

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

An Easy-to-use SAS® Macro for a Descriptive Statistics Table

USERS PROGRAM



An Easy-to-use SAS® Macro for a Descriptive Statistics Table

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath

Stanford University, Palo Alto, CA 94040

ABSTRACT

This paper introduces an easy-to-use macro to generate a descriptive statistics table. The table reports counts and percentages for categorical variables as well as means, standard deviations, medians, and quantiles for continuous variables. For variables with missing values, the table also includes the count and percentage missing. Customization options allow for the analysis of stratified data, the specification of variables’ output order, and user-defined formats. Additionally, this macro incorporates the SAS Output Delivery System (ODS) to automatically output a Rich Text Format (RTF) file, which can be further edited by a word processor for the purpose of publication.

YOU WILL FIND THE MACRO USEFUL IF

- You are tired of copying output from the Proc Freq or Proc Means procedures and pasting it into your tables.
- You need to produce summary tables repeatedly.
- You are spending a lot of your time generating the same summary table for different subpopulations.

SPECIFY MACRO OPTIONS

Options	Descriptions
%let yourdata=;	Name of SAS data set containing variables to be summarized.
%let output_data=;	Name of SAS data set containing summarized statistics, and the output RTF file.
%let decimal_max=;	Specify how many decimal points you need: 0, 1, 2, 3. This does not apply to count data.
%let varlist_cat=;	List of categorical variables. Leave empty if none.
%let varlist_cont=;	List of continuous variables. Leave empty if none.
%let formatsfolder=;	Location of SAS formats. Leave empty if none.
%let yourfolder=;	Location where your data set is saved. Leave empty for the SAS work library.
%let output_order=;	List of all UNIQUE variables from varlist_cat and varlist_cont in the order to be shown in the output table. Leave empty for default order, i.e., order entered in <i>varlist_cont</i> and <i>varlist_cat</i> .
%let group_by=;	Specify whether you want to output results by categories, e.g., gender. Leave empty to obtain statistics for the whole population. If a group-by variable is specified, a category for unformatted missing data can be created by user’s option. See the <i>group_by_missing</i> option.
%let group_by_missing=;	Specify whether or not output statistics for those observations with unformatted missingness in the group-by variable: 0, 1. Required if the <i>group_by</i> option is used. Value 1 creates a category for missing group-by variable. Change to 0 if not interested in reporting summary statistics for those missingness.

DATA EXAMPLE

```
data testdata;
  input ID sex $ age race BMI;
  datalines;
1      F      19      1      18.5
2      .      38      1      19.5
3      F      72      1      25.1
4      F      21      1      22.0
5      .      20      2      33.2
6      F      24      2      17.5
7      F      28      2      28.2
8      F      33      2      29.5
9      F      .       2      25.3
10     F      65      2      29.2
11     .      77      1      20.5
12     F      .       1      24.6
13     F      21      2      19.5
14     F      27      1      18.5
15     F      39      1      18.4
16     M      55      2      19.0
17     M      71      1      21.7
18     M      .       1      34.4
19     .      35      1      26.3
20     M      28      1      22.9
21     M      31      1      27.3
22     M      29      2      23.7
23     M      25      2      32.1
24     .      26      1      25.3
25     M      38      1      16.8
;
Run;
```

```
proc format;
  value $gender
    "F"="Female"
    "M"="Male"
  ;
  value race_cat
    1="White"
    2="Black"
  ;
  value age_cat
    low-40="<=40"
    40<-high=">40"
    .="Missing age"
  ;
  value $gender_new
    "F"="Female"
    "M"="Male"
    " "="Missing Gender"
  ;
run;
```

```
data testdata;
  set testdata;
  format sex gender. race race_cat.
          age age_cat.;
run;
```


An Easy-to-use SAS® Macro for a Descriptive Statistics Table

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath

Stanford University, Palo Alto, CA 94040

EXAMPLE 1: A WHOLE POPULATION

```
%let yourdata=testdata;           /*name of your SAS data set*/
%let output_data=test_summary1;    /*name of output SAS data set*/
%let formatsfolder=;               /*location of your SAS formats*/
%let yourfolder=;                  /*location of your SAS data set*/
%let decimal_max=1;                /*desired number of decimal points*/
%let varlist_cat = age race sex;    /*list of categorical variables*/
%let varlist_cont = age;            /*list of continuous variables*/
%let output_order = age race sex;   /*output order of all UNIQUE variables*/
%let group_by=;                    /*name of stratification variable*/
%let group_by_missing=;             /*whether to remove observations missing
                                     the stratification variable.*/
%Table_summary;                  /*call the macros*/
```

