

One Macro to Produce Descriptive Statistic Summary Tables with P-Values

Rajaram Venkatesan and Harinarayan Gopichandran, Cognizant Technology Solutions

ABSTRACT

In clinical trial reporting, the most popular type of tables is those that have descriptive statistics (n, mean, SD, median, min and max, CI, and p-values) or tables having a frequency (%) count and descriptive statistics of categorical and continuous variables. These are the bread and butter of reporting. However, producing these tables is simple yet trivial, and sometimes cumbersome and time-consuming, as many variables and many conditions might be requested.

The solution is to create a simple macro and easy to understand macro, which allows the user to develop and produce descriptive summary tables within minutes. This can be used to produce or validate most safety tables without any problems. This allows users to create many types of tables (demographics and baseline characteristics, laboratory, vital signs, and ECG data) with minimal effort. It also means that when statisticians would like to change the table afterward, it can be done with the minimum of effort. This will not only save a lot of time but also improve quality.

INTRODUCTION

In the pharmaceutical industry, most common tables are descriptive summary tables, which consist of categorical variables (frequency counts and percentages) and percentages and/or continuous variables (n, means, standard derivation, medians, min and max, confidence limits, p-values, etc.). Developing these tables sometime cumbersome and time consuming, as many variables, formatting, decimal alignment and many conditions required. Hence, a standard macro was developed to minimize time, increase quality, easy to use, multiple format options, and automatic decimal alignment.

In this paper, we introduce our standard and simple SAS® macro functionality, parameters and usage.

ROUNDING

By default, descriptive statistics mean, median, the confidence interval (CI), quartiles and geometric mean will be rounded to 1 additional decimal place of original data. Standard deviation, standard error of the mean and coefficient of variation (CV) will be displayed to the 2 decimal places of original data. The min and max should be displayed to the same number of decimal places as the original data.

P-values will be reported to the fourth (4th) decimal place and presented in the format of "0.xxxx" or "<0.0001".

MACRO CALL AND ITS PARAMETERS

To invoke the macro, issue the macro call:

```
%ASUMSTAT (INDSNM=, POPDSNM=, OUTDSNM=, DISPLBL=, ANLYZVRS=, COLMVAR=,  
           GRPCOLM=, INTOTCAT=, BYVARBL=, PGVARBL=, MISNOTN=, DISFMT=,  
           UNIQSUBC=, STACKSPC=, SPCBVAL=, ALPHA=, INALLCAT=, SMRYLBL=,  
           SETDECN=, TOTCLMYN=, BRKAVRS=, STATST=);
```

Where

INDSNM The input dataset name. If you want to subset dataset you can also pass required condition as it is used in WHERE statement, any valid SAS® syntax is acceptable. Dataset and where condition should be separated by \.

General format

INDSNM = ADaM.ADSLITTFL='Y'

Required: Yes

Default: No default value.

POPDSNM The population/denominator dataset name. If you want to subset dataset you can also pass required condition as it is used in WHERE statement, any valid SAS® syntax is acceptable. Dataset and where condition should be separated by \. This dataset will be used for percentage calculation.

Required: No

Default: No default value.

OUTDSNM The output dataset name.

Required: Yes

Default: No default value.

DISPLBL Display label for each variable separated by pipe (|). If you do not want to specify display label in the output you can pass "NULL" or "_NULL_"

Required: Yes

Default: No default value.

ANLYZVRS List of analysis variables. This is no restriction to maximum number of variables. Variables should be separated by pipe (|). Each variable has special format to indicate variable, format, and type of data (i.e., continuous or categorical) and these will be separated by backslash (\).

General format

ANLYZVRS=VARIABLE\FORMAT\TYPE|VARIABLE\FORMAT\TYPE

Where

Variable=Name of the variable

Format=SAS® acceptable format

Type=Continuous (CON) or Categorical (CAT)

Example: ANLYZVRS=AGE\8.\CON|SEX\SEX.\CAT

For categorical variables, 'n' and percent will be displayed. For continuous variables, statistics which are requested in DISFMT will be displayed. Display order in the table will be based on the variable list.

Required: Yes

Default: No default value.

COLMVAR Specify column (usually treatment) variable. You can also specify to transpose descriptive statistics and apply format for variables and these will be separated by

backslash (\). Only one variable will be allowed.

General format

COLMVAR=VARIABLE\YES or NO\FORMAT

Where

Variable=Name of the variable

Format=SAS® acceptable format (Optional)

Example: COLMVAR =TRTPN\NO\TRTN.

