An Easy-to-use SAS® Macro for a Descriptive Statistics Table
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 formats_folder=; 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 varlist_cont and varlist_cat.
%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 group_by_missing 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 group_by 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 < missing age
  <40 high=40
  ;
  run;

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

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.
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

```sas
%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*/
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type of statistic</th>
<th>ALL (n=25)</th>
<th>FEMALE (n=12)</th>
<th>MALE (n=8)</th>
<th>UNSPECIFIED MISSING (n=5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>Continuous</td>
<td>mean and std</td>
<td>37.4 (18.3)</td>
<td>34.9 (18.8)</td>
<td>39.6 (17.1)</td>
<td>39.2 (22.3)</td>
</tr>
<tr>
<td>AGE</td>
<td>Continuous</td>
<td>median and IQR</td>
<td>30.0 (25.0, 39.0)</td>
<td>27.5 (21.0, 39.0)</td>
<td>31.0 (28.0, 55.0)</td>
<td>35.0 (26.0, 38.0)</td>
</tr>
<tr>
<td>AGE</td>
<td>Continuous</td>
<td>min and max</td>
<td>19.0 (72.0)</td>
<td>19.0 (70.0)</td>
<td>25.0 (70.0)</td>
<td>20.0 (77.0)</td>
</tr>
<tr>
<td>AGE</td>
<td>Continuous: missing</td>
<td>count and percent</td>
<td>3 (12.0%)</td>
<td>2 (16.7%)</td>
<td>1 (12.5%)</td>
<td>1 (20.0%)</td>
</tr>
<tr>
<td>AGE</td>
<td>Less than 40</td>
<td>count and percent</td>
<td>5 (20.0%)</td>
<td>4 (50.0%)</td>
<td>3 (37.5%)</td>
<td>1 (20.0%)</td>
</tr>
<tr>
<td>RACE</td>
<td>White</td>
<td>count and percent</td>
<td>15 (60.0%)</td>
<td>6 (50.0%)</td>
<td>5 (62.5%)</td>
<td>4 (80.0%)</td>
</tr>
<tr>
<td>RACE</td>
<td>Black</td>
<td>count and percent</td>
<td>10 (40.0%)</td>
<td>6 (50.0%)</td>
<td>3 (37.5%)</td>
<td>2 (40.0%)</td>
</tr>
<tr>
<td>SEX</td>
<td>Missing</td>
<td>count and percent</td>
<td>5 (20.0%)</td>
<td>6 (50.0%)</td>
<td>3 (37.5%)</td>
<td>1 (20.0%)</td>
</tr>
<tr>
<td>SEX</td>
<td>Female</td>
<td>count and percent</td>
<td>12 (48.0%)</td>
<td>6 (50.0%)</td>
<td>5 (62.5%)</td>
<td>4 (80.0%)</td>
</tr>
<tr>
<td>SEX</td>
<td>Male</td>
<td>count and percent</td>
<td>8 (32.0%)</td>
<td>6 (50.0%)</td>
<td>3 (37.5%)</td>
<td>2 (40.0%)</td>
</tr>
</tbody>
</table>

EXAMPLE 2: A STRATIFIED POPULATION

```sas
%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*/
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type of statistic</th>
<th>ALL (n=20)</th>
<th>FEMALE (n=12)</th>
<th>MALE (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>Continuous</td>
<td>mean and std</td>
<td>36.8 (17.7)</td>
<td>34.9 (18.8)</td>
<td>39.6 (17.1)</td>
</tr>
<tr>
<td>AGE</td>
<td>Continuous</td>
<td>median and IQR</td>
<td>29.0 (25.0, 39.0)</td>
<td>27.5 (21.0, 39.0)</td>
<td>31.0 (28.0, 55.0)</td>
</tr>
<tr>
<td>AGE</td>
<td>Continuous</td>
<td>min and max</td>
<td>19.0 (72.0)</td>
<td>19.0 (70.0)</td>
<td>25.0 (70.0)</td>
</tr>
<tr>
<td>AGE</td>
<td>Continuous: missing</td>
<td>count and percent</td>
<td>3 (15.0%)</td>
<td>2 (16.7%)</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>RACE</td>
<td>Black</td>
<td>count and percent</td>
<td>9 (45.0%)</td>
<td>6 (50.0%)</td>
<td>3 (37.5%)</td>
</tr>
<tr>
<td>RACE</td>
<td>White</td>
<td>count and percent</td>
<td>11 (55.0%)</td>
<td>6 (50.0%)</td>
<td>5 (62.5%)</td>
</tr>
</tbody>
</table>
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 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.