BIOS 511 Lab 8
Creating Graphs in SAS - I

Please read the following instructions carefully before beginning this lab.

Instructions:

- All tasks should be completed in a single SAS program named lab-08-PID.sas where PID is your student PID number. Please make sure to include an appropriate header in your SAS program.

- All the output that your SAS program produces in this lab should be delivered to a *single PDF file* named lab-08-PID-output.PDF.

- In your code, before starting a new task/part, include a block comment with the task/part number

```
/*************************************************************
                        SAS Code for Task # X / Step # Y
 *************************************************************/
```

- Recommendation: While initially writing the SAS program, do not concern yourself with creating a permanent output file. Simply view results in the results window to verify that your program has created the desired output. Once your program is essentially complete, add the appropriate ODS statements to create the permanent PDF file based on the requirements above.

- You will upload the SAS program, SAS log, and PDF output file to document completion of the lab.

- The submitted log should reflect a clean run of the complete program (i.e., it should not contain log messages from when the program was being developed).

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

BIOS 511 Lab 8
Creating Graphs in SAS - I

Task 1: For this task, you will practice making a vertical bar chart.

Step 1: Type AND run the following code to produce a vertical bar chart that counts the number of subjects in both treatment groups in the ECHO Trial:

```
title "Task 1 - Step 1: Number of Subjects by Treatment Group";
proc sgplot data = echo.DM;
 vbar armcd;
run;
```

- The ECHO.DM dataset has one observation per subject and the VBAR statement summarizes (counts) the number of subjects having each value of ARMCD to determine the height of the bars.

- The frequencies are plotted on the y-axis when one uses the VBAR statement. The frequencies are plotted on the x-axis when one uses the HBAR statement.

- Feel free to substitute the HBAR statement if you prefer.

Step 2: Type AND run the following code that modifies the previous SGPLOT step to change the bar fill attributes by using the FILLATTRS option and setting the fill color to light red (you will add another PROC step to your submitted program and do so for each step in each task).

```
title "Task 1 - Step 2: Number of Subjects by Treatment Group";
proc sgplot data = echo.DM;
 vbar armcd / fillattrs=(color=lightRed);
run;
```

- Note that colors in SAS can be referenced in a very simple way: red, lightRed, veryLightRed, darkRed, and veryDarkRed.

- Users can use the qualifiers VERYLIGHT, LIGHT, DARK, and VERYDARK to modify any color and the text itself is NOT case sensitive.

- The FILLATTRS options is relevant to bar charts but similar options exist for virtually every type of graph.

    - When making a scatter plot with the SCATTER, REG, or LOESS statements, one can use the MARKERATTRS option to control the symbol properties for each point (e.g., type of symbol, size, color, whether or not it is solid).

    - When making a line plot with the SERIES, REG, or LOESS statements, one can use the LINEATTRS option to control properties of the line (e.g., color, thickness, pattern of hashing, etc.).

  - For all of these options, the SAS syntax is very similar to what is shown above for the FILLATTRS option.

- With the VBAR and HBAR statements, the FILLATTRS option allows one to change the color of the bars and their transparency level.

- **Look in the SAS documentation for more details on the VBAR statement and change the transparency level by adding the appropriate option (use transparency level = 0.5).**

Step 3: Type AND run the following code that modifies the previous SGPLOT step to provide a "skin" to the bars.

```
title "Task 1 - Step 3: Number of Subjects by Treatment Group";
proc sgplot data = echo.DM;
 vbar armcd / fillattrs=(color=lightRed) dataskin=pressed;
run;
```

- Isn't that fancy! Look in the SAS documentation for the other possible values of the DATASKIN option and experiment to see which ones, if any, you like.

Step 4: Type AND run the following code that modifies the previous SGPLOT step to present a percent on the Y-axis and to change the X-axis label by using the XAXIS statement.

```
title "Task 1 - Step 4: Number of Subjects by Treatment Group";
proc sgplot data = echo.DM;
 vbar armcd / fillattrs=(color=lightRed) dataskin=pressed stat=percent;
 xaxis label ='Treatment Group';
run;
```

- One could also use a LABEL statement to temporarily modify the label for the ARMCD variable to achieve the same effect.

- The XAXIS (and YAXIS) statement allows one to do many more things aside from setting an axis label.

- **Add a YAXIS statement to the plot and set some Y-axis attribute that you feel makes the graph more visually appealing. *Look in the SAS documentation for more details on the YAXIS statement. You are free to modify any X- and Y-axis attributes you choose but you must add a Y-axis statement.***

Step 5: Note that the title of the graph from Step 4 is embedded in the image. Sometimes you will want this to be the case but many (most) times you will not. You can control this behavior by using the NOGTITLE option on an ODS statement for the destination currently being written to (or when you open a new destination).

