

Going Beyond Proc Tabulate

Jim Edgington, LabOne, Inc., Lenexa, KS
Carole Lindblade, LabOne, Inc., Lenexa, KS

ABSTRACT

PROC Tabulate is one of the most powerful and versatile of the SAS reporting tools. It is one of very few SAS procedures that allows you to both compute statistics and have extensive control over the display of the results. It is also compatible with the ODS report delivery system built into version 8 SAS.

Even with these advantages, there are limits to what the basic PROC can do. PROC Tabulate has a limited set of default statistics that do include some commonly used descriptive statistics, including median. It also is limited to only four types of denominators for various percentage calculations. Finally, until version 8, PROC Tabulate had to be used in isolation, as the results were trapped in the output and not available to the rest of the SAS program.

INTRODUCTION

With the ability to load data sets, you can both include all the statistical power of Univariate, ANOVA, or any of the other highly computational PROCs that are part of the SAS system, while maintaining the look and feel that your users have come to expect. And new in version 8, the ability to export data sets from PROC Tabulate, allows you to send your data to PROC Report, or even export the results directly to an MS Excel spreadsheet. Some previous limitations encountered using PROC Tabulate have been eliminated, and powerful alternatives are now available.

This paper assumes that you have a basic understanding of PROC Tabulate. The purpose of this paper is to extend beyond what Tabulate can provide by itself; teaching the reader how to use other parts of SAS and integrate that knowledge to extend the capabilities of PROC Tabulate.

EXTENDING STATISTICS AVAILABLE IN PROC TABULATE

The following statistics are included with PROC Tabulate and are available with any cross of the analysis data set. These statistics include the number of observations, number of missing observations, maximum, minimum, range, sum, percentage of observations, and the percentage of the sum. Also included in Tabulate are higher-level statistics including sum of squares, variance, coefficient of variance, standard deviation, standard error, and various T-tests. While this covers the basics of statistical reporting, SAS has a vast array of statistics not included in this list.

Here is an example of how you can load non-standard statistics into PROC Tabulate: Calculations that are beyond PROC Tabulate's abilities are performed using another PROC or a DATA step. The resulting data set must contain a unique index of the class variables. A unique index means that the data set has one and only one row for each combination of class variables. It also must contain the calculated value that is the result of the DATA step or PROC. Then create the PROC Tabulate as normal, with the following exception: Instead of using Tabulate to calculate the values, sum the total of all the analysis variables.

Since you have a single record for the lowest level of the class combination, this sum is equal to the result you calculated by the alternative method.

In this example we want to calculate a median, a function not found in PROC Tabulate, but rather in PROC Univariate, a procedure with very poor reporting characteristics.

```
data mydata;
  input year $4. sex $1. test $8. result best4.
  @;
  output;
  input year $4. sex $1. test $8. result best4.
  @;
  output;
  input year $4. sex $1. test $8. result best4.
  @;
  output;
  input year $4. sex $1. test $8. result best4.
  @;
  output;
  input year $4. sex $1. test $8. result best4.;
output;
cards;
1998MCocaine      101998FPCP      151998MMarijuan
121998FMorphine   01998FAlcohol      0
1998MCocaine      01998FPCP      221998FMarijuan
121998MMorphine   31998MAlcohol      2
1998MCocaine      01998FPCP      51998FMarijuan
21998MMorphine   61998FAlcohol      7
1998MCocaine      31998FPCP      251998MMarijuan
21998FMorphine   41998MAlcohol      9
1998FCocaine      321998MPCP      121998FMarijuan
01998FMorphine   01998MAlcohol      0
1999MCocaine      111999FPCP      01999MMarijuan
01999FMorphine   01999FAlcohol      1
1999MCocaine      101999FPCP      01999FMarijuan
01999MMorphine   01999MAlcohol      3
1999MCocaine      101999FPCP      01999FMarijuan
01999MMorphine   01999FAlcohol      6
1999MCocaine      301999FPCP      01999MMarijuan
01999FMorphine   01999MAlcohol      5
1999FCocaine      01999MPCP      01999FMarijuan
01999FMorphine   01999MAlcohol      9;
run;

proc sort data=mydata; by year sex test; run;

proc univariate data=mydata noprint;
  by year sex test;
  var result;
  output out=stats n=n max=max min=min mean=mean
  median=median;
run;

proc tabulate data=stats;
  class year sex test;
  var n max min mean median;
  table year='Year' * sex='Gender',
         test*(n='N'*sum=' ' max='Max'*sum=' '
         min='Min'*sum=' '
         mean='Mean'*sum=' '
         median='Median'*sum=' ');
  /rts=20;
  title 'Example of Loading Statistics into PROC
```

Example of Loading Statistics into PROC Tabulate

CALCULATING NON-STANDARD PERCENTAGES

While these often provide sufficient percentage calculations, there are times when a different method of calculating percentages is desirable. One example is when dealing with missing values.

