BIOS 511 Lab 9
Advanced Dataset Programming & Creating Graphs in SAS - II

Please read the following instructions carefully before beginning this lab.

Instructions:

- **<u>The two tasks should be completed in separate SAS programs</u>** named lab-09-PID-Task1.sas and lab-09-PID-Task2.sas, respectively, where PID is your student PID number. Please make sure to include an appropriate headers in both SAS programs.

    - The first SAS program should produce a permanent dataset named ADLB and store that dataset in some location on your computer (we will not use this dataset after lab-09 and so I do not recommend placing the dataset with other ECHO datasets).

    - The second SAS program will use the ADLB dataset to create three graphs. The graphs should be written to a single PDF file named lab-09-PID-output.PDF.

- You will upload the two SAS programs, two SAS logs, your ADLB.SAS7BDAT dataset, and PDF output file to document completion of the lab.

- The submitted logs should reflect clean runs of the complete programs (i.e., they should not contain log messages from when the programs were being developed).

Logs that contain ERRORs, WARNINGs, etc. will result in a point deduction of at least 10 points.

BIOS 511 Lab 9
Advanced Dataset Programming & Creating Graphs in SAS - II

Task 1: Analysis Dataset Programming

- For this task you are given the LB (Laboratory Parameters Dataset) from the ECHO Trial. This dataset is to be used as the input dataset for this task.
- Using this dataset along with the DM dataset, you will create an analysis laboratory dataset (named ADLB) that will then be used to make some graphs for the lab.
- The ADLB dataset should have the same number of observations as the LB dataset.
- The LB dataset contains all variables included in Table 1 ***except*** for those variables that are found in the rows highlighted in yellow.
- The variables found in the rows highlighted in yellow must be programmed in order to complete this task. Thus, to complete this task you must create the following variables: LBSEQ, LBNRIND, LBBLFL, BASE, BASECAT, CHANGE, and PCT_CHANGE and merge on the variables required from DM (e.g., AGE).
- Table 2 at the end of the lab provides an example for how the variables LBBLFL, LBNRIND, BASE, BASECAT, CHANGE, and PCT_CHANGE should be populated if ADLB is correctly created. Only select variables are shown in the diagram.
- Instructions for the content of each ***new*** variable are provided in Table 1 along with the required variable attributes (e.g., type, length, and label).
    - Your final ADLB dataset should match attributes with Table 1 and should not contain any variables other than those in Table 1.
    - An ideal solution will order the columns in the dataset as shown in Table 1.
- Before starting programing, spend 20 minutes brainstorming how you should go about programming the ADLB dataset. Use the example from Table 2 to help you think through creation on LBNRIND, LBBLFL, BASE, BASECAT, CHANGE, and PCT_CHANGE. Your program will require multiple DATA steps.
- Here are some additional points to consider when programming:
    - In general, you should not assume that every subject's first dose date was the date of the Week 0 laboratory specimen collection. For example, see subject ECHO-031-008. This has implications on which observations should have LBBLFL='Y'.
    - LBSEQ can be programmed by sorting the data and incrementing its value restarting at 1 for each new subject once the data have been properly sorted. To keep track of the LBSEQ variable values from one observation to the next, it will need to be included in a RETAIN statement.
    - Some variables require other variables to be created at an earlier step (either in an earlier DATA step or at an earlier point in a given DATA step). For example, one cannot create LBNRIND unless the SEX variable has been added to the dataset already. Think about the order in which variables should be created.
    - Creating LBBLFL can be challenging. *One strategy for doing this* is to identify the set of *observations* that have non-missing test results AND where the specimen was collected on or before the date of first dose. Among this set of observations (for each subject and test) there may be more than one test result (e.g. Screening and Week 0). The baseline observation should be the observation associated with the specimen collected *latest in time* among that set.
    - One cannot create BASE/BASECAT unless it is known which observation should be marked as LBBLFL="Y" because the value of BASE and BASECAT are equal to the value of LBSTRESN and LBNRIND from that observation.
    - The best way to perform Quality Control (QC) of your ADLB dataset is to manually check the calculations for a few subjects and tests and to ensure you match Table 2