Copy AND run the following code from Step 4 after running the ODS statement below. Compare the graph created to that from Step 4:

```
ods <destination> nogtitle;
```

- Note that the title will no longer be embedded in the image from this point forward in the program. You can revert to the default behavior by running the following code:

  ```
  ods <destination> gtitle;
  ```

  You do not have to include this code in your submitted program but experiment with it.

Task 2: For this task, you will practice making a **grouped** bar chart and using some additional options.

Step 1: Type AND run the following code to produce a vertical bar chart that counts the number of subjects in both sexes and groups by treatment:

```
title1 "Task 2 - Step 1: Number of Subjects by Sex and Treatment
Group";
proc sgplot data = echo.DM;
 vbar sex / group=armcd stat=percent;
 label armcd = 'Treatment Group';
run;
```

- Note that the graph has one bar per value of SEX but the bars for sex are subdivided according to the variable listed in the GROUP option.
- This is referred to as a **stacked** or **group-stacked** bar char.
- A LABEL statement is used to change the legend label. This can also be done by adding a KEYLEGEND statement. The KEYLEGEND statement can be used to customize the legend in more detail.
- **Modify the code to use the KEYLEGEND statement and use the appropriate option to put the legend to the right of the bar chart instead of below the bar chart. *Look in the SAS documentation for more details on the KEYLEGEND statement.***

Step 2: Type AND run the following code that modifies the previous SGPLOT step so that groups are arranged in clusters instead of being stacked.

```
title1 "Task 2 - Step 2: Number of Subjects by Sex and Treatment
Group";
proc sgplot data = echo.DM;
 vbar sex / group=armcd groupdisplay=cluster stat=percent;
 label armcd = 'Treatment Group';
run;
```

- One can modify how groups are displayed using the GROUPDISPLAY option. This option can have values equal to STACK or CLUSTER.

- When GROUPDISPLAY=CLUSTER, bars for each group are created side-by-side for each value of the analysis variable (i.e., SEX in this case).

Task 3: For this task you must create a dataset named WORK.DM_USA that contains the following variables: USUBJID ARMCD SEX AGE SYSBP DIABP HEIGHT WEIGHT HR BMI. **No code will be given to assist you with this portion of the lab.**

- The dataset should contain one observation for each ECHO trial subject that was enrolled in the United States (COUNTRY='USA').

- The variables ARMCD, SEX, and AGE should be pulled from the DM dataset.

- The variables SYSBP, DIABP, HEIGHT, WEIGHT, and HR should contain the vital sign measurements from the **SCREENING** visit. These data are stored in the VS dataset.

- The BMI variable should be derived according to the following formula:

$$\text{BMI} = \frac{(\text{weight in kilograms})}{\text{height in meters}^2}$$

- If done correctly (and sorted by USUBJID), the first 10 observations of the WORK.DM_USA dataset should match the following:

Table 1: First 10 Observations from WORK.DM_USA

Unique Subject Identifier	Age	Sex	Planned Arm Code	Diastolic Blood Pressure	Height	Heart Rate	Systolic Blood Pressure	Weight	Body Mass Index
ECHO-011-001	59	M	ECHOMAX	112	174.4	64	139	69.9	22.98
ECHO-011-002	61	M	PLACEBO	106	175.5	63	135	72.0	23.38
ECHO-011-003	55	M	ECHOMAX	109	187.7	62	146	92.1	26.14
ECHO-011-004	67	M	PLACEBO	103	179.9	59	158	69.8	21.57
ECHO-011-005	83	F	PLACEBO	106	155.9	61	140	40.3	16.58
ECHO-011-006	68	F	PLACEBO	97	177.9	59	144	85.6	27.05
ECHO-011-007	62	F	PLACEBO	91	161.3	59	136	45.5	17.49
ECHO-011-008	43	F	ECHOMAX	103	162.4	50	138	53.1	20.13
ECHO-011-009	66	F	ECHOMAX	117	180.1	60	144	70.7	21.80
ECHO-011-010	58	M	ECHOMAX	96	172.6	51	145	73.9	24.81

- This dataset will be used as input for **Tasks 4-6** of this lab.

BIOS 511 Lab 8
Creating Graphs in SAS - I

Task 4: For this task, you will practice making a scatter plot.

