An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Abstract

In this poster, we introduce **an easy-to-use macro that allows SAS users to report counts and percentages for categorical variables as well as means, standard deviations, medians, and quantiles for continuous variables with an option to include trend p-values.** For variables with missing values, the table also includes the count and percentage missing.

We will show how we can easily customize the code to generate a table stratified by a group variable, as well as, show how we can specify the order of the variables shown in the output. It automatically outputs a Rich Text Format (RTF) file, which can be further edited by a word processor.

You will find the macro useful if

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

Output statistics by group variable (severity) including trend p-values

Variable	Description	Type of statistic	ALL (n=80)	1 (n=22)	2 (n=31)	3 (n=27)	Trend P-value
AGE	Continuous	mean and std	46.6 (22.0)	44.3 (22.3)	44.0 (22.2)	51.4 (21.6)	0.2599
AGE	Continuous	median and IQR	39.0 (27.0, 69.0)	38.0 (27.0, 69.0)	33.0 (24.0, 65.0)	46.0 (31.0, 72.0)	
AGE	Continuous	min and max	(19.0, 90.0)	(19.0, 90.0)	(19.0, 88.0)	(21.0, 89.0)	
AGE	18-24	count and percent	14 (17.5%)	5 (22.7%)	8 (25.8%)	1 (3.7%)	0.0660
AGE	25-64	count and percent	42 (52.5%)	11 (50.0%)	15 (48.4%)	16 (59.3%)	0.4955
AGE	65+	count and percent	24 (30.0%)	6 (27.3%)	8 (25.8%)	10 (37.0%)	0.4343
SEX	Female	count and percent	38 (47.5%)	15 (68.2%)	15 (48.4%)	8 (29.6%)	0.0021
SEX	Male	count and percent	35 (43.8%)	4 (18.2%)	14 (45.2%)	17 (63.0%)	0.0021
SEX	Missing	count and percent	7 (8.8%)	3 (13.6%)	2 (6.5%)	2 (7.4%)	
BMI	Continuous	mean and std	25.6 (5.7)	23.5 (5.9)	25.4 (5.4)	27.6 (5.3)	0.0126
BMI	Continuous	median and IQR	25.3 (20.5, 29.3)	22.5 (18.5, 25.1)	26.0 (19.7, 29.2)	27.3 (23.7, 30.3)	
BMI	Continuous	min and max	(16.2, 39.5)	(16.2, 39.3)	(17.5, 34.8)	(16.8, 39.5)	

NOTE: there are 0 observations deleted due to missing values in the variable severity

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Trend Test

Continuous variables

Model-based p-values from simple linear regression. Test for the linear trend in the mean of continuous variable across levels of a single factor or covariate.

Use Proc GLM procedure with the contrast statement as appropriate.

Reading

- Page 84. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models Eric Vittinghoff - Springer – 2012
- http://support.sas.com/kb/22/912.html

Categorical variables

Cochran-Armitage Test for trend in binomial proportions across levels of a single factor or covariate. This test is appropriate for a two-way table where one variable has two levels and the other variable is ordinal. The binomial proportion is computed as the proportion of observations in the first row.

Use Proc Freq procedure with the TREND option.

Reading

- http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_freq_a0000000663.htm

Data Example

```
Data rawdata;
 input ID sex $ age severity 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   3   29.5
9    F   22   3   25.3

 . . . . . .
 . . . . . .

75   F   45   3   33.2
76   M   27   2   18.5
77   M   22   2   24.0
78   M   47   3   29.1
79   F   79   3   25.4
80   F   89   3   39.0
;
Run;
```

```
Proc format;
 value $gender
 "F" = "Female"
 "M" = "Male"
 ;

 value age_cat
 low-<25 = "18-24"
 25-<65  = "25-64"
 65-high = "65+"
 ;

Run;

Data testdata;
 set rawdata;
 format sex gender. age age_cat.;
Run;
```

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

How to 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 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 to 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.
%let trendtest=;	Specify whether to output trend p values: change to 1 if needed. Only works when group_by option is specified, group_by_missing=0 and stratification variable contains more than 2 categories.

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Example 1: Descriptive table for the whole population

Variable	Description	Type of statistic	All (n=80)
AGE	Continuous	mean and std	46.6 (22.0)
AGE	Continuous	median and IQR	39.0 (27.0, 69.0)
AGE	Continuous	min and max	(19.0, 90.0)
AGE	18-24	count and percent	14 (17.5%)
AGE	25-64	count and percent	42 (52.5%)
AGE	65+	count and percent	24 (30.0%)
SEVERITY	1	count and percent	22 (27.5%)
SEVERITY	2	count and percent	31 (38.8%)
SEVERITY	3	count and percent	27 (33.8%)
SEX	Missing	count and percent	7 (8.8%)
SEX	Female	count and percent	38 (47.5%)
SEX	Male	count and percent	35 (43.8%)

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Code for Example 1:
Descriptive table for the 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 severity sex;      /*list of categorical variables*/
%let varlist_cont=age;                  /*list of continuous variables*/
%let output_order=age severity 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.*/
%let trendtest=;                        /*whether to report trend p values*/

