BIOS 511 Lab 10
Advanced Data Step Programming

Please read the following instructions carefully before beginning this lab. You will need to start by downloading the PC dataset from the LAB-10 assignment on the Sakai site. The PC and DM datasets will be the basis of all activities for this lab.

READ ALL OF THE FOLLOWING INSTRUCTIONS BEFORE STARTING THE LAB:

- Task 2 and 3 should be completed in two SAS programs named lab-10-PID-Task-X.sas where PID is your student PID number and X is the task number. Please make sure to include an appropriate headers in both SAS programs.

- Upload the SAS programs and SAS logs to the Sakai site to document completion of the lab *if you choose to complete the lab*.

- The purpose of this lab is to help you program/understand *complex* tasks.

    - You are encouraged to try and write programming of your own.

    - I recommend that you spend approximately 30m on a task (or more) attempting to write code to complete the task yourself.

    - If you are unsuccessful writing code from scratch after some effort, turn to the solution programs that are available on the Sakai site for guidance. There is no shame in that! I want to encourage you to think about these fairly hard problems on your own and in groups, but it is quite useful to learn by studying examples too.

    - If you are unable to write your own code, my expectation is that you will write comments that explain what the various sections of those programs do. For full credit on the lab, you must do this. *It is not enough to simply copy + paste the programs into you own and run them.*

BIOS 511 Lab 10
Advanced Data Step Programming

PC – Pharmacokinetic (PK) Concentrations data for the ECHO clinical trial (# obs = 3102)

The PC dataset contains data collected about tissue concentrations of analytes (usually study drugs and/or their metabolites) as a function of time after dosing the study drug. In the case of the ECHO trial, the PC dataset structure is one observation per subject per time point at which the ECHO-MAX study drug metabolite was measured. Pharmacokinetic concentration data is used to help understand how quickly study drug is metabolized and therefore removed from the body. Such information informs how much of the drug is absorbed by the body and how often study drug needs to be taken. The variables contained in the PC dataset along with their types, lengths, and labels are included in the table below. A description of the values stored in each variable is also provided.

Variables contained in the PC dataset					
#	**Variable**	**Type**	**Len**	**Label**	**Description**
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	PCSEQ	Num	8	Sequence Number	A unique identifier for observations within a subject. The values of PCSEQ should start at 1 for each subject and should increment by 1 for each successive observation for the subject.
4	PCTESTCD	Char	8	Pharmacokinetic Test Short Name	Always equal to “ECHOMETA”.
5	PCTEST	Char	40	Pharmacokinetic Test Name	Always equal to “ECHO-MAX Metabolite”
6	PCSTRESC	Char	200	Character Result/Finding in Std. Format	Character version of PC test result. PCSTRESC will be non-missing whenever the test was performed.
7	PCSTRESN	Num	8	Numeric Result/Finding in Standard Units	Numeric version of PC test result. PCSTRESN will be non-missing whenever the test was performed and test result was above the lower limit of quantification for the test.
8	PCSTRESU	Char	25	Standard Units	Always equal to “ng/mL”.
9	PCSTAT	Char	8	Completion Status	Will be equal to “NOT DONE” when test was not performed for some reason.
10	PCREASND	Char	200	Reason Test Not Done	Only non-missing when PCSTAT = ‘NOT DONE’. In these cases, PCREASND will indicate the reason why the test was not performed.
11	PCLLOQ	Num	8	Lower Limit of Quantitation	Numeric lower limit of quantification for the pharmacokinetic assay. Always equal to 0.01.
12	VISITNUM	Num	8	Visit Number	Pharmacokinetic tests were performed at baseline visit. Always equal to 1.
13	VISIT	Char	50	Visit Name	Pharmacokinetic tests were performed at baseline visit. Always equal to “Week 0”.
14	PCDTC	Char	20	Date/Time of Specimen Collection	Date of Pharmacokinetic testing in YYYY-MM-DD format.
15	PCTPT	Char	50	Planned Time Point Name	Character time-point indicating elapsed time from dose of study drug to blood sample draw.
16	PCTPTNUM	Num	8	Planned Time Point Number	Numeric time-point indicating elapsed time from dose of study drug to blood sample draw.

BIOS 511 Lab 10
Advanced Data Step Programming

Task 1 – No output or program should be turned in for this task.

Write a PROC PRINT step to print out the data for two subjects: ECHO-011-001 and ECHO-019-018. There is no precise requirement regarding exactly which variables to print. The purpose of this task is to help you understand the data structure for the PC dataset. A similar approach should be taken for any new dataset you encounter in practice.

Note the following:

- When a concentration is below the lower limit of quantification (PCLLOQ) for the assay used, the value of PCSTRESN will be missing but the value of PCSTRESC will be non-missing to indicate that the pharmacokinetic concentration in the serum blood sample was too low to measure using the assay.

- If a test was not performed at a time point, both PCSTRESC and PCSTRESN will be missing. This is not the same thing as the test result being too low for the assay to quantify it precisely. When the test is not performed, PCSTAT = "NOT DONE' and the reason will be stored in PCREASND. (e.g., PCREASND='Blood Sample Hemolyzed').