1999MCocaine	111999FPCP
01999MMarijuan	01999FMorphine
01999FAlcohol	1
1999MCocaine	101999FPCP
01999FMarijuan	01999MMorphine
01999MAlcohol	3
1999MCocaine	101999FPCP
01999FMarijuan	01999MMorphine
01999FAlcohol	6
1999MCocaine	301999FPCP
01999MMarijuan	01999FMorphine
01999MAlcohol	5
1999FCocaine	01999MPCP
01999FMarijuan	01999FMorphine
01999MAlcohol	9
1998MCocaine	.1998FPCP
.1998MMarijuan	.1998FMorphine
.1998MAlcohol	.
1998FCocaine	.1998MPCP
.1998FMarijuan	.1998FMorphine
.1998MAlcohol	.
1999MCocaine	.1999FPCP
.1999MMarijuan	.1999FMorphine
.1999FAlcohol	.
1999MCocaine	.1999FPCP
.1999FMarijuan	.1999MMorphine
.1999MAlcohol	.
:	

```
proc sort data=mydata; by year sex test;
```

```
data tabload(keep = year sex test cnt_test
tot_test cnt_miss);
  set mydata;
  by year sex test;
  retain tot_test cnt_test cnt_miss 0;
```

```
if first.test then do;
  cnt_test = 0;
  cnt_miss = 0;
end;
```

```

if result ne . then do;
    cnt_test = cnt_test + 1;
    tot_test = tot_test + 1;
end;
else cnt_miss = cnt_miss + 1;

```

```

    if last.test then output;
run;
```

```
proc sort data=tabload;
  by descending year descending sex descending
  test;
```

```
data mydata;
  input year $4. sex $1. test $8. result
best4. @;
  output;
  input year $4. sex $1. test $8. result
best4. @;
  output;
  input year $4. sex $1. test $8. result
best4. @;
  output;
  input year $4. sex $1. test $8. result
best4. @;
  output;
  input year $4. sex $1. test $8. result
best4.;
output;
cards;
1998MCocaine      101998FPCP
151998MMarijuan  121998FMorphine
01998FAlcohol     0
1998MCocaine      01998FPCP
221998FMarijuan  121998MMorphine
31998MAlcohol     2
1998MCocaine      01998FPCP
51998FMarijuan   21998MMorphine
61998FAlcohol     7
1998MCocaine      31998FPCP
251998MMarijuan  21998FMorphine
41998MAlcohol     9
1998FCocaine      321998MPCP
121998FMarijuan   01998FMorphine
01998MAlcohol     0
```

```
proc tabulate data=tabload1 missing;
  class year sex test;
  var cnt_test cnt_per cnt_miss;
  table year='Year' * sex='Gender',
        test*(cnt_test='Tests'*sum= ' '*f=5.
              cnt_per='% '*sum= ' '*f=6.1
              cnt_miss='Miss'*sum= ' '*f=5.)
        /rts=16;
  title 'Example of Loading Percentages into
PROC Tabulate';
run;
```

		test									
		Alcohol		Cocaine		Marijuan		Morphine		PCP	
		Tests	%	Tests	%	Tests	%	Tests	%	Tests	%
Year	Gender										
1998	F	2	4.0	0	1	2.0	1	3	6.0	1	3
	M	3	6.0	2	4	8.0	1	2	4.0	1	2
1999	F	2	4.0	1	1	2.0	0	3	6.0	1	3
	M	3	6.0	1	4	8.0	2	2	4.0	1	2

