A SAS® Macro to Produce Summary Statistics Tables Allowing Additional Stratification

Amanda (Xiao) Liu, Cerus Corporation, Concord, CA

David Tappe, Cerus Corporation, Concord, CA

ABSTRACT

The majority of tables generated in clinical studies involve the comparison of summary statistics of continuous and/or categorical variables across treatments. There are many published macros that can perform these tasks with both efficiency and accuracy. However, when the needed analysis involves a stratification variable (e.g. study sites) and corresponding p-values adjusted by strata, it can be cumbersome to have to run your code repetitively by strata and then integrate with the overall p-values. In this paper, we introduce a user-friendly macro that not only produces the summary statistics of continuous and categorical data by treatment group, but also provides the flexibility of adding a stratified variable of interest as well as selecting the proper test to use in computing p-values with or without strata as a factor.

INTRODUCTION

For clinical trial reporting, summary tables are the most common means used to explore whether treatment groups are comparable or differ in a meaningful way. Many published macros have discussed how to create summary tables for the comparison of continuous and/or categorical variables across therapy groups. When a stratification variable such as study site is requested, unless strata handling is built into the macro, it can be tedious to handle manually by having to run the code repetitively by stratum level and then combine the outputs back together.

The SAS Macro %SUMMARY_STRATA described in this paper allows the flexibility of handling one stratified variable by specifying the variable in a macro parameter. It also can accommodate analyses where there is no stratification and/or a single treatment by leaving stratification and/or treatment parameters blank.

%SUMMARY_STRATA uses basic SAS procedures, PROC SQL and DO LOOP processing to generate descriptive statistics and p-values using the appropriate statistical tests as follows:

- 2-way Analysis of Variance (ANOVA) for continuous variables and Cochran-Mantel-Haenszel (CMH) test of general association for categorical variables if stratification is present
- Choice of 1-way ANOVA or Kruskal Wallis test for continuous variables without stratification
- Choice of Chi-square or Fisher's exact test for categorical variables without stratification

The output of the macro is a SAS dataset that can be fed into PROC REPORT or manipulated further in subsequent SAS code. With this macro, users can produce a summary dataset with or without strata as a factor quickly by setting the appropriate parameters in a single call of the macro for each analysis variable.

MACRO SYNTAX

```
%SUMMARY_STRATA(indsn=,
               select=,
               trt=,
               strat=,
               var=,
               parmname=,
               parmord=,
               analtype=,
               test=,
               idvar=,
               outdsn=);
```

MACRO PARAMETERS

The 11 macro parameters, descriptions, and whether required or optional are summarized in the table below:

Parameter	Description & Parameter Values	Required / Optional
INDSN	Specifies the name of the macro input dataset.	Required
SELECT	Put conditions on INDSN e.g., SELECT=(if saftypopflg=1).	Optional
TRT	Defines the group variable, typically treatment group.	Optional
STRAT	Defines the stratification variable. Leave undefined if not needed. e.g., STRAT=Site	Optional
VAR	Variable to be summarized. e.g., VAR=Sex	Required
PARMNAME	Text that describes the variable to be summarized, for use in reporting. By default, set to the value of &VAR. e.g., PARMNAME=%str(Age at enrollment, in years)	Optional
PARMORD	Unique number to assign each variable being summarized. Useful for ordering variables when creating final summary table. e.g., PARMORD=01	Optional
ANALTYPE	Type of summary statistics to be performed. ANALTYPE=cont (Display N, Mean(SD), Median, 25th to 75th Percentile, and Min to Max). ANALTYPE=cat (Frequency and Column Percentage will be displayed).	Required
TEST*	Statistical test to be performed when computing p-values. When STRAT is not missing and TRT is not missing, the default setting of TEST is used and 2-way ANOVA for continuous and CMH for categorical variables will be performed automatically.	Optional
	When STRAT is missing and TRT is not missing possible values for TEST are: CHISQ, EXACT, ANOVA, KW.	Required
IDVAR	Variable defining the unique entity to be summarized, e.g., the variable for subject ID.	Required
OUTDSN	Specifies the output dataset name.	Required

*See details of macro parameter TEST in Table 2.

Table 1. Summary of Macro Parameters in %SUMMARY_STRATA

The optimal statistical test to perform is dependent on many factors, including the type of data, stratification, distribution assumptions and sparseness of the data. Table 2 below provides a general guideline for use in selecting the most appropriate test under these different scenarios when using %SUMMARY_STRATA.

Variable Type	Stratified?	Distribution Assumption	Test Name / Value of parameter TEST	SAS PROC
Continuous	Yes	Normally distributed dependent variable	2-way ANOVA / ANOVA (Default)	GLM (include treatment & strata)
	No	Normally distributed dependent variable	1-way ANOVA / ANOVA	GLM
		Normal distribution not assumed	Kruskal Wallis / KW	NPAR1WAY
Categorical	Yes	Strata are independent and data not sparse	CMH / CMH (Default)	FREQ with CMH and WEIGHT
	No	Data not sparse (e.g., cell count >=5)	Chi-Square / CHISQ	FREQ with CHISQ
		Sparse data (e.g., cell count < 5)	Fisher's Exact / EXACT	FREQ with FISHER

Table 2. Specifications of Macro Parameter TEST in %SUMMARY_STRATA

EXAMPLES

The output dataset will contain a column named Variable, and $COL_{11} - COL_{nm}$ containing the computed statistics for the nth treatment and mth stratum. If &TRT is specified and the specified variable contains at least 2 distinct values, an additional column containing the computed p-value will also be included. Note that when macro parameters &TRT and &STRAT are both blank in the macro call, a single column of an overall summary statistics will be created. The tables displayed in these examples can be generated directly from the concatenated dataset(s) produced by calls of %SUMMARY_STRATA for each variable being analyzed.