Variable	Description	Type of statistic	All (n=25)
AGE	Continuous	mean and std	37.4 (18.3)
AGE	Continuous	median and IQR	30.0 (25.0, 39.0)
AGE	Continuous	min and max	(19.0, 77.0)
AGE	Continuous: missing	count and percent	3 (12.0%)
AGE	Missing age	count and percent	3 (12.0%)
AGE	<=40	count and percent	17 (68.0%)
AGE	>40	count and percent	5 (20.0%)
RACE	White	count and percent	15 (60.0%)
RACE	Black	count and percent	10 (40.0%)
SEX	Missing	count and percent	5 (20.0%)
SEX	Female	count and percent	12 (48.0%)
SEX	Male	count and percent	8 (32.0%)

EXAMPLE 2: A STRATIFIED POPULATION

```
%let varlist_cat = race;           /*list of categorical variables*/
%let varlist_cont = age;            /*list of continuous variables*/
%let output_order = age race;       /*output order of all UNIQUE variables*/
%let group_by=sex;                  /*name of stratification variable*/
%let group_by_missing=1;            /*keep observations missing the
                                     stratification variable.*/
%Table_summary;                  /*call the macros*/
```

Variable	Description	Type of statistic	ALL (n=25)	FEMALE (n=12)	MALE (n=8)	UNSPECIFIED MISSING (n=5)
AGE	Continuous	mean and std	37.4 (18.3)	34.9 (18.8)	39.6 (17.1)	39.2 (22.3)
AGE	Continuous	median and IQR	30.0 (25.0, 39.0)	27.5 (21.0, 39.0)	31.0 (28.0, 55.0)	35.0 (26.0, 38.0)
AGE	Continuous	min and max	(19.0, 77.0)	(19.0, 72.0)	(25.0, 71.0)	(20.0, 77.0)
AGE	Continuous: missing	count and percent	3 (12.0%)	2 (16.7%)	1 (12.5%)	
RACE: Black	Black	count and percent	10 (40.0%)	6 (50.0%)	3 (37.5%)	1 (20.0%)
RACE: White	White	count and percent	15 (60.0%)	6 (50.0%)	5 (62.5%)	4 (80.0%)

```
%let group_by_missing=0;           /*remove observations missing the
                                     stratification variable.*/
```

Variable	Description	Type of statistic	ALL (n=20)	FEMALE (n=12)	MALE (n=8)
AGE	Continuous	mean and std	36.8 (17.7)	34.9 (18.8)	39.6 (17.1)
AGE	Continuous	median and IQR	29.0 (25.0, 39.0)	27.5 (21.0, 39.0)	31.0 (28.0, 55.0)
AGE	Continuous	min and max	(19.0, 72.0)	(19.0, 72.0)	(25.0, 71.0)
AGE	Continuous: missing	count and percent	3 (15.0%)	2 (16.7%)	1 (12.5%)
RACE: Black	Black	count and percent	9 (45.0%)	6 (50.0%)	3 (37.5%)
RACE: White	White	count and percent	11 (55.0%)	6 (50.0%)	5 (62.5%)



An Easy-to-use SAS® Macro for a Descriptive Statistics Table

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath

Stanford University, Palo Alto, CA 94040

Contact information

Your comments and questions are valued and encouraged.

Contact the author at:

Yuanchao Zheng

Division of Nephrology, Department of Medicine

Stanford University School of Medicine

1070 Arastradero Rd., Suite 100

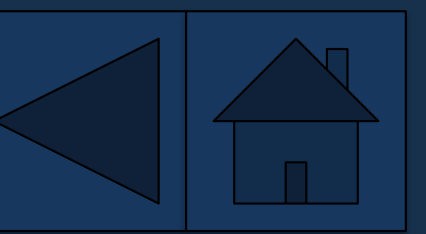
Palo Alto, CA 94304

Email: yuanchao@stanford.edu

Where to download this Macro?

Free download at

<https://github.com/ggzheng/SAS2017>



SAS[®] GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL