

# A SAS® Macro for Quick Descriptive Statistics

Jenna C. Carlson, University of Pittsburgh, Pittsburgh, PA

## ABSTRACT

Arguably, the most required table in publications is the description of the sample table, fondly referred to among statisticians as “Table 1”. This table displays means and standard errors, and counts and percentages for the variables in the sample, often stratified by some variable of interest (e.g. disease status, recruitment site, sex, etc.). While this table is extremely useful, the construction of it can be time consuming and, frankly, rather boring. The following macro quickly constructs this table of descriptive statistics for continuous and categorical variables stratified by some grouping variable.

## INTRODUCTION

Descriptive statistics are an essential component of data analysis. They are often reported in manuscripts and are frequently used in data cleaning and primary data analysis. Especially common in epidemiologic studies are descriptive statistics for each analysis variable, stratified by a grouping variable (e.g. treatment group, sex, etc.). To increase efficiency and accuracy in constructing tables, I have created a SAS® macro to automatically generate tables of descriptive statistics.

The SAS® macro described below computes statistics stratified by a grouping variable – mean and standard deviation for quantitative variables and count and percent for categorical variables. It also computes a p-value to test if the distribution of the analysis variable differs across levels of the grouping variable. For quantitative variables, this is an F-test from ANOVA and for categorical variables a Chi-Squared test. The output table contains variable name, p-value, and the given statistics stratified by grouping variable

## USING THE MACRO

To invoke the macro the user must supply the data set, grouping variable, and the names of quantitative and categorical variables:

```
%table1 (DSET=,  
         GRP=,  
         CATVARLIST=,  
         QUANTVARLIST=)
```

## MACRO PARAMETERS

The parameters that must be passed to the %table1 macro are as follows.

DSET =	dataset to be used for analysis which must include all analysis variables
GRP =	grouping variable by which to stratify descriptive statistics (must be categorical, can be coded as numeric or character)
CATVARLIST =	list of categorical variable names (must be categorical, and all coded as numeric or character)
QUANTVARLIST =	list of quantitative variable names (must be coded numeric)

## EXAMPLE

This example table of descriptive statistics uses the SAS® dataset of infant birth weights, *Sashelp.BWeight*, which contains 50,000 birth weight observations from 1997 from the National Center for Health Statistics. The variables in this data set are given in Figure 1. Note that each variable is coded numerically.

Variables in Creation Order				
#	Variable	Type	Len	Label
1	Weight	Num	8	Infant Birth Weight
2	Black	Num	8	Black Mother
3	Married	Num	8	Married Mother
4	Boy	Num	8	Baby Boy
5	MomAge	Num	8	Mother's Age
6	MomSmoke	Num	8	Smoking Mother
7	CigsPerDay	Num	8	Cigarettes Per Day
8	MomWtGain	Num	8	Mother's Pregnancy Weight Gain
9	Visit	Num	8	Prenatal Visit
10	MomEdLevel	Num	8	Mother's Education Level

**Figure 1. Variables from SasHELP.BWeight**

Using the dataset above, we will calculate descriptive statistics for six variables (married, boy, momsmoke, momedlevel, momage, and momwtgain) stratified by mother's race (black). The invocation of the %table1 macro is given below.

```
%table1 (DSET=sasHELP.bweight,
         GRP=black,
         CATVARLIST=married boy momsmoke momedlevel,
         QUANTVARLIST=momage momwtgain)
```

The %table1 macro calculates p-values from Chi-Squared Goodness of Fit tests for each of the categorical variables (boy, married, momedlevel, and momsmoke) and displays them along with the counts and percentages for each level of the aforementioned categorical variables as seen in Figure 2.

		Black Mother					
		0			1		
Variable	P		N	(%)	N	(%)	
boy	0.3702	0	20229	(48.32%)	3979	(48.87%)	
		1	21629	(51.67%)	4163	(51.12%)	
married	<.0001	0	9053	(21.62%)	5316	(65.29%)	
		1	32805	(78.37%)	2826	(34.70%)	
momedlevel	<.0001	0	14009	(33.46%)	3440	(42.25%)	
		1	10009	(23.91%)	2120	(26.03%)	
		2	11463	(27.38%)	986	(12.11%)	
		3	6377	(15.23%)	1596	(19.60%)	
momsmoke	<.0001	0	36117	(86.28%)	7350	(90.27%)	
		1	5741	(13.71%)	792	(9.72%)	

**Figure 2. Descriptive Statistics for Categorical Variables**

Similarly, Figure 3 shows the output table for the quantitative variables momage and momwtgain, this time calculating p-values from an F-test (in the two-group case this is identical to a two-sample t-test) and displaying them alongside means and standard deviations stratified by mother's race.