BIOS 511 Lab 9
Advanced Dataset Programming & Creating Graphs in SAS - II

Table 1: Contents of the ADLB Dataset (Number of Observations = 7224)

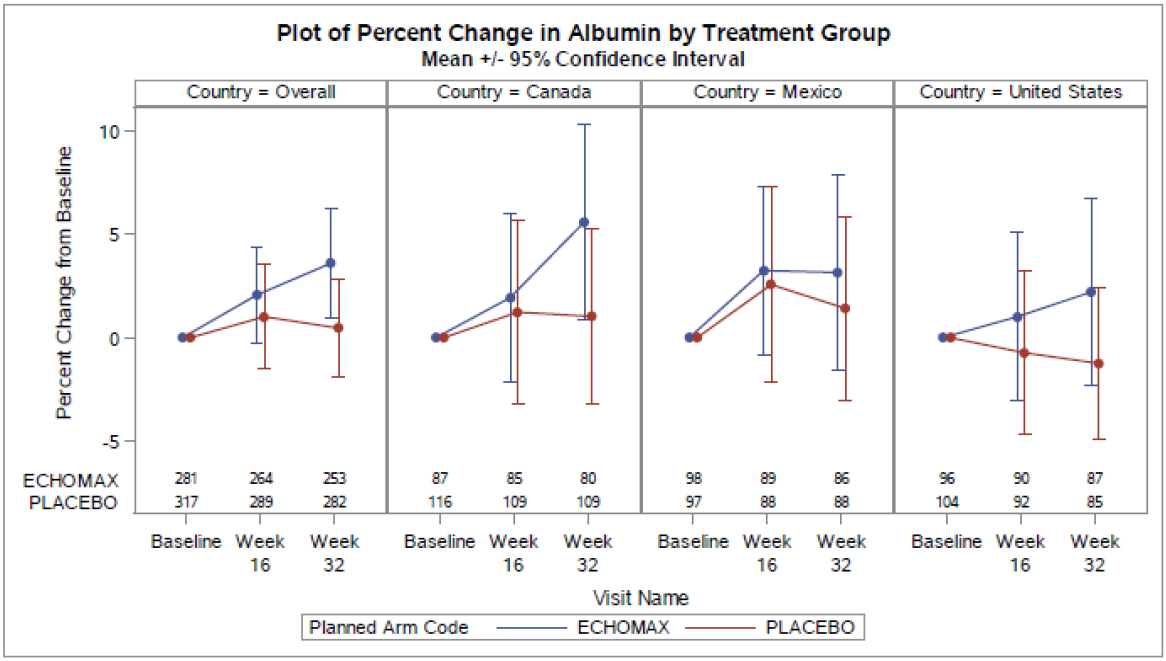
#	Variable	Type	Len	Label	Description of variable / derivation
1	STUDYID	Char	10	Study Identifier	Values are always equal to "ECHO".
2	USUBJID	Char	30	Unique Subject Identifier	Values are of the form "ECHO-XXX-YYY" where XXX and YYY are integers.
3	AGE	Num	8	Age	Variable comes from ECHO.DM
4	SEX	Char	10	Sex	Variable comes from ECHO.DM
5	RACE	Char	50	Race	Variable comes from ECHO.DM
6	COUNTRY	Char	20	Country	Variable comes from ECHO.DM
7	ARMCD	Char	20	Planned Arm Code	Variable comes from ECHO.DM
8	ARM	Char	50	Description of Planned Arm	Variable comes from ECHO.DM
9	LBSEQ	Num	8	Sequence Number	A unique identifier for observations within a subject. The values of LBSEQ should start at 1 for each subject and should increment by 1 for each successive observation for the subject. Prior to creating this variable, the data should be ordered by USUBJID, LBTESTCD, LBTEST, VISITNUM, VISIT, and LBDTC.
10	LBTESTCD	Char	8	Lab Test or Examination Short Name	Variable comes from ECHO.LB. Values are "ALB", "HCT", and "CA".
11	LBTEST	Char	40	Lab Test or Examination Name	Variable comes from ECHO.LB. Values are "Albumin", "Hematocrit", and "Calcium".
12	LBCAT	Char	50	Category for Lab Test	Variable comes from ECHO.LB. Values are "CHEMISTRY" or "HEMATOLOGY".
13	LBSTRESN	Num	8	Numeric Result/Finding in Standard Units	Variable comes from ECHO.LB
14	LBSTRESU	Char	20	Standard Units	Variable comes from ECHO.LB
15	LBNRIND	Char	5	Reference Range Indicator	Values should be derived using the LBSTRESN variable. Values should be "L" for low, "N" for normal, or "H" for high. If the variable LBSTRESN is missing, then this variable will also be missing. Hematocrit: Females – Normal Range [0.349,0.445] (inclusive) Males – Normal Range [0.388,0.500] (inclusive) Calcium – Normal Range [2.1, 2.7] (inclusive) Albumin – Normal Range [35, 55] (inclusive) Note that a value equal to the lower or upper limit is normal. For example, if a female subject has a non-missing hematocrit test value below 0.349 at a visit then the value of LBNRIND should be "L" for that visit.
16	LBSTAT	Char	10	Completion Status	Variable comes from ECHO.LB Equal to "NOT DONE" if the lab test was not done or missing otherwise

