

Macro to Generate Summary and Descriptive Statistics Tables

Yogesh Pande, Merck Sharp & Dohme Corp., Rahway, New Jersey

ABSTRACT

The most popular type of tables in the pharmaceutical industry are those that have descriptive statistics (n mean std median maximum) or tables having frequency (%) count and descriptive statistics of categorical and continuous variables. Often requested by statisticians, medical writers, and clinical research scientists, these tables have to be generated in lesser time spans and with stringent quality concerns. Using SAS[®] macro language and SAS[®] base, this paper will introduce a macro that generates a table having categorical variables (for frequency count) and continuous variables (for descriptive statistics). For laboratory, vital signs, and ECG data, this macro can generate tables having descriptive statistics for "Change from Baseline."

Keywords: Summary, Descriptive, Macro, continuous, categorical

1. DATA CHECKS AND MACRO OVERVIEW

- 1) This macro is useful for generating frequency tables and/or tables containing summary statistics.
- 2) The user can have a maximum of five titles and five footnotes in the table.
- 3) Below are listed data checks this macro will complete prior to creating a table:

DATA CHECKS:

- (a) The code will check whether the data has one subject per observation or multiple observations per subject.
- (b) If the data has multiple observations per subject, the code will check whether each subject has multiple observations within same visit.
- (c) If it has multiple observations within a same visit, the macro will generate the following warning in the log - "This dataset has multiple observations within same visit. Please make this dataset having one observation per visit." Before using this macro, the programmer has to remove all duplicate records appearing within the same visit.

In case of laboratory, vital sign, or ECG data where each subject has multiple parameters and under these parameters each subject has multiple visits, this macro will check under any particular parameter if each subject has multiple records within similar visit. If there are any, the warning message as quoted above will be displayed in the log and the macro will stop executing. Following is a screen shot of laboratory data;

Subject	Lab parameter	Lab visit	Period
101	Basophiles	Screening	Pre-Treatment
101	Basophiles	Baseline	Pre-Treatment
101	Basophiles	Visit 1	In-Treatment
101	Basophiles	Visit 2	In-Treatment
101	Basophiles	Visit 2	In-Treatment
101	Basophiles	Visit 3	In-Treatment
101	Basophiles	Visit 4	In- Treatment

Table 1: Laboratory data showing parameter, visits associated with it and the treatment period.

EXPLANATION OF TABLE 1:

In the laboratory data as shown above in parameter Basophile, seven observations are generated. The highlighted portion in the above table is the point of interest. This macro will check whether subject number 101 has multiple observations for a similar visit (in this case Visit 2). For this situation, the macro will generate the following warning in the log - "This dataset has multiple observations within visit/parameter. Please make this dataset having one observation per visit." After receiving this warning, the programmer has to decide which observation within those duplicate records (Visit 2) needs to be selected for use of the above data in the macro.

- (d) The code will also check whether the dataset is an empty/no records/no observation. If the dataset is empty and if the user has entered 'Y' value for parameter "Empty", then an empty report will be generated.

- 4) To successfully use this macro, the following two datasets are required: &DSIN which is the main dataset having the variables that will be used to generate the table and &DSO having one observation per subject. The &DSO dataset is required in order to identify the total number of subjects (N=XXX) in each treatment group.

Parameter	Statistics	Treatment		
		Trt1 (N=XXX)	Trt2 (N=XXX)	Trt3 (N=XXX)

Table 2: Header showing total number of subjects in each treatment group, this N=XXX is from &DSO data

- 5) Please ensure that the three categorical and three continuous variables to be included in the summary table have labels associated with them in the dataset (&dsin). This macro captures those labels to actually describe the variable in the generated summary table.
- 6) Frequently within a table having laboratory data and vital sign data, the summary statistics are generated for parameter value and change from baseline value associated with each parameter. This macro can generate summary statistics including change from baseline. For the screen shot, below, please see Table 5

In the following pages, please see the screen shot of the various types of tables this macro can generate.

Parameter	Statistics	Treatment	
		Trt1	Trt2
Categorical variable -1	No	XX(XX.X)	XX(XX.X)
	Yes	XX(XX.X)	XX(XX.X)
Categorical variable -2	> 3 to =< 6 months	XX(XX.X)	XX(XX.X)
	> 6 months	XX(XX.X)	XX(XX.X)
Categorical variable -3	No	XX(XX.X)	XX(XX.X)

Table 3: This table illustrates that, (if asked in the macro call) the macro can generate only a frequency table.