Required: Yes

Default: TRTPN\N

Refer: [Example 2](#)

GRPCOLM

Second stratification variable, placed above the COLMVAR variable. One variable can be specified on the GRPCOLM parameter. GRPCOLM may or may not format. Format will be separated by backslash (\).

General format GRPCOLM =VARIABLE\FORMAT

Where

Variable=Name of the variable

Format=SAS® acceptable format (Optional)

Example: GRPCOLM=TRTPN\TRTN.

Required: No

Default: No default value.

Refer: [Example 5](#)

PGVARBL

Name of variable(s) used to specify Page By on report. More than one variable can be specified with spaces between. The PGVARBL variable(s) appears on a BY line at the top of each page of the table.

Required: No

Default: No default value.

Refer: [Example 7](#)

BYVARBL

Name of variable(s) used to specify By variable on report. More than one variable can be specified with pipe (|) between. Each variable optionally combined with its format by backslash (\). The BYVARBL variable(s) appears first column in the table. BYVARBL can be clubbed with statistics and/or display label.

General format BYVARBL=VARIABLE\FORMAT|VARIABLE\FORMAT

Where

Variable=Name of the variable

Format=SAS® acceptable format (Optional)

Example: BYVARBL=AVISTIN\VISMAP.

Required: No

Default: No default value.

Refer: [Example 7](#)

DISFMT

Control display format in report for continuous and categorical separated by pipe (|). Continuous statistics order will be based on selected position will be arranged.

General format DISFMT=Continuous|Categorical

Following options are available for Continuous

```
*****
INPUT | DISPLAY LABEL | EXAMPLE OUTPUT
*****
1 | n | 462
2 | Mean | 145.0
3 | SD | 16.40
4 | SE | 0.76
*****
MEAN WITH STANDARD DEVIATION & STANDARD ERROR
*****
5 | Mean (SD) | 145.0 (16.40)
6 | Mean (SE) | 145.0 (0.76)
7 | Mean ± SD | 145.0 ± 16.40
8 | Mean ± SE | 145.0 ± 0.76
*****
MEDIAN & MEDIAN WITH RANGE & INTER QUARTILE RANGE
*****
9 | Median | 142.8
10 | Median (Min, Max) | 145.0 (103, 190)
11 | Median (Q1, Q3) | 145.0 (132.7, 160.0)
*****
RANGE
*****
12 | Min | 103
13 | Max | 190
14 | Min, Max | 103, 190
15 | (Min, Max) | (103, 190)
```

```

16 | Min to Max          | 103 to 190
*****
                                INTER QUARTILE RANGE
*****
17 | Q1                   | 132.7
18 | Q3                   | 160.0
19 | Q1, Q3              | 132.7, 160.0
20 | (Q1, Q3)            | (132.7, 160.0)
21 | Q1 to Q3           | 132.7 to 160.0
*****
                                COEFFICIENT OF VARIATION & GEOMETRIC MEAN (GM)
*****
22 | %CV                  | 11.30
23 | Geometric Mean      | 5.00
24 | %CV for GM          | 11.00
*****
                                CONFIDENCE INTERVAL FOR MEAN & GM
*****
25 | Lower                | 143.8
26 | Upper                | 146.3
27 | Lower, Upper        | 143.8, 146.3
28 | (Lower, Upper)     | (143.8, 146.3)
29 | Lower to Upper     | 143.8 to 146.3
30 | Lower for GM        | 142.9
31 | Upper for GM        | 145.4
32 | Lower, Upper for GM | 142.9, 145.4
33 | (Lower, Upper) for GM | (142.9, 145.4)
34 | Lower to Upper for GM | 142.9 to 145.4
*****

```

Following options are available for Categorical

- 1 = displays count only
- 2 = displays count and percentage as XX (XX.X)
- 3 = displays count and percentage with percentage symbol as XX (XX.X%)
- 4 = displays count, population and percentage as XX/XX (XX.X)
- 5 = displays count, population and percentage with percentage symbol as XX/XX (XX.X%)

Display of percentages when count=0. The default condition displays percentages for 0 counts.

- X = displays count of 0, no percentages.
- X\0 = displays count of 0, and 0 percentage.

Required: Yes

Default: 1 7 9 14|2\0

Refer: [Example 6](#)

SMRYLBL Statistics will be displayed with the default labels. User can override default values. Maximum allowed character is 100.

Each statistic label can be relabeled

SMRYLBL=STAT1=LABEL1|STAT2= LABEL2| ... |STATy= LABELy.

