STSCI 5010 HW2

(Due: 10/9/2018 at 11:59PM)

General instructions:

- Download the Excel file (FBPandHIV.xlsx) from the course Blackboard website. Note that within the Excel file there is a sheet for the Data and a different sheet for the Codebook. Open the file in Excel so that you become familiar with the variables and their values.
- Assign a new SAS library named hw2, where you keep all your HW2 files.
- Use SAS programming (not point-and-click) approach for the entire homework.
- Only produce HTML output for this homework.
- Use SAS comments to briefly document your code. If comments are missing, up to 5 points are deducted.
- Use title statements to mark your question numbers, e.g., "Question 1. Use a footnote to include answers to each question, if applicable. If titles and footnotes are missing, up to 5 points are subtracted.
- Submit four files (compressed with 7-zip or a similar tool, named Last-name_first-name_HW2.7z) *to the course website*:
    - Your SAS code (in one single file), named last-name_first-name_HW2_SAS_code.sas.
    - Your SAS HTML output, named last-name_first-name_HW2_HTML_output.mht.
    - Your SAS log file, named last-name_first-name_HW2_log.log.
    - The Excel file created from Question 4.

  If the HTML file or the log file is missing, 5 points are deducted.

Question 1. Read, convert and define formats (40 points)

a. While reading in the data from FBPandHIV.xlsx into a SAS dataset called FBP_HIV (stored in a library called HW2),
   - Calculate the participant's BMI (Body Mass Index) at baseline (Base_BMI), follow-up (Post_BMI) and the change in BMI (Delta_BMI), where
     $$BMI = \text{Weight}(kg)/height \text{ (meter)}^2$$
     $$\text{Delta\_BMI} = \text{Post\_BMI} - \text{Base\_BMI}$$
b. Create three user-defined formats (you need to give a name to each format) based on the following rules:

   - User-defined format for the Karnofsky scores (PreKarn and PostKarn values in the dataset). You are required to take care of

low and high boundaries and possible missing values if any (in the case of missing value, label it "Unknown"):
- When the Karnofsky score <= 24, label it "Sick 24 or less"
- When the Karnofsky score is in [25, 74], label it "Disabled 25-74"
- When the Karnofsky score >= 75, label it "Healthy 75 or greater"

- User-defined format for "Yes" and "No:"
  - If the value is 1, label it "Yes."
  - If the value is 0, label it "No."

- User-defined format for male and female:
  - If the value is 1, label it "Female."
  - If the value is 0, label it "Male."

- User-defined format for BMI. You are required to take care of the low and high boundaries and possible missing values if any (in the case of missing value, label it "Unknown"):
  - If a BMI value < 18.5, label it "Underweight."
  - If a BMI value is in [18.5, 25.0), label it "Normal Weight."
  - If a BMI value is in [25.0, 30.0), label it "Overweight."
  - If a BMI value >= 30.0, label it "Obese."

c. Display the first 10 observations of the FBP_HIV dataset. Apply the user-defined formats to all applicable variables so that their values are displayed as labelled.

Question 2. Crosstabulation (20 points)

Based on the FBP_HIV dataset, produce crosstabulations of the following variables, being sure to specify appropriate output row and column labels for each category of each variable with the user-defined formats.
   a. FBP and Base_BMI.
   b. FBP and Post_BMI.
   c. ARV and Base_BMI.
   d. ARV and Base_BMI, stratified by gender.

Question 3. PROC Summary and PROC Means (30 points)
   a. The average baseline CD4 count and follow-up CD4 count for each level of FBP using PROC SUMMARY. You are required to produce a report

with PROC SUMMARY. Keep one decimal point (same below).
b. The median and Qrange of PreKarn, PostKarn, PreCD4, and PostCD4.
c. Use PROC MEANS to test if the change in BMI from baseline to follow-up is statistically significant from zero. What do you conclude from this result? (Hint: Consider using the t and probt options)

Question 4. Creating an Excel from a SAS dataset (10 points)

Using the nutrition dataset from Lab1 and one data step, create an Excel file named **Health** and save it to the directory to which the hw2 libref points with the following requirements:

- Health.xlsx has one worksheet named **Nutrition**.
- While you are creating the Excel file, reverse the gender in the worksheet, i.e. if, in the original dataset nutrition.sas7bdat,the gender variable for a person is M, change it to F in health.xlsx and vice versa.
- Submit the Excel file with other files.