17	LBREASND	Char	50	Reason Test Not Done	Variable comes from ECHO.LB Equal to the reason the lab test was not performed.
18	LBBLFL	Char	1	Baseline Flag	For each subject and lab test, set to "Y" for the observation associated with the latest non-missing test result among those test results based on specimens collected on or before the date of first dose. Otherwise, LBBLFL will be missing. This variable identifies the baseline lab test value for each subject and lab test. Note that the "Week 0" visit is intended to be the baseline visit but not all subjects will have all lab tests performed at each visit.
19	BASE	Num	8	Baseline Lab Test Value	For each subject and test, set to the value of LBSTRESN from the observation where LBBLFL = 'Y' (if one exists). The value of BASE should be populated on all observations for the associated lab test. Thus, if a subject has a value of 35 for the Albumin lab test for the baseline measurement, the value of BASE should be 35 for all Albumin observations for the subject.
20	BASECAT	Char	1	Baseline Reference Range Indicator	For each subject and test, set to the value of LBNRIND from the observation where LBBLFL = 'Y' (if one exists). The value of BASECAT should be populated on all observations for the associated lab test.
21	CHANGE	Num	8	Change from Baseline	Equal to LBSTRESN-BASE.
22	PCT_CHANGE	Num	8	Percent Change from Baseline	Equal to (LBSTRESN-BASE)/BASE*100
23	VISITNUM	Num	8	Visit Number	Variable comes from ECHO.LB A numeric version of VISIT: Screening → -1 Week 0 → 1 Week 16 → 3 Week 32 → 5
24	VISIT	Char	50	Visit Name	Variable comes from ECHO.LB Values are "Screening", "Week 0", "Week 16", and "Week 32".
25	LBDTC	Char	20	Date/Time of Specimen Collection	Variable comes from ECHO.LB Values have format "YYYY-MM-DDTHH:MM".

BIOS 511 Lab 9
Advanced Dataset Programming & Creating Graphs in SAS - II

Task 2: Creating a Graph from Summarized Data

- For this task, you will create three paneled graphs (one for each lab test). The paneled graphs should resemble the graph below.



- Each graph should be 4.5 inches in height and 8 inches in width. This should be set using a single ODS GRAPHICS statement.