B1998MCocaine	31998FPCP
251998MMarijuan	21998FMorphine
41998MAlcohol	9
A1998FCocaine	321998MPCP
121998FMarijuan	01998FMorphine
01998MAlcohol	0
B1999MCocaine	111999FPCP
01999MMarijuan	01999FMorphine
01999FAlcohol	1
A1999MCocaine	101999FPCP
01999FMarijuan	01999MMorphine
01999MAlcohol	3
B1999MCocaine	101999FPCP
01999FMarijuan	01999MMorphine
01999FAlcohol	6
A1999MCocaine	301999FPCP
01999MMarijuan	01999FMorphine
01999MAlcohol	5
B1999FCocaine	01999MPCP
01999FMarijuan	01999FMorphine
01999MAlcohol	9
A1998MCocaine	.1998FPCP
.1998MMarijuan	.1998FMorphine
.1998MAlcohol	.
B1998FCocaine	.1998MPCP
.1998FMarijuan	.1998FMorphine
.1998MAlcohol	.
A1999MCocaine	.1999FPCP
.1999MMarijuan	.1999FMorphine
.1999FAlcohol	.
B1999MCocaine	.1999FPCP
.1999FMarijuan	.1999MMorphine
.1999MAlcohol	.
i	

```
proc sort data=mydata; by comp_cd year sex
test; run;
```

```
data tabload(keep = comp_cd year sex test
cnt_test tot_comp cnt_miss);
set mydata;
by comp_cd year sex test;
retain tot_comp cnt_test cnt_miss 0;

if first.comp_cd then tot_comp = 0;
if first.test then cnt_test = 0;

if result ne . then do;
    cnt_test = cnt_test + 1;
```

```

cnt_per = cnt_test / total * 100;
run;

proc tabulate data=tabload1;
  class comp_cd year sex test;
  var cnt_test cnt_per cnt_miss;
  table year='Yr' * sex='Sex',

comp_cd='Co.'*test*(cnt_test='Test'*sum='
'*f=4.
               cnt_per='%'*sum=' '*f=3.1
cnt_miss='Mis'*sum= ' '*f=3.)
  /rts=11;
  title 'Example of Loading Percentages into
PROC Tabulate';
run;

```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
;                                                                                               Co-
;                                                                                               -
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
;                                                                                               A                               B
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
;                                                                                               Test
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
; Alcohol Cocaine Marijuana Morphine PCP Alcohol Cocaine Marijuana Morphine PCP
; Test % Mis Test % Mis Test % Mis Test % Mis Test % Mis Test % Mis Test % Mis Test % Mis Test % Mis Test % Mis
Yr Sex
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1998,F 2,8,0 0 1,4,0 0 2,8,0 0 2,8,0 1 2,8,0 2 0,0,0 11 1,4,0 12 1,4,0 13 2,8,0 13
; N 1,4,0 3 2,8,0 4 1,4,0 5 1,4,0 6 2,8,0 7 2,8,0 8 1,4,0 9 2,8,0 10 1,4,0 11 2,8,0 12 1,4,0 13 2,8,0 14
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
1999,F 0,0,0 6 1,4,0 6 1,4,0 7 2,8,0 8 2,8,0 15 1,4,0 15 2,8,0 16 2,8,0 16 2,8,0 17
; N 0,0,0 9 1,4,0 10 1,4,0 11 2,8,0 12 2,8,0 13 1,4,0 14 2,8,0 15 2,8,0 16 2,8,0 17
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
; var cnt_test cnt_per cnt_miss;
; table year='Yr' * sex=Sex';

```

New in version 8, PROC Tabulate can now export a data set containing all the information to reproduce the Tabulate output. In order to use this feature, you must understand how the data set is structured. Take the Tabulate function from the last example and add OUT=MyTab to the initial line and the results of the PROC Tabulate is stored in a data set. Follow this with a PROC CONTENTS and a PROC PRINT to see what has been created.

```
comp_cd='Co.'*test*(cnt_test='Test'*sum='
'*f=4.
                                cnt_per='%'*sum=' '*f=3.1
cnt_miss='Mis'*sum= ' '*f=3.)
                                /rts=11;
                                title 'Example of Loading Percentages into
PROC Tabulate';
run;
```

```
proc contents data=mytab; run;
proc print data=mytab uniform; run;
```

```
Observations: 37
Variables: 10
Indexes: 0
Observation Length: 64
Deleted Observations: 0
Compressed: NO
Sorted: NO
```

