kilograms, appropriately calculated for everyone in the data set. To make the permanent SAS data set, do the following:
    - a. Create a folder called b510 on your h: drive if you are working at a campus computing site, or as a folder in c:\temp if you are working at another computer.
    - b. Submit a libname statement to define this folder as a library to SAS, using the V8 engine specification.
    - c. Submit a data step to save HRTRATE2 as a permanent SAS data set.

```

libname b510 v8 "c:\temp\b510";
data b510.hrtrate2;
    set hrtrate;
    /*set up missing values and do recodes here*/
    if units = "LB" then wtkg=weight/2.2046;
    if units = "KG" then wtkg=weight;

run;

```

6. Get descriptive statistics on your data set using Proc Means. Note the number of observations, min, max and mean for each variable. Do they make sense? Print the first 10 observations to check the data entry. Get a Proc Contents.

```

proc means data=b510.hrtrate2;
run;
proc print data=b510.hrtrate2(obs=10);
run;
proc contents data=b510.hrtrate2;
run;

```

7. Open the file in INSIGHT and check the distribution of Pulse2 using the Distribution(Y) menu. Print out the histogram and descriptive statistics for Pulse2. Check any outliers (missing values, etc).
8. Create BMI (body mass index) by calculating Weight in kilograms/Height in meters-squared. To do this you will first need to transform Height into height in meters (divide inches by 39.37).
9. Create a new variable called AGEGRP that divides age into 3 categories. You can choose how to set up the categories. Be sure that only those with a value for AGEYR get a value for your new variable AGEGRP. You can check your new variable using Proc Freq. Note: You should put the SAS statements necessary to do this recode into your data step and rerun it to create the new variable, agegrp.

```

proc means data=b510.hrtrate2;
    class agegrp;
    var age;
run;

```

10. Get frequency distributions for Sex, Activity, Age, Agegrp, and Ran, using Proc Freq. If you notice any weird values, go back fix them in your SAS code, and rerun it. Open HRTRATE2 in INSIGHT and check the distribution of BMI. What does this distribution look like?
11. Get descriptive statistics for all variables separately for those who ran and those who did not run, using the Class statement with Proc Means.
  - a. What are the mean and standard deviation of Pulse1 for those who ran and those who did not run?
  - b. Do an independent samples t-test to compare the mean of Pulse1 for those who did not run compared to those who ran.
  - c. Write out the null and alternative hypotheses for this test. Are the variances for these two groups equal? What is the value of the t-statistic?

What are the degrees of freedom? What is the p-value? What do you conclude?

- d. What are the mean and standard deviation of Pulse2 for those who ran and those who did not run?
- e. Do another independent samples t-test for Pulse2, comparing the mean for those who ran and did not run. Again, write out the null and alternative hypotheses for this test, the t-statistic, the degrees of freedom and the p-value. What do you conclude about the mean of Pulse2?

```
proc ttest data=b510.hrtrate2;  
  class ran;  
  var pulse1 pulse2; run;
```

12. Do a paired t-test to compare the mean of pulse1 vs. pulse2 for all people in the class.
  - a. What are the null and alternative hypotheses for this test?
  - b. What is the value of the t-statistic?
  - c. Is there a test of equal variances here? Why not?
  - d. What is the p-value? What do you conclude?
13. Do this same paired t-test separately for those who ran and those who did not run. Use a BY statement in SAS to do this. Be sure to sort the data set first.
  - a. What are the null and alternative hypotheses now? Remember, you should now have 2 tests, one for those who ran and one for those who did not run.
  - b. What is the value of the t-statistic for each group?
  - c. The degrees of freedom?
  - d. The p-value of the test?
  - e. What do you conclude about the mean of pulse1 vs. pulse2 for those who did not run? For those who ran?

```
proc ttest data=b510.hrtrate2;  
  paired pulse2*pulse1;  
run;  
proc sort data=b510.hrtrate2;  
  by ran;  
run;  
proc ttest data=b510.hrtrate2;  
  by ran;  
  paired pulse2*pulse1;  
run;
```

14. Now, create a new variable in your data step, called Pulsedif, that is calculated as Pulse2-Pulse1. Include it in your data step where you created the new data set hrtrate2. Rerun the data step so this new variable gets created.
  - a. Do an independent samples t-test to compare pulsedif for those who ran vs. those who did not run.
  - b. What kind of t-test will this be?
  - c. What are the null and alternative hypotheses for this test?
  - d. Are the variances for the two groups the same?