

BIOS 511 Lab 7
Reading/Writing Non-SAS Datasets

Please read the following instructions carefully before beginning this lab.

Instructions:

- All tasks should be completed in a single SAS program named lab-07-PID.sas where PID is your student PID number. Please make sure to include an appropriate header in your SAS program.
- For this lab you will turn in your program, log, an excel file (task 4) and a CSV file (task 5). Failure to follow the instructions for naming the excel and csv files will result in you not receiving credit for those parts of the lab.
- In your code, before starting a new task, include a block comment with the task number:

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

BIOS 511 Lab 7

Reading/Writing Non-SAS Datasets

Task 1: Reading in the genotype data.

- Your collaborator has provided you a text file (ECHO_GENOTYPE.DAT) that contains the genotype for a subset of subjects from the ECHOMAX treatment group.
- The file is CSV delimited and contains a header row that lists the variable names.

An excerpt from the top of the data file is given below:

Table 1: Excerpt from ECHO_GENOTYPE.DAT file

```
site,subject,genotype,reason
11,1,B,
11,3, ,Insufficient Sample
11,8, ,Insufficient Sample
11,9,A,
11,10,B,
11,11,A,
11,12,B,
11,13,B,
11,14,A,
11,16, ,Insufficient Sample
12,1,B,
12,3,B,
12,9,A,
12,12,A,
12,13,B,
12,17,B,
12,20,A,
```

The file contains the following variables:

- Site – numeric site number
- Subject – numeric subject number within each site
- Genotype – character version of genotype – either “A” or “B” if genotype assay was successful
- Reason – reason genotype assay wasn’t successful

Write a DATA step that reads the data from the ECHO_GENOTYPE.DAT file and creates a SAS dataset named WORK.GENOTYPE. The SAS dataset should have the same four variables and should create a variable named USUBJID that allows the WORK.GENOTYPE dataset to be merged with the DM dataset (recall the structure of USUBJID which includes both site and subject identifiers in a single variable).

To create the USUBJID variable you will need to convert and concatenate the numeric SITE and SUBJECT variables values to form a character variable. A SITE value of 1 must convert to the three-character value ‘001’. This can be done using the PUT function and the Z3 format.

BIOS 511 Lab 7

Reading/Writing Non-SAS Datasets

Task 2: Dataset processing.

For this part of the task, you need to create a dataset named WORK.GENOTYPE2 with one observation per subject that has the GENOTYPE and REASON variables that are found in the WORK.GENOTYPE dataset, the SEX variable that is found in the ECHO.DM dataset, and the CHANGE_DBP and CHANGE_SBP variables that must be created from the ECHO.VS dataset. The dataset should also contain the USUBJID variable.

The CHANGE_DBP and CHANGE_SBP variables are to be the subject's change from baseline to week 32 values for diastolic and systolic blood pressure respectively. To compute change from baseline in diastolic blood pressure, one needs to compute the following:

$$\text{CHANGE_DBP} = (\text{Value of diastolic blood pressure from the week 32 observation}) - (\text{Value of diastolic blood pressure from the baseline observation (where VSBLFL='Y')})$$

Change from baseline for systolic blood pressure is to be created in similar fashion. The WORK.GENOTYPE2 dataset should contain one observation for each subject that has an observation in the WORK.GENOTYPE dataset. An excerpt of the correct WORK.GENOTYPE2 dataset is given below.