Example1: Demographic and baseline characteristic summary table by treatment stratified by site.

```
%SUMMARY_STRATA(indsn=demog,
                select=(if saftypopflg=1),
                trt=treatment,
                strat=site,
                var=age,
                parmname=%str(Age, (yrs)),
                parmord=01,
                analtype=cont,
                test=ANOVA,
                idvar=usubjid,
                outdsn=age_out);

%SUMMARY_STRATA(indsn=demog,
                select=(if saftypopflg=1),
                trt=treatment,
                strat=site,
                var=race,
                parmname=%str(Race, n(%)),
                parmord=02,
                analtype=cat,
                test=CMH,
                idvar=usubjid,
                outdsn=race_out);
```

| | -------- Site A -------- | | -------- Site B -------- | | |
Variable	Test (N=146)	Control (N=159)	Test (N=172)	Control (N=168)	p-value*
Age, (yrs)					
N	146	159	172	168	0.311
Mean (SD)	46.8 (14.3)	46.3 (13.9)	46.9 (14.4)	45.1 (15.8)	
Median	49.0	48.0	49.5	46.0	
25th to 75th Percentile	35.0 to 57.0	38.0 to 56.0	38.0 to 58.0	37.0 to 56.5	
Min to Max	8.0 to 77.0	7.0 to 74.0	7.0 to 85.0	6.0 to 75.0	
Race, n(%)					
Caucasian	133 (91.1)	149 (93.7)	157 (91.3)	150 (89.3)	0.878
African-American	6 (4.1)	3 (1.9)	4 (2.3)	8 (4.8)	
Asian	0 (0.0)	2 (1.3)	2 (1.2)	2 (1.2)	
Hispanic	3 (2.1)	4 (2.5)	7 (4.1)	5 (3.0)	
Other	4 (2.7)	1 (0.6)	2 (1.2)	3 (1.8)	

* P-value for treatment differences using 2-way ANOVA (with treatment and site in the model) for Age and CMH test of general association (stratified by site) for Race.

Output 1. Demographics and Baseline Characteristics by Treatment and Site

Example 2: Demographics and baseline characteristics summary table by treatment.

```
%SUMMARY_STRATA(indsn=demog,
               select=(if saftypopflg=1),
               trt=treatment,
               strat=,
               var=age,
               parmname=%str(Age, (yrs)),
               parmord=01,
               analtype=cont,
               test=anova,
               idvar=usubjid,
               outdsn=age_out);

%SUMMARY_STRATA(indsn=demog,
               select=(if saftypopflg=1),
               trt=treatment,
               strat=,
               var=race,
               parmname=%str(Race, n(%)),
               parmord=02,
               analtype=cat,
               test=exact,
               idvar=usubjid,
               outdsn=race_out);
```

Variable	Test (N=318)	Control (N=327)	p-value*
Age, (yrs)			
N	318	327	0.382
Mean (SD)	46.9 (14.4)	45.7 (14.9)	
Median	49.0	47.0	
25th to 75th Percentile	36.0 to 57.0	38.0 to 56.0	
Min to Max	7.0 to 85.0	6.0 to 75.0	
Race, n(%)			
Caucasian	290 (91.2)	299 (91.4)	0.899
African-American	10 (3.1)	11 (3.4)	
Asian	2 (0.6)	4 (1.2)	
Hispanic	10 (3.1)	9 (2.8)	
Other	6 (1.9)	4 (1.2)	

* P-value for treatment differences using 1-way ANOVA for Age and Fisher's exact test for Race because of small counts.

Output 2. Demographics and Baseline Characteristics by Treatment

Example 3: Demographics and baseline characteristics summary table (Overall).

```
%SUMMARY_STRATA(indsn=demog,
               select=(if saftypopflg=1),
               trt=,
               strat=,
               var=age,
               parmname=%str(Age, (yrs)),
               parmord=01,
               analtype=cont,
               test=,
               idvar=usubjid,
               outdsn=age_out);
```

4

```
%SUMMARY_STRATA(indsn=demog,
                select=(if saftypopflg=1),
                trt=,
                strat=,
                var=race,
                parmname=%str(Race, n(%)),
                parmord=02,
                analtype=cat,
                test=,
                idvar=usubjid,
                outdsn=race_out);
```

Variable	Overall (N=645)
Age, (yrs)	
N	645
Mean (SD)	46.3 (14.6)
Median	48.0
25th to 75th Percentile	38.0 to 57.0
Min to Max	6.0 to 85.0
Race, n(%)	
Caucasian	589 (91.3)
African-American	21 (3.3)
Asian	6 (0.9)
Hispanic	19 (2.9)
Other	10 (1.6)

Output 3. Demographics and Baseline Characteristics (Overall)

CONCLUSION

This paper has demonstrated how the macro %SUMMARY_STRATA generates summary statistics with or without strata by specifying key macro parameters. Depending on the final reporting needs, the output dataset from the macro can be directly used to generate the needed summary table by PROC REPORT.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Amanda Liu
Cerus Corporation
2550 Stanwell Drive
Concord, CA 94520
Work Phone: (925)-288-6085
E-mail: aliu@cerus.com

David Tappe
Cerus Corporation
2550 Stanwell Drive
Concord, CA 94520
Work Phone: (925)-288-6064
E-mail: dtappe@cerus.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.