Step 1: Type AND run the following code to produce a scatter plot of height values by BMI values:

```
title1 "Task 4 - Step 1: Scatter plot of Height by Body Mass Index";
proc sgplot data = DM_USA;
 scatter x=height y=BMI;
run;
```

- Note that the X- and Y-axis labels are equal to the labels of the variables plotted by default. You could use XAXIS and YAXIS statements to set them as well.

Step 2: Type AND run the following code that modifies the previous SGPLOT step to change the scatter plot points (a.k.a., markers) to be dark blue filled circles.

```
title1 "Task 4 - Step 2: Scatter plot of Height by Body Mass Index";
proc sgplot data = DM_USA;
 scatter x=height y=BMI /
             markerattrs=(symbol=circleFilled color=darkBlue);
 xaxis label="Height (cm)" values=(150 to 210 by 10);
 yaxis values=(5 to 35 by 5);
run;
```

- Note that and XAXIS statement is used here to change the default axis label (the label of the HEIGHT variable) and to specify the lower and upper range for the axis as well as where tick marks occur (i.e., the “by” portion of the VALUES option).

- Note that the VALUES option requires that the range be specified in parentheses.

Step 3: Type AND run the following code that modifies the previous SGPLOT step to identify the scatter plot points by group.

```
title1 "Task 4 - Step 3: Scatter plot of Height by Body Mass Index";
proc sgplot data = DM_USA;
 format sex $gend.;
 scatter x=height y=BMI / markerattrs=(symbol=circleFilled) group=sex;
 xaxis label="Height (cm)" values=(150 to 210 by 10);
 yaxis values=(5 to 35 by 5);
run;
```

- Note that when using the GROUP option to request that scatter plot points be identified by group, by default the points are distinguished by color and so one would generally not specify the COLOR suboption for the MARKERATTRS option as doing so will make the groups indistinguishable.

- **The provided code makes use of a FORMAT statement so that the values M and F for the SEX variable are replaced with Male and Female in the graph legend. Write the appropriate PROC FORMAT step above the PROC SGPLOT step to create the needed character format.**

Task 5: For this task, you will practice using an ODS GRAPHICS statement.

Step 1: Type AND run the following code to produce a scatter plot of height values by BMI values:

```
title1 "Task 5 - Step 1: Scatter plot of Height by Body Mass Index";
proc sgplot data = DM_USA;
 scatter x=height y=BMI;
run;
```

Step 2: Now, type and run the following code to produce the same scatter plot after submitting the ODS GRAHICS statement that requests a 4x4 image without a border:

```
ods graphics / height=4in width=4in noborder;
title1 "Task 5 - Step 2: Scatter plot of Height by Body Mass Index";
proc sgplot data = DM_USA;
 scatter x=height y=BMI;
run;
ods graphics / reset=all;
```

- Note that after the SGPLOT step, we *may* submit an ODS graphics statement to reset all the ODS graphics options to their default values.

- Much more can be done with the ODS GRAPHICS statement than what we have shown here. ***Read in the SAS documentation for more details on the ODS GRAPHICS statement.***

BIOS 511 Lab 8
Creating Graphs in SAS - I

Task 6: For this task, you will learn how a BY statement impacts the production of graphics. Type and run the following code to produce a scatter plot and regression line (line of best fit) for each treatment group separately.

```
proc format;
 value $ trt
  "ECHOMAX" = "Investigational Treatment"
  "PLACEBO" = "Placebo";
run;

proc sort data = DM_USA out = DM_USA2;
      by armcd;
run;

option nobyline;
title1 "Task 6: Scatter plot of Height by Body Mass Index";
title2 "Treatment Group = #byval(armcd)";
proc sgplot data = DM_USA2 noautolegend;
 by armcd;
 format armcd $trt.;
 reg x=height y=BMI / markerattrs=(size=4 symbol=diamondFilled color=Blue)
                      lineattrs=(pattern=2 thickness=2 color=darkRed);
run;
option byline;
```

- In order to produce a separate graph for each treatment group (i.e., use a BY statement), the input dataset must first be sorted by the variable.

- Because we are using a BY statement in the analysis PROC, we can make use of the #BYVAL() command to dynamically create titles for each plot using the value of the variable listed in the BY statement.

- We have turned off the default “by-line” title using the NOBYLINE system option.

- The formatted values of ARMCD are used in the second title because of the FORMAT statement in the PROC SGPLOT step.

- The REG statement requests that a regression curve (linear by default) be fit to the data and plotted on top of the scatter plot points.

- Because the REG statement generates a scatter plot overlaid with a line plot, both the MARKERATTRS and LINEATTRS options can be used to control the format of the graph.