- Each graph should have 4 panels. The first panel should present the data for all countries and the 2nd through 4th panels should present the data for specific countries.

    - In order to obtain a dataset that can be analyzed and then plotted to have such an effect one can use OUTPUT statements in the DATA step as follows.

```
data INPUT;
 set lab09.ADLB;
  output;
  country = 'A';
  output;
run;
```

    This code will create a dataset named WORK.INPUT with two times the observations as the ADLB dataset with each observation in the input dataset being duplicated apart from the fact that the COUNTRY variable will have COUNTRY="A" (or whatever value you assign it) for the second observation.

- The scatter plot points shown in the graphs should be equal to the mean value of the treatment group (overall or for a given country). The high/low bars should correspond the upper and lower 95% confidence limits for the mean and can be requested from PROC MEANS by using the LCLM and UCLM options on the OUTPUT statement.

- For each graph and in each panel, the sample size for each treatment group is shown at the bottom of the image. This is achieved by using the COLAXISTABLE statement (equivalent XAXISTABLE statement exists in SGPLOT). We have not discussed this statement in class and so you will need to consult the SAS documentation. Hint: You will need to use the CLASS option on this statement but should not have to use any other options.
- To create the three different graphs, you will essentially need to copy+paste your code from the first image and modify it for the second and third. We will learn a more efficient way soon using the SAS MACRO language.
- Feel free to customize the graphs further using the STYLEATTRS statement that we learned about when learning about PROC SGPLOT.

BIOS 511 Lab 9
Advanced Dataset Programming & Creating Graphs in SAS - II

Table 2: Example of correct ADLB derivations using one subject as an example (only select variables shown from ADLB dataset)

USUBJID	LBTESTCD	LBCAT	LBSTRESN	LBSTRESU	LBSTAT	VISIT	LBDTC	LBBLFL	LBNRIND	BASE	BASECAT	CHANGE	PCT_CHANGE
ECHO-011-003	ALB	CHEMISTRY	46.000	g/L		Screening	2016-04-18T08:00		N	51.000	H	-5.000	-9.804
ECHO-011-003	ALB	CHEMISTRY	51.000	g/L		Week 0	2016-05-04T08:00	Y	H	51.000	H	0.000	0.000
ECHO-011-003	ALB	CHEMISTRY	40.000	g/L		Week 16	2016-08-24T08:00		N	51.000	H	-11.000	-21.569
ECHO-011-003	ALB	CHEMISTRY	47.000	g/L		Week 32	2016-12-14T08:00		N	51.000	H	-4.000	-7.843
ECHO-011-003	CA	CHEMISTRY	2.350	mmol/L		Screening	2016-04-18T08:00	Y	N	2.350	N	0.000	0.000
ECHO-011-003	CA	CHEMISTRY	.		NOT DONE	Week 0	2016-05-04T08:00			2.350	N		
ECHO-011-003	CA	CHEMISTRY	2.050	mmol/L		Week 16	2016-08-24T08:00		L	2.350	N	-0.300	-12.766
ECHO-011-003	CA	CHEMISTRY	2.300	mmol/L		Week 32	2016-12-14T08:00		N	2.350	N	-0.050	-2.128
ECHO-011-003	HCT	HEMATOLOGY	0.347	fraction of 1		Screening	2016-04-18T08:00		L	0.469	N	-0.122	-26.013
ECHO-011-003	HCT	HEMATOLOGY	0.469	fraction of 1		Week 0	2016-05-04T08:00	Y	N	0.469	N	0.000	0.000
ECHO-011-003	HCT	HEMATOLOGY	0.353	fraction of 1		Week 16	2016-08-24T08:00		L	0.469	N	-0.116	-24.733
ECHO-011-003	HCT	HEMATOLOGY	0.414	fraction of 1		Week 32	2016-12-14T08:00		N	0.469	N	-0.055	-11.727

Note: This subject was male and had first dose date equal to 2018-05-04.