Where

STATn is an output statistics key words and its default values:

N = Number of observation (Default: N=n)

MEAN = Mean (Default: MEAN=Mean)

MEDIAN = Median (Default: MEDIAN=Median)

STD or SD = Standard deviation (Default: STD=SD)

STDERR or SE = Standard error (Default: STDERR=SE)

Q1 = 25% Quartile (Default: Q1=Q1)

Q3 = 75% Quartile (Default: Q3=Q3)

GMEAN OR GM = Geometric mean (Default: GMEAN=Geometric Mean)

CV = Co-efficient of variance (Default: CV=%CV)

GCV = Co-efficient of variance log based (Default: GCV=%CV for GM)

LCL = Lower confidence limit (Default: LCL=Lower)

UCL = Upper confidence limit (Default: UCL=Upper)

GLCL = Lower confidence limit log based (Default: GLCL=Lower)

GUCL = Upper confidence limit log based (Default: GUCL=Upper)

MIN = Minimum (Default: MIN=Min)

MAX = Maximum (Default: MAX=Max)

Example:

SMRYLBL = MIN=Minimum|MAX=Maximum.

If output format is Min, Max and you want relabel to Range then specify as follow

SMRYLBL = MIN=Range|MAX=Range.

Required: No

Default: Refer above

SETDECN Set decimal places for continuous variables. Specify a decimal "offset" for specific statistics or standard option (STDD, STDF, ASISD, and ASISF). In general, statistics are assigned that number of places as per data and rounding rule, with the exception of P-value statistics.

The SETDECN parameter uses the following syntax:

SETDECN=<Standard option>\STAT1=N|STAT2=M| ... |STATy=X

Where

Standard option:

STDD = Standard [Rounding](#) based on dataset

STDF = Standard [Rounding](#) based on format specified in ANLYZVRS

ASISD = Maximum number of decimals in dataset for all statistics

ASISF= Same decimal places in format specified in ANLYZVRS for all statistics

STATn is an output statistics key words:

MEAN = Mean

MEDIAN = Median

STD or SD = Standard deviation

STDERR or SE = Standard error

Q1 or Q3 or Q1Q3 = Quartiles

GMEAN or GM = Geometric mean

CV = Co-efficient of variance

GCV = Co-efficient of variance log based

CI or CL or CLM = Confidence limit

MINMAX or MIN or MAX = Minimum and maximum

PCT or PERCENT = Percentage

N, M...X are decimal offsets. The number of decimal places used for the statistic is the maximum number of decimals in dataset or format plus the offset.

Example:

SETDECN=STD=2|CV=3 or SETDECN=STDD\STD=2|CV=3

If maximum number of decimals in dataset were 1, then the STD would display three places and the CV would display four.

Required: No

Default: STDD

STACKSPC

If ANLYZVRS to be stacked with statistics or summary category then specify number spaces to stack. If you are going to report using ODS RTF or TAGSETS.RTF then specify RTF with \ between. If you specify RTF then special RTF Tag will be added appropriately.

General format STACKSPC=Number of spaces\Options\< RTF>

There are three options to stack

STAT = Display label stack with summary statistics

BYVAR = By variable stack with display label

ALL = Display label and by variable stack with summary statistics

Required: No

Default: Number of spaces\STAT

Refer: [Example 4](#)

SPCBVAL Include space before summary values to adjust counts.

Required: No

Default: 0

MISNOTN Summary statistics missing value notation for continuous variables.

Required: No

Default: NE

STATST Request p-value for variables in ANLYZVRS. Available tests are Student t-test, Wilcoxon, Kruskal- Wallis, Analysis of Variance, Analysis of Co-variance, Chi-square, Fisher's exact test and Cochran–Mantel–Haenszel.

General format

STATST=Test-Test Option\Scores-Identifier-Covariates

Test:

TTEST to request Student t-test

WILCOXON to request Wilcoxon test

KRUSKAL to request Kruskal- Wallis test

ANOVA to request Analysis of Variance

ANCOVA to request Analysis of Co-variance

CHISQ to request Chi-square test

FISHER or EXACT to request Fisher's exact test

CMH to request Cochran–Mantel–Haenszel test

Test Option:

If t-test is requested then following options are applicable to test

- a. DIFF (To test the difference of means – Default)
- b. RATIO (To test the ratio of means)

If Wilcoxon/Kruskal- Wallis is requested then following options are applicable for p-value display

- a. PL_WIL (Left sided p-value - normal assumption)
- b. PR_WIL (Right sided p-value - normal assumption)
- c. P2_WIL (2-sided p-value - normal assumption)

- d. PTL_WIL(Left sided p-value - t-assumption)
- e. PTR_WIL (Right sided p-value - t-assumption)
- f. PT2_WIL (2-sided p-value - t-assumption)
- g. P_KW (Kruskal-Wallis – Default)