Table 2: Excerpt from WORK.GENOTYPE2 dataset (from SDM viewer)

VIEWTABLE: Work.Genotype2						
	USUBJID	SEX	genotype	reason	change_dbp	change_sbp
1	ECHO-011-001	M	B		-6	-7
2	ECHO-011-003	M		Insufficient Sample	-9	-9
3	ECHO-011-008	F		Insufficient Sample	-8	-5
4	ECHO-011-009	F	A		-8	-12
5	ECHO-011-010	M	B		-12	-2
6	ECHO-011-011	F	A		-6	-9
7	ECHO-011-012	M	B		-10	-7
8	ECHO-011-013	M	B		-3	-4
9	ECHO-011-014	M	A		-8	-3
10	ECHO-011-016	F		Insufficient Sample	-9	-4
11	ECHO-012-001	F	B		-4	-7
12	ECHO-012-003	F	B		-14	-7
13	ECHO-012-009	F	A		-8	-7
14	ECHO-012-012	M	A		-5	-10
15	ECHO-012-013	F	B		-5	-11
16	ECHO-012-017	M	B		-9	-6
17	ECHO-012-020	M	A		-2	-2
18	ECHO-013-001	M	A		-4	-10
19	ECHO-013-002	M	A		-5	0
20	ECHO-013-007	M	A		-6	1
21	ECHO-013-008	M	B		-9	-3
22	ECHO-013-009	M	B		-11	-5
23	ECHO-013-011	M	A		-11	-9
24	ECHO-014-004	F	B		5	-7
25	ECHO-014-012	F	A		-9	-9
26	ECHO-014-013	F	B		-6	-6

BIOS 511 Lab 7

Reading/Writing Non-SAS Datasets

Task 3: Data Analysis

For this part, you must analyze the change from baseline data for each genotype and blood pressure measurement (i.e., systolic BP and diastolic BP). Write a PROC MEANS step to analyze the data and produce a table similar to what is shown below.

Summary of Change from Baseline in Systolic and Diastolic Blood Pressure by Genotype

Genotype	N Obs	Variable	Label	N	Mean	Std Dev	Median	Minimum	Maximum
	16	change_dbp	Change in Diastolic Blood Pressure at 32 Weeks	16	-6.00	3.01	-6.00	-11.00	0.00
		change_sbp	Change in Systolic Blood Pressure at 32 Weeks	16	-5.81	2.86	-5.00	-10.00	0.00
A	136	change_dbp	Change in Diastolic Blood Pressure at 32 Weeks	136	-6.51	3.44	-6.50	-16.00	3.00
		change_sbp	Change in Systolic Blood Pressure at 32 Weeks	136	-6.69	3.63	-7.00	-18.00	1.00
B	130	change_dbp	Change in Diastolic Blood Pressure at 32 Weeks	130	-5.97	3.77	-6.00	-14.00	5.00
		change_sbp	Change in Systolic Blood Pressure at 32 Weeks	130	-5.33	3.54	-5.00	-14.00	2.00

To reproduce this table, you will need to use a CLASS statement and an option for the CLASS statement that instructs PROC MEANS to summarize the data for missing values of class variables. You will also need to use the MAXDEC option for the PROC MEANS statement to control the number of decimals that are displayed.

In addition to displayed output, direct the PROC MEANS results (only the number of non-missing observations and mean) to a SAS dataset using the OUTPUT statement. If correctly done, the results should mirror the following dataset (named WORK.RESULTS shown in SDM viewer).

	genotype	n	dbp	sbp
1	U	16	-6	-5.8125
2	A	136	-6.507352941	-6.691176471
3	B	130	-5.969230769	-5.330769231

Task 4: Exporting Data (Excel)

Aside from producing a basic data summary, you are tasked with providing a dataset in Excel format to another collaborator for downstream analysis. This collaborator wants the data in Excel format. Using PROC EXPORT, write the contents of the WORK.GENOTYPE2 dataset to an Excel workbook named "ECHO_GENOTYPE.XLSX" and name the worksheet "ECHO".

Task 5: Exporting Data (CSV)

Another collaborator wants analysis results from part 3 in a file in CSV format. Using a DATA step, write the contents of the WORK.RESULTS dataset to a CSV file named "ECHO_GENOTYPE.CSV". Include a header row in the file which contains the variable names. In the data step that writes the CSV file, change the missing genotype value to a value of "U" before the data are written.