Parameter	Statistics	Treatment		
		Trt1	Trt2	Trt3
Drug exposure	n	XXX	XXX	XXX
	Mean (SD)	X.X(X.X)	X.X(X.X)	X.X(X.X)
	Median	XX.X	XX.X	XX.X
	Min-Max	XX- XX	XX- XX	XX- XX
Ethnicity n (%)	Hispanic or Latino	XX(XX.X)	XX(XX.X)	XX(XX.X)
	Not Hispanic or Latino	XX(XX.X)	XX(XX.X)	XX(XX.X)
Gender n (%)	Female	XX(XX.X)	XX(XX.X)	XX(XX.X)
	Male	XX(XX.X)	XX(XX.X)	XX(XX.X)
Height (cm)	n	XXX	XXX	XXX
	Mean (SD)	X.X(X.X)	X.X(X.X)	X.X(X.X)
	Median	XX.X	XX.X	XX.X
	Min-Max	XX- XX	XX- XX	XX- XX

This parameter description is actually the label associated with the variable.

Table 4: This table illustrates that the macro can generate combined tables having summary statistics and frequency table

Table 5 Laboratory data summary statistics by visit and treatment

			Treatment			
			trt1 (N=XXX)		trt2 (N=XXX)	
Parameter	Visit		Value	Change from baseline	Value	Change from baseline
Laboratory Parameter	Baseline	n	XXX	0	XXX	0
		Mean (SD)	X.X(X.X)		X.X(X.X)	
		Median	XX.X		XX.X	
		Min-Max	XX- XX		XX- XX	
	Visit 1	n	XXX	XXX	XXX	XXX
		Mean (SD)	X.X(X.X)	X.X(X.X)	X.X(X.X)	X.X(X.X)
		Median	XX.X	XX.X	XX.X	XX.X
		Min-Max	XX- XX	XX- XX	XX- XX	XX- XX
	Visit 2	n	XXX	XXX	XXX	XXX
		Mean (SD)	X.X(X.X)	X.X(X.X)	X.X(X.X)	X.X(X.X)
		Median	XX.X	XX.X	XX.X	XX.X
		Min-Max	XX- XX	XX- XX	XX- XX	XX- XX

There are no data in this listing/table

Table 6: This is how the empty report will be printed in the RTF format.

2. MACRO PARAMETERS:

Please see the table, giving each and every macro parameters explanation, with details which macro parameter is mandatory and which macro parameter is optional.

Parameter	Explanation	Value range	Mandatory/optional
Dsin	The dataset that has the variables required for the summary table	Dataset name in the existing library	Mandatory
Select	This key parameter puts additional conditions on DSIN	Condition for DSIN data (Ex: Select=%str(if treat=1)	Mandatory
Subj	The variable that gives the subject numbers	Variable name	Mandatory
TYPE	If there is a variable in DSIN data having multiple parameters, then this key parameter should not be kept blank	Example : Vital sign data TYPE=vital, for Laboratory data TYPE=lab	Mandatory only when data ("DSIN") is having multiple parameters, if the data is not having multiple parameters, then "TYPE" should be kept blank.
PARAM	This PARAM gives the variable name that has multiple parameters	The variable name	Mandatory only if TYPE is filled, otherwise keep it blank
Unit	This macro variable will give the variable name that gives the unit associated with the lab, vital sign or ECG parameter.	Variable name that gives the unit.	Mandatory if TYPE is filled.

Parameter	Explanation	Value range	Mandatory/optional
Value	This key parameter will have a variable name that gives the actual value associated with any Laboratory, vital or ECG parameter	Variable name that gives the value.	Mandatory if TYPE is filled.
CHG	This key parameter will have a variable name that gives the change from baseline value associated with any Laboratory, vital or ECG parameter	Variable name that gives change from baseline.	Mandatory if TYPE, is filled.
timvar= Varfmt=	TIMVAR is the variable that gives visit information VARFMT is the format name associated with TIMVAR	TIMVAR: Visit variable VARFMT: Format associated with visit variable.	Mandatory only when DSIN data has visit variable. If visit variable is present in the data VARFMT parameter become mandatory
Dso=	Every table has N=, which gives total subjects in that treatment group. This dataset is used to calculate the total subjects within each treatment group	Dataset name having one subject/ observation. (Ex: Demographic data, baseline characteristic data). This data should have a variable that gives treatment group.	Mandatory
poptrt=	This is a condition for DSO data within PROC SQL, this condition selects subject population, treated group or randomized group.	Example : how to enter the condition while using the macro Poptrt=%str(where treated=1)	Mandatory
Trt= nTrt=	This gives the user to type-in the variables that defines the treatment groups of the trial and ntrt parameter user has to put how many treatment groups the trial has	Trt = The variable name that defines the treatment group Ntrt = Number of treatment groups (Ex: ntrt=2; ntrt=3, ntrt=4 ...)	Mandatory
Title	Number of titles a table can have (Integer value expected)	1-5	Mandatory
t1, t2, t3, t4, t5	Actual string, if title=1 then t1 should be filled, if title=2 then t1 and t2 should be filled and so on	Character string	At least t1 should be filled if "title=1". Rest all other parameter optional
Footnote	Number of footnotes a table can have (Integer value expected)	1-5	Number of footnotes required for the table maximum a table can have is 5 footnotes. If no footnotes are required then this macro key parameter should be kept empty.
f1, f2, f3, f4, f5	Actual string, if footnote=1 then f1 should be filled, if footnote=2 then f1 and f2 should be filled and so on	Character string	Optional, if footnote is not required.
Empty	If the data set is empty, and user wants to create a empty report, this parameter should be filled	Empty='Y': Macro will print a blank report with a string "There are no data in this table". Empty='N': Macro will not create an empty report even though dataset has no observation. Default value of this parameter is 'N'	Optional