If AN(C)OVA is requested then following options are applicable for p-value display

Class variables (default: COLMVAR)

If chi-square is requested then following options are applicable for p-value display

- a. P_PCHI (Pearson chi-square for 2-way tables – Default)
- b. P_LRCHI (Likelihood-ratio chi-square)
- c. P_AJCHI (Continuity-adjusted chi-square for 2-way tables)
- d. P_MHCHI (Mantel-Haenszel chi-square)

If Fisher’s exact test is requested then following options are applicable for p-value display

- a. XPL_FISH (Left sided p-value for 2-way tables)
- b. XPR_FISH (Right sided p-value for 2-way tables)
- c. XP2_FISH (2-sided p-value for 2-way and larger than 2-way tables – Default)

If Cochran–Mantel–Haenszel test is requested then following options are applicable for p-value display and specify scores option

- a. P_CMHCOR (Nonzero correlation)
- b. P_CMHRMS (Row mean scores difference – Default if MODRIDIT)
- c. P_CMHGA (General association – Default if Other)

Scores: Applicable to CHM only. TABLE, RANK, RIDIT or MODRIDIT (Default: TABLE)

Identifier: Any identification alphanumeric with or without special characters

Covariates: Applicable to ANCOVA and CMH

Required: No

Default: Refer above

Restriction: STATST is not applicable if GRPCOLM is present

Refer: [Example 3](#)

UNIQSUBC Patient variable name to calculate unique subject/patient count. If patient variable is specified unique count will be used while calculating numerator and denominator count.

Required: No

Default: No default value.

INTOTCAT Optional display of a TOTAL value for categorical variables

The SETDECN parameter uses the following syntax:

INTOTCAT=T or B\Percentage Y or N\<Label>

Where

T = Total row is displayed above the all categories.
B = Total row is displayed below the all categories.
Y = Total row with percentage is displayed (Default: N).
Label = Total row display label (Default: Total).

Required: No

Default: Refer above

Refer: [Example 4](#)

TOTCLMYN Controls creation and display of a TOTAL column on the final table. The TOTAL group combines data from all values of the COLMVAR parameter. Set TOTCLMYN=N to suppress display of a TOTAL group.

Required: No

Default: N

INALLCAT Request to include all categories in the format event the category is not present in the dataset. Set INALLCAT=Y to display all categories.

Required: No

Default: N

Restriction: If format is not exist then this will set to INALLCAT=N.

Refer: [Example 4](#)

ALPHA Set the level of confidence limits. The value of α must be between 0 and 1. A confidence level of α produces $100(1-\alpha)\%$ confidence limits. The default of ALPHA=0.05 produces 95% confidence limits.

Required: No

Default: 0.05

BRKAVRS Option to generate blank line between data rows. When BRKAVRS=Y, the report generates a blank line when the value for the first term variable changes. To suppress the blank lines, set BRKAVRS=N.

Required: No

Default: Yes

EXAMPLES

Example 1

This example uses default options.

```
%asumstat(indsnm=adam.adsl\trt01pn notin (1,0),  
  outdsnm=tfin,  
  displbl=Age (years)|Sex - n (%),  
  anlyzvrs=age\8.\con|sex\sex.\cat);
```

Demographic Parameter	Statistics	Active (N=XXX)	Placebo (N=XXX)
Age (years)	n	75	77
	Mean ± SD	76.8 ± 5.92	76.8 ± 4.53
	Median	75.0	76.0
	Min, Max	70, 98	70, 90
Sex – n (%)	Male	23 (30.7)	29 (37.7)
	Female	52 (69.3)	48 (62.3)

Example 2

This example uses statistics are transposed in COLMVAR.

```
%asumstat(indsnm=adam.adsl\trt01pn notin (1,0),
           outdsnm=tfin,
           colmvar=trt01pn\y\trtn.,
           displbl=Age (Years)|Sex,
           anlyzvrs=age\8.\con|sex\sex.\cat);
```

	Active				Placebo			
	n	Mean ± SD	Median	Min, Max	n	Mean ± SD	Median	Min, Max
Age (Years)	75	76.8 ± 5.92	75.0	70, 98	77	76.8 ± 4.53	76.0	70, 90
Sex								
Female	23 (30.7)				29 (37.7)			
Male	52 (69.3)				48 (62.3)			

Example 3

This example uses STATST option to request t-test and chi-square p-values.

```
%asumstat(indsnm=adam.adsl\trt01pn notin (1,0),
           outdsnm=tfin,
           displbl=Age (years)Sex - n (%),
           anlyzvrs=age\8.\con|sex\sex.\cat,
           stattst=ttest-diff- [a]|chisq-p_pchi- [b]);
```