		Black Mother			
		0		1	
Variable	P	Mean	(SD)	Mean	(SD)
momage	<.0001	0.71	(5.66)	1.11	(5.82)
momwtgain	<.0001	1.01	(12.58)	0.86	(14.16)

**Figure 3. Descriptive Statistics for Quantitative Variables**

Using this SAS® macro, tables like the ones above can be generated very quickly without complicated or time-consuming programming. However, this macro can be easily altered to change the style of output and tests computed, if desired.

## CONTACT INFORMATION

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

Jenna C. Carlson  
University of Pittsburgh  
Pittsburgh, PA 15219  
jennaccarlson@gmail.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.

## APPENDIX – %TABLE1 MACRO CODE

```
proc format;
  picture pctfmt low-high= ' 009.99%' (prefix='(');
  picture sdfmt low-high= ' 009.99)' (prefix='(');
  picture meanfmt low-high= '009.99';
run;
%macro getCatVars;
%do I=1 %to %sysfunc(countw("&CATVARLIST"));
  data _NULL_;
    call symputx('CATVAR', "%qscan(&CATVARLIST, &I, ' '");
  run;
  ods select none;
  proc tabulate data=&DSET out=mytblc;
    class &GRP &CATVAR;
    table &CATVAR all='Total', &GRP=' '* (n='Freq' colpctn='(%)'*f=pctfmt.) /
box='Variables';
  run;
  proc freq data=&DSET;
    table &CATVAR*&GRP / nocum norow nopercnt chisq;
    ods output chisq = pvaluesc (where=(Statistic="Chi-Square"));
  run;
  ods select all;
  data pvaluesc;
    set pvaluesc;
    _TABLE_=1;
    keep _TABLE_ Prob;
  run;
  data mytblc2_&CATVAR;
    length var $32;
    merge mytblc pvaluesc;
    by _table_;
    var="&CATVAR";
    values=&CATVAR;
    if missing(&CATVAR) then delete;
  run;
```

```

%end;
%mend getCatVars;
%macro getQuantVars;
%do I=1 %to %sysfunc(countw("&QUANTVARLIST"));
  data _NULL_;
    call symputx('QUANTVAR', "%qscan(&QUANTVARLIST, &I, ' ' )");
  run;
  ods select none;
  proc tabulate data=&DSET out=mytblq(rename=(&QUANTVAR._mean=mymean
&QUANTVAR._std=mysd));
    class &GRP;
    var &QUANTVAR;
    table &QUANTVAR, &GRP='*' (mean='Mean' std='(SD)'*f=sdfmt.) /
box='Variables';
  run;
  proc glm data=&DSET;
    class &GRP;
    model &QUANTVAR = &GRP;
    ods output OverallAnova = pvaluesq (where=(Source="Model"));
  run;quit;
  ods select all;
  data pvaluesq;
    set pvaluesq;
    _TABLE_=1;
    keep _TABLE_ ProbF;
  run;
  data mytblq2_&QUANTVAR;
    length var $32;
    merge mytblq pvaluesq;
    by _table_ ;
    var="&QUANTVAR";

  run;
%end;
%mend getQuantVars;
%macro table1(DSET=,GRP=,CATVARLIST=,QUANTVARLIST=);
%getCatVars;
data mytblc_all;
  set mytblc2_;;
  keep var prob values &GRP N PctN_10;
run;
proc report data=mytblc_all(rename=(N=myN)) nofs missing;
  column var prob values &GRP,(myN PctN_10);
  define var / 'Variable' group;
  define prob / 'P' group;
  define values / '' group width=6;
  define &GRP / across width=3;
  define myN / 'N' width=3;
  define PctN_10 / format=pctfmt. '(%)' width=3;

run;
%getQuantVars;
data mytblq_all;
  set mytblq2_;;
run;
proc report data=mytblq_all nofs;
  column var probF &GRP,(mymean mysd);
  define var / 'Variable' group;
  define probF / 'P' group;
  define &GRP / across width=3;
  define mymean / 'Mean' format=meanfmt. width=3;
  define mysd / '(SD)' format=sdfmt. width=3;

run;
%mend table1;

```