%Table_summary;                         /*call the macros*/
```

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Example 2: Output statistics by group variable (severity)

Variable	Description	Type of statistic	ALL (n=80)	1 (n=22)	2 (n=31)	3 (n=27)
AGE	Continuous	mean and std	46.6 (22.0)	44.3 (22.3)	44.0 (22.2)	51.4 (21.6)
AGE	Continuous	median and IQR	39.0 (27.0, 69.0)	38.0 (27.0, 69.0)	33.0 (24.0, 65.0)	46.0 (31.0, 72.0)
AGE	Continuous	min and max	(19.0, 90.0)	(19.0, 90.0)	(19.0, 88.0)	(21.0, 89.0)
AGE	18-24	count and percent	14 (17.5%)	5 (22.7%)	8 (25.8%)	1 (3.7%)
AGE	25-64	count and percent	42 (52.5%)	11 (50.0%)	15 (48.4%)	16 (59.3%)
AGE	65+	count and percent	24 (30.0%)	6 (27.3%)	8 (25.8%)	10 (37.0%)
SEX	Missing	count and percent	7 (8.8%)	3 (13.6%)	2 (6.5%)	2 (7.4%)
SEX	Female	count and percent	38 (47.5%)	15 (68.2%)	15 (48.4%)	8 (29.6%)
SEX	Male	count and percent	35 (43.8%)	4 (18.2%)	14 (45.2%)	17 (63.0%)
BMI	Continuous	mean and std	25.6 (5.7)	23.5 (5.9)	25.4 (5.4)	27.6 (5.3)
BMI	Continuous	median and IQR	25.3 (20.5, 29.3)	22.5 (18.5, 25.1)	26.0 (19.7, 29.2)	27.3 (23.7, 30.3)
BMI	Continuous	min and max	(16.2, 39.5)	(16.2, 39.3)	(17.5, 34.8)	(16.8, 39.5)

NOTE: there are 0 observations deleted due to missing values in the variable severity

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Code for Example 2:
Output statistics by group variable (severity)

```
%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 sex;               /*list of categorical variables*/
%let varlist_cont=age bmi;              /*list of continuous variables*/
%let output_order=age sex bmi;          /*output order of all UNIQUE variables*/

%let group_by=severity;                 /*name of stratification variable*/
%let group_by_missing=0;                /*whether to remove observations missing the stratification variable.*/
%let trendtest=;                        /*whether to report trend p values*/

%Table_summary;                         /*call the macros*/
```

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Example 3: Output statistics by group variable (severity) including trend p-values

Variable	Description	Type of statistic	ALL (n=80)	1 (n=22)	2 (n=31)	3 (n=27)	Trend P-value
AGE	Continuous	mean and std	46.6 (22.0)	44.3 (22.3)	44.0 (22.2)	51.4 (21.6)	0.2599
AGE	Continuous	median and IQR	39.0 (27.0, 69.0)	38.0 (27.0, 69.0)	33.0 (24.0, 65.0)	46.0 (31.0, 72.0)	
AGE	Continuous	min and max	(19.0, 90.0)	(19.0, 90.0)	(19.0, 88.0)	(21.0, 89.0)	
AGE	18-24	count and percent	14 (17.5%)	5 (22.7%)	8 (25.8%)	1 (3.7%)	0.0660
AGE	25-64	count and percent	42 (52.5%)	11 (50.0%)	15 (48.4%)	16 (59.3%)	0.4955
AGE	65+	count and percent	24 (30.0%)	6 (27.3%)	8 (25.8%)	10 (37.0%)	0.4343
SEX	Female	count and percent	38 (47.5%)	15 (68.2%)	15 (48.4%)	8 (29.6%)	0.0021
SEX	Male	count and percent	35 (43.8%)	4 (18.2%)	14 (45.2%)	17 (63.0%)	0.0021
SEX	Missing	count and percent	7 (8.8%)	3 (13.6%)	2 (6.5%)	2 (7.4%)	
BMI	Continuous	mean and std	25.6 (5.7)	23.5 (5.9)	25.4 (5.4)	27.6 (5.3)	0.0126
BMI	Continuous	median and IQR	25.3 (20.5, 29.3)	22.5 (18.5, 25.1)	26.0 (19.7, 29.2)	27.3 (23.7, 30.3)	
BMI	Continuous	min and max	(16.2, 39.5)	(16.2, 39.3)	(17.5, 34.8)	(16.8, 39.5)	

NOTE: there are 0 observations deleted due to missing values in the variable severity

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

Code for Example 3:
Output statistics by group variable (severity) including trend p-values

```
%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 sex;               /*list of categorical variables*/
%let varlist_cont=age bmi;              /*list of continuous variables*/
%let output_order=age sex bmi;          /*output order of all UNIQUE variables*/

%let group_by=severity;                 /*name of stratification variable*/
%let group_by_missing=0;                /*whether to remove observations missing the stratification variable.*/
%let trendtest=1;                       /*whether to report trend p values*/

%Table_summary;                         /*call the macros*/
```

An Easy-to-Use SAS® Macro for a Descriptive Statistics Table with P-Values

Yuanchao Zheng, Jin Long, Maria E. Montez-Rath
Stanford University, Palo Alto, CA

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 Marco

https://github.com/ggzheng/CSP2018

https://github.com/ggzheng/SAS2017