

SAS Programming (BIOL-40190) – Programming Assignment 2 (35 points)

This assignment is due by the date shown on the Assignments page in Blackboard.

The document **Programming Assignment Submission Instructions.pdf** provides the details for submitting your completed assignment.

1-6 – Create the data set STUDY using your corrected code for Assignment 1.

7 – DEMOG1062 is a permanent SAS data set located on the server in the directory **/courses/dc4508e5ba27fe300/c_629/saslib**

Create a new data set called PAT_INFO by merging STUDY and DEMOG1062 by their two common variables. Also add items in 8-12 to PAT_INFO.

Note: Your code should create a single data set called PAT_INFO, which contains the merge code and items 8-12. PAT_INFO should contain 15 observations and 21 variables.

8 – Using 1 programming statement, create a variable called pt_id by concatenating Site and Pt and adding a hyphen between the two variables. An example value of pt_id should look like: Z-99. Label the variable 'Site-Patient'. If either Site or Pt are missing then pt_id should be missing.

9 – Using 2 functions (one function will be embedded within the other) to create a variable dose_qtr by concatenating the letter 'Q' to the number which corresponds to the quarter of the year in which the dose date falls. Values of dose_qtr should look like Q1, Q2, etc. If the dose date is missing then dose_qtr should be missing. Use 1 programming statement for this item.

10 – Use a function to create a variable mean_result which is the mean of result1, result2, and result3. The mean should be calculated using all non-missing values of the three variables. Format mean_result to 2 decimal places.

11 – Create a variable BMI which is calculated as: $\text{Weight} \div (\text{Height})^2 \times 703$. If either Weight or Height are missing then BMI should be missing. Format BMI to 1 decimal place.

12 – Create a variable est_end which is the Estimated Termination Date for the patient. Use an assignment statement. Do not use a function. If Protocol Amendment is missing then est_end should be missing. If Protocol Amendment is A then est_end is 120 days after Dose Date. If Protocol Amendment is B then est_end is 90 days after Dose Date. Apply a format so that the est_end is displayed as mm/dd/yyyy. Label the variable 'Estimated Termination Date'.

13 – Using the data set PAT_INFO, generate the following output using PROC PRINT:

Listing of Baseline Patient Information for Patients Having Weight > 250									
Study Site	Site Name	Patient	Age	Sex	Race	Height	Weight	Date of First Dose	Dose Lot Number
J	Aurora Health Associates	08	63	Male	Asian	74	280	02/08/98	P0526
		10	79	Female	Black	68	258	01/03/98	P1122
		12	77	Male	Caucasian	76	358	01/17/98	P0526
R	Sherwin Heights Healthcare	08	80	Male	Caucasian	72	386	01/24/98	P0526
		09	77	Male	Black	74	387	12/27/97	S0576

14 – Use the data set PAT_INFO and one PROC MEANS to do the following:
Create output stratified by Sex for the variables Result1, Result2, Result3, Height, and Weight.
The display should show the number of non-missing values, mean, standard error, minimum value, maximum value and be formatted to one decimal point.
Also create an output data set that contains the median value of Weight stratified by Sex.
The variable that contains the median value of weight should be called med_wt.
Your output data set should contain two observations and two variables, Sex and med_wt.

15 – Combine the data sets PAT_INFO and the output data set from item 14 by the variable Sex and create a new variable called wt_cat as follows:

If the patient's weight is less than or equal to the median weight for all patients of that sex, then wt_cat=1.
If the patient's weight is more than the median weight for all patients of that sex, then wt_cat=2.
Label this variable 'Median Weight Category'.
Create and apply a descriptive format to wt_cat:
For wt_cat=1, the descriptor is '<= Median Weight'
For wt_cat=2, the descriptor is '> Median Weight'
Hint: Your new data set should contain 15 observations.

16 – Using your data set from Item 15 and one PROC FREQ to do the following:
Show the frequency distributions of (1) Dose Lot Numbers and (2) Median Weight Category. Exclude missing values from the frequency distributions. (3) Generate a two-way table for Race by Weight. Include missing values in the frequency distribution.

Use formats to group Race and Weight variables as follows:

If Race is Caucasian then display the race as 'White'.

If Race is anything else (including missing) then display the race as 'Other'.

Note: Give this new race format a different name than the race format created in the first part of the assignment.

Group Weight into the following 4 categories: < 200, 200 to < 300, >= 300, Missing

17 – Using your data set from Item 15 and one PROC UNIVARIATE to do the following:
Generate summary statistics for Height stratified by Median Weight Category.
Identify the extreme values using the Site-Patient identifier variable.

Before submitting your assignment, please ensure that your code and files meet the following criteria:

- *** Your code conforms to the item specifications and coding methodology.
- *** Your code is in sequential order and contains comments clearly identifying by item number the code for each item.
- *** Your code properly accounts for missing values.
- *** Missing values in your data set are left as missing and not artificially populated with values such as the word "Missing".
- *** Your code conforms to good programming practices by using white space, proper indentation, comments, etc.
- *** Your completed program may also need other statements or procedures that are implied but not specified (i.e. filename or options statements or PROC FORMAT).
- *** Your log file does not contain any ERROR or WARNING messages.
- *** Your files are flat files with the proper file extensions (.sas and .log).