Parameter	Explanation	Value range	Mandatory/optional
Cont	Number of continuous variables that needs to be in table. The continuous variable will have descriptive statistics (n, mean, standard deviation, median, minimum and maximum)	1-3	Mandatory, only if descriptive statistics required. If PARAM parameter is filled then CONT should be kept blank
cont1, cont2, cont3	These parameters will have the actual variable name as in the dataset	Variable names from the data	Cont is not missing then these parameters are Mandatory, otherwise optional
Cat	Number of categorical variables that needs to be in table. The categorical variable will have frequency counts along with the percentage	1-3	Mandatory, only if frequency counts are required. If PARAM parameter is filled then CAT should be kept blank
cat1,cat2, cat3	These parameters will have the actual variable name as in the dataset	Variable names from the data	CAT is not missing then these parameters are Mandatory, otherwise optional
Out	This key parameter is for the user to put the output file name.	Ex: Out=table1	Mandatory

3. MACRO CODE

Please follow the link for the macro code that can generate tables as described above.

https://sites.google.com/site/sasfilesyogesh/home/pharmasug2011_macro-code

4. EXAMPLE OF HOW TO CALL A MACRO

*---Laboratory summary statistics including change from baseline---;

```
%summary(dsin=Labdata, type=Lab, subj=sid, timvar=timeregf, varfmt=timereg.,param=prfparam
,unit=prfunit, value=prfvalue, chg=change, trt=trtgrp, ntrt=2
,select=%str(if treated=1 and visit >= 84),empty=N,dso=demo
,poptrt=%str(where treated=1)
,title=2
,t1=Table 5 Laboratory data summary statistics
,t2=by visit and treatment,out=table5);
```

The above example macro call will generate Table 5.

*---Demographic summary statistics (2 continuous and 2 categorical variables)-----;

```
%summary(dsin=demographic, subj=sid, select=%str(if treated=1)), tcaf=trtgrp,ntrt=3
,cont=2,cont1=hgtcm,cont2=wgtkg,cat=2,cat1=sexf,cat2=racef,empty=N
,dso=demographic, poptrt=%str(where treated=1)
,title=2
,t1=Table 2 Demographic summary statistics
,t2=by visit and treatment,out=table2);
```

5. CONCLUSION

This paper has explained how a macro can generate descriptive statistics of various styles by using various macro key parameters. An attempt has been made to explain and show the inquisitive reader that there are many types of descriptive statistics tables one can program. It also gives an idea to all SAS® programmers about the initial data checks that needs to be performed before they start programming or using this macro to generate the summary/descriptive statistics table. I hope this paper works as a handy tool to generate 'Good Quality' summary and descriptive statistics tables.

6. REFERENCE

1. SAS/MACRO Software: version 9.1.3, SAS® Institute Inc., Cary NC
2. SAS/BASE Software: Version 9, SAS® Institute Inc., Cary NC
3. SAS® Guide to Report Writing: Examples, Second Edition
4. SAS® Institute, Inc. (1990), SAS/Graph Software: Reference, First Edition, Cary, NC: SAS Institute, Inc.

CONTACT INFORMATION: Your comment and questions are valued and encouraged.

Yogesh Pande
Sr. Scientific Programmer
Rahway NJ
Merck Sharp & Dohme Corp.
Ph: 732-594-5272.
E-mail: yogesh.pande@merck.com

SAS® and all other SAS® Institute Inc. product or service names are registered trademarks or trademarks of SAS® Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.