Demographic Parameter	Statistics	Active (N=XXX)	Placebo (N=XXX)	P-value
Age (years)	n	75	77	0.9807 [a]
	Mean ± SD	76.8 ± 5.92	76.8 ± 4.53	
	Median	75.0	76.0	
	Min, Max	70, 98	70, 90	
Sex – n (%)	Male	23 (30.7)	29 (37.7)	0.3634 [b]
	Female	52 (69.3)	48 (62.3)	

Example 4

This example demonstrate STACKSPC, INTOTCAT and INALLCAT option.

```
proc format;
  value sexn
    1='Male'
    2='Female'
```

```

3='Unknown';
run;
%asumstat(indsnm=adam.adsl\trt01pn notin (1,0),
  outdsnm=tfin,
  displbl=Age (years)Sex - n (%),
  anlyzvrs=age\8.\con|sexn\sexn.\cat,
  inallcat=y,
  stackspc=3);

```

Demographic Parameter Statistics	Active (N=XXX)	Placebo (N=XXX)
Age (years)	75	77
n	76.8 ± 5.92	76.8 ± 4.53
Mean ± SD	75.0	76.0
Median	70, 98	70, 90
Min, Max		
Sex - n (%)		
Total	75	77
Male	23 (30.7)	29 (37.7)
Female	52 (69.3)	48 (62.3)
Unknown	0 (0.0)	0 (0.0)

Example 5

This example demonstrate GRPCOLM option.

```

%asumstat(indsnm=adam.adsl\trt01pn notin (1,0),
  outdsnm=tfin,
  grpcol=trt01pn\trtn.,
  colmvar=sexn\n\sexn.,
  displbl=Age (years),
  anlyzvrs=age\8.\con);

```

Demographic Parameter	Statistics	Active		Placebo	
		Male	Female	Male	Female
Age (years)	n	28	47	31	46
	Mean ± SD	76.1 ± 4.55	77.2 ± 6.62	76.3 ± 4.28	77.1 ± 4.71
	Median	75.0	75.0	76.0	76.0
	Min, Max	70, 86	70, 98	70, 89	70, 90

Example 6

This example uses DISFMT option customized continuous and categorical display format and its order.

```

%asumstat(indsnm=adam.adsl\trt01pn notin (1,0),
  outdsnm=tfin,
  colmvar=trt01pn\n\trtn.,
  displbl=Age (Years)Sex - n (%),
  anlyzvrs=age\8.\con|sex\sex.\cat,
  disfmt=1 2 3 12 9 13|5\0);

```

	Statistics	Active	Placebo
Age (Years)	n	75	77
	Mean	76.8	76.8
	SD	5.92	4.53
	Min	70	70
	Median	75.0	76.0
	Max	98	90
Sex – n (%)	Female	23/75 (30.7)	29/77 (37.7)
	Male	52/75 (69.3)	48/77 (62.3)

Example 7

This example uses PGVARBL and BYVARBL.

```
%asumstat(indsnm=adam.advs,
          outdsnm=tfin,
          colmvar=trt01pn\n\trtn.,
          displbl=_null_,
          anlyzvrs=aval\8.\con,
          byvarbl=avisitn,
          pgvarbl=param);
```

Vital Signs: Systolic BP

Visit	Statistics	Active (N=XXX)	Placebo (N=XXX)
Baseline	n	xxx	xxx
	Mean ± SD	xx.xx ± x.xx	xx.xx ± x.xx
	Median	xx.xx	xx.xx
	Min, Max	xx, xx	xx, xx
Week 1	n	xxx	xxx
	Mean ± SD	xx.xx ± x.xx	xx.xx ± x.xx
	Median	xx.xx	xx.xx
	Min, Max	xx, xx	xx, xx

CONCLUSION

One macro handles both categorical and continuous descriptive statistic summary tables along with statistical analysis. This macro is generalized so that they can be used in many situations and as I mentioned in the introduction, this macro will not only save a lot of time but also improve the quality.

REFERENCES

Zhong, Wayne. "One Macro Call to a Table with both Frequency and Summary Elements from a Subject-Level Dataset" *Proceedings of the PharmaSUG 2012, Paper CC18*.

Wang, Chunmao. "Efficiently Produce Descriptive Statistic Summary Tables with SAS Macros." *Proceedings of the PharmaSUG 2015, Paper PO11*.

CONTACT INFORMATION

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

Rajaram Venkatesan

Cognizant Technology Solutions
rajaram.venkatesan@cognizant.com

Harinarayan Gopichandran
Cognizant Technology Solutions
harinarayan.g@cognizant.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.