```
Data Set Page Size:      8192
Number of Data Set Pages: 1
First Data Page:        1
Max Obs per Page:       127
Obs in First Data Page:  37
Number of Data Set Repairs: 0
File Name:               D:\TEMP\SAS Temporary Files\_TD74\mytab.sas7bdat
Release Created:          8.0101M0
Host Created:             WIN_NT
```

#	Variable	Type	Len	Pos	Label
6	_PAGE_	Num	8	0	Page for Observation
7	_TABLE_	Num	8	8	Table for Observation
5	_TYPE_	Char	4	54	Type of Observation
10	cnt_miss_Sum	Num	8	32	
9	cnt_per_Sum	Num	8	24	
8	cnt_test_Sum	Num	8	16	
1	comp_cd	Char	1	40	
3	sex	Char	1	45	
4	test	Char	8	46	
2	year	Char	4	41	

obs	comp	cd	year	sex	test	TYPE	PAGE	TABLE	cnt_ test Sum	cnt_ per Sum	cnt_ miss Sum
-----	------	----	------	-----	------	------	------	-------	------------------	-----------------	------------------

1	A	1998	F	Alcohol	1111	1	1	2	8	0
2	A	1998	F	Cocaine	1111	1	1	1	4	0
3	A	1998	F	Marijuan	1111	1	1	2	8	0
4	A	1998	F	Morphine	1111	1	1	2	8	1
5	A	1998	F	PCP	1111	1	1	2	8	2
6	B	1998	F	Cocaine	1111	1	1	0	0	1
7	B	1998	F	Marijuan	1111	1	1	1	4	12
8	B	1998	F	Morphine	1111	1	1	1	4	13
9	B	1998	F	PCP	1111	1	1	2	8	13
10	A	1998	M	Alcohol	1111	1	1	1	4	3
11	A	1998	M	Cocaine	1111	1	1	2	8	4
12	A	1998	M	Marijuan	1111	1	1	1	4	5
13	A	1998	M	Morphine	1111	1	1	1	4	5
14	A	1998	M	PCP	1111	1	1	1	4	5
15	B	1998	M	Alcohol	1111	1	1	2	8	14
16	B	1998	M	Cocaine	1111	1	1	2	8	14
17	B	1998	M	Marijuan	1111	1	1	1	4	14
18	B	1998	M	Morphine	1111	1	1	1	4	14
19	B	1998	M	PCP	1111	1	1	0	0	15
20	A	1999	F	Alcohol	1111	1	1	0	0	6
21	A	1999	F	Marijuan	1111	1	1	1	4	6
22	A	1999	F	Morphine	1111	1	1	1	4	7
23	A	1999	F	PCP	1111	1	1	2	8	8
24	B	1999	F	Alcohol	1111	1	1	2	8	15
25	B	1999	F	Cocaine	1111	1	1	1	4	15
26	B	1999	F	Marijuan	1111	1	1	2	8	16
27	B	1999	F	Morphine	1111	1	1	2	8	16
28	B	1999	F	PCP	1111	1	1	2	8	17
29	A	1999	M	Alcohol	1111	1	1	2	8	8
30	A	1999	M	Cocaine	1111	1	1	2	8	9
31	A	1999	M	Marijuan	1111	1	1	1	4	10
32	A	1999	M	Morphine	1111	1	1	1	4	10
33	B	1999	M	Alcohol	1111	1	1	1	4	18
34	B	1999	M	Cocaine	1111	1	1	2	8	19
35	B	1999	M	Marijuan	1111	1	1	1	4	19
36	B	1999	M	Morphine	1111	1	1	1	4	20
37	B	1999	M	PCP	1111	1	1	1	4	20

SUBHEAD (HEADER 2)

Each row in the data table represents a unique level of the Class definition of the PROC Tabulate report.

The various CLASS variables from the PROC Tabulate are included in the data set. (In this example, the CLASS Variables are YEAR, SEX, and TEST.) The analysis variables are calculated, and the value of the cell is assigned. Since there are three analysis variables, three variables are created. (cnt_test_Sum, cnt_per_Sum, cnt_miss_Sum). Three new variables are also created: _type_, _page_, and _table_. Each of these variables represents an individual cell.

The _type_ variable is created in a manner similar to the _type_ variable found in PROC Summary. This is a text variable that stores a binary number that indicates which members of the Class variables contributed to its total. The first character represents the first Class variable, the second character, the second class variable, and so on. In our example, all these values are '1' or true. Had we used PROC Tabulate's ability to subtotal, the noncontributing factors would have been '0'. This will allow the user to extract the results from PROC Tabulate and use the data in various other PROCs.

CONCLUSION

PROC Tabulate is one of the most powerful reporting tools in the SAS system. Even so, Tabulate has its limitations. By using other SAS PROCs and the Data step procedure, you can extend the power of Tabulate beyond its original limitations. In this exercise, we have shown how to load the results of other PROCs by calculating the results as distinct values of the class variables and summing those results. We have also shown various methods of calculating percentages other than the default statistics found in PROC Tabulate. And finally, we have shown that in SAS Version 8 the results of Tabulate can be exported into a SAS data set to be used elsewhere in your SAS program.

REFERENCES

Much of the specific information on the new features of PROC Tabulate, in particular the exporting of the data set and the new percentage functions found in version 8, were researched at the SAS web sit. (www.sas.com)

ACKNOWLEDGMENTS

We would like to thank Lisa Ousley with her help with the final editing of this paper.

CONTACT INFORMATION (HEADER 1)

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

Jim Edgington
LabOne, Inc.
10101 Renner Blvd
Lenexa, KS 66203
W: (913) 577-1335
Fax: (913) 8884160
Email: jim.edgington@labone.com

Carole Lindblade
LabOne, Inc.
10101 Renner Blvd
Lenexa, KS 66203
W: (913) 577-1350
Fax: (913) 8884160
Email: carole.lindblade@labone.com