- The PCTPT variable is character but stores the time point at which the serum blood sample was collected. For example, PCTPT= 0.25 HOURS POST-DOSE indicates that the sample was collected 0.25 hours (or 15 minutes) after the dose of study drug was taken. The PCTPTNUM variable is a numeric version of PCTPT, which can be used for sorting and array indexing.

- If one needs to create a numeric variable storing the number of hours elapsed since the dose of study drug, that information needs to be extracted from the PCTPT variable using SAS functions such as INPUT, SCAN, COMPRESS, etc. There are many different ways to do this. You will need to accomplish this for other parts of the lab but you need not do so here.

BIOS 511 Lab 10
Advanced Data Step Programming

Task 2

Compute at least the number missing, number non-missing, mean, standard deviation, and median of the pharmacokinetic concentration values (e.g. based on the PCSTRESC and PCSTRESN variables) at each time point *for males, females, and overall*. Use the PROC of your choice to do this. Depending on how you go about this, you may need multiple PROC steps.

Your printed output does not need to resemble the output produced by the solution program. The solution program uses PROC REPORT which is a topic that we do not cover in BIOS 511. The REPORT procedure can make detailed reports/tables that are much more visually appealing than can be produced with the PRINT procedure. Take BIOS 669 if you would like to learn more about the REPORT Procedure. Otherwise the programming techniques used in the solution program should be reviewed as a training exercise.

Prior to computing the summary statistics, impute concentration values that are below the lower limit of quantification to be equal to the actual lower limit of quantification. That is to say, when PCSTRESC = "<0.01", impute the missing numeric test result to be 0.01. Do this in such a way that your code will always work even if the lower limit of quantification is not the same for each subject (e.g. if the value of the PCLLOQ variable differs by subject and/or the value of PCSTRESC would sometimes equal things such as "<0.25"). It is common for study sites to use different assays that do not always have the same lower detection limits. You can assume PCSTRESC values will be of the form "<XXX" where XXX is a number having one or more decimal places.

Answer the following questions based on the output for the requested analysis. These are practice problems similar to what one might see on the final exam and the answers are included so that you can check yourself. The output that your program produces should display these answers in an easily finable way as would be required for an exam.

Question 1: At which time point does the highest mean PK concentration occur (based on the overall analysis)?

Answer 1: 2.00 Hours Post-Dose

Question 2: What is the value of the highest mean concentration?

Answer 2: 9.01

Question 3: For the time point identified for Question 1, how many test results were missing because the test was NOT DONE for any reason?

Answer 3: 3

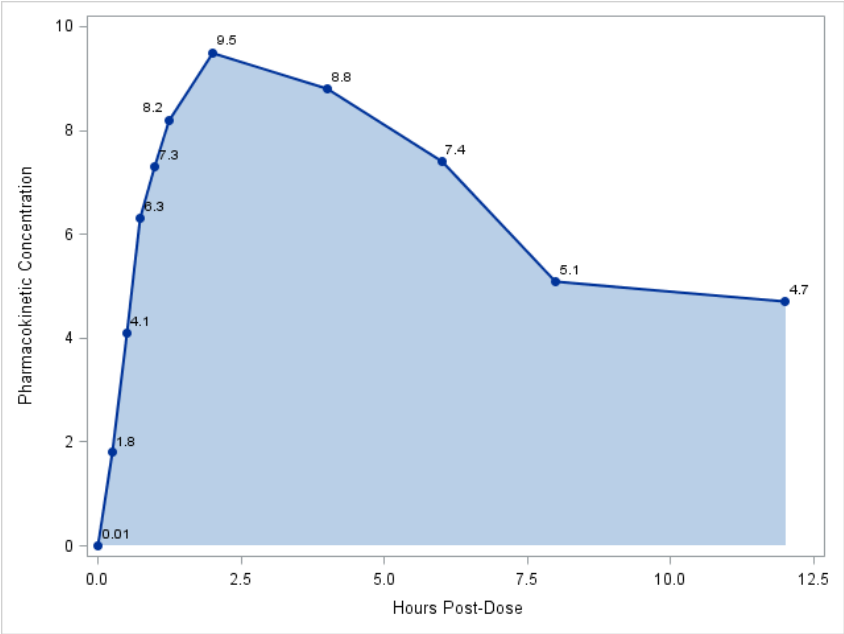
BIOS 511 Lab 10
Advanced Data Step Programming

Task 3

Produce a CSV file named ECHO_AUC.csv that contains one observation per subject and that contains the following variables: USUBJID, AUC12.

The variable AUC12 should store the area under a subject's pharmacokinetic concentration curve over the 12 hour period following the dose of study drug. The following graph illustrates the area that should be computed for subject ECHO-011-001.

Figure 1: Pharmacokinetic Concentration Curve for USUBJID=ECHO-011-001



The area under the curve between any two time points (e.g. 6.00 hours post-dose and 8.00 hours post-dose) corresponds to a trapezoid. The formula for the area of the trapezoid is 0.5*(value #1 + value #2)*(elapsed time between two time points). Here, value #1 is the concentration value from the earlier time point and concentration #2 is the concentration value from the later time point. For this subject, the AUC12 contribution for the area between 6.00 hours post-dose and 8.00 hours post-dose would be as follows:

AUC12 Contribution = 0.5*(7.4 + 5.1)*2.00 = 12.5

There are many ways to go about producing this dataset. Multiple solutions are provided in the lab-10-solution-task3.sas program.