Using PROC FORMAT and other little tweaks to enable PROC TABULATE's hidden capabilities of optimizing statistical reporting

Heli Ghandehari[1], Victor A Lopez[1]
[1]Baxter Healthcare Corporation, Westlake Village, CA 91362

PROC TABULATE is arguably the most efficient approach for calculating statistics and generating output, all within one procedure. However, developers must often stray from TABULATE when display specifications require values to be reported as concatenated pairs. For example, a common reporting requirement is for a mean and standard deviation to be grouped within a single cell, with the latter enveloped by brackets. Similarly, a range could be requested with the minimum and maximum delimited by a dash, or perhaps a confidence interval nestled within parentheses. The combinations are endless, but the underlying solution is simple and universal. This presentation demonstrates the utility of PROC FORMAT's PICTURE statement when applied in combination with TABULATE's computational and reporting capabilities to create customized statistical tables.

Introduction

Consider the following example using the CLASS dataset from the SASHELP library. Suppose that one's task is to generate a report with a simple count (xx) and percentage (yy.y) of the gender breakdown within the classroom. Now suppose that it is also required to present this information as a concatenated pair resembling the following: xx (yy.y). While the calculations and report generation can be done simultaneously with PROC TABULATE, most SAS® programmers will turn to a DATA step as soon as symbols or concatenations are required.

Contrary to popular belief, creating the same table with PROC TABULATE is simple and extensible. The first step is to utilize PROC FORMAT's PICTURE statement which allows inclusion of symbols within and around a formatted value. The *paren*. format defined below will envelope any numeric value within parentheses. The ROUND option indicates that the numeric value will be rounded according to the specified digital selectors of '009.9' (in this case, equivalent to a W.d format of 5.1).

The next step involves setting up the TABULATE procedure in such a way that accommodates the visual merging of two cells. Within the STYLE statement of each statistic, cell justification is adjusted to allow for the frequency (*n*) and percentage (*pctn*) values to be placed side by side. The BORDERRIGHTSTYLE and BORDERLEFTSTYLE options are then set to hidden in order to suppress the cell boundaries between the two values. Finally, the *paren*. format is applied to the percentage.

Below is a comparison of the DATA step approach (Output 1) versus the proposed PROC FORMAT/PROC TABULATE approach (Output 2).

Classic DATA step approach

The DATA step approach would require 1) a procedure to perform the calculations, 2) a DATA step to add parentheses around the percentage and concatenate the values, and 3) a reporting procedure.

```
** Calculate frequency and percentage;
ods output OneWayFreqs=one (keep=sex frequency percent);
proc freq data=sashelp.class;
  table sex;
run;

** Add parentheses and concatenate;
data two;
  set one;
  cntpct =
strip(put(frequency,best.))||'('||strip(put(percent,8.1))||')';
run;

** Optional transposition;
data three (keep=male female);
  merge two (where=(sex='M') rename=(cntpct=male  ))
        two (where=(sex='F') rename=(cntpct=female));
run;

** Output;
proc report data=three nowd split='|';
  columns female male;
    define female / 'Female|n (%)' style=[cellwidth=200 just=C];
    define male   / 'Male|n (%)'   style=[cellwidth=200 just=C];
run;
```

Female n (%)	Male n (%)
9 (47.4)	10 (52.6)

Output 1. PROC REPORT – Number and percent of students by Gender (ODS RTF)

Proposed approach

The proposed approach would require 1) PROC FORMAT with PICTURE statement, and 2) a reporting procedure.

```
** Define format;
proc format;
  picture paren (round)
    low-high = '(009.9)'
    ( prefix = '(' );
run;

** Calculate, concatenate and output;
proc tabulate data=sashelp.class;
  class sex;
  table sex=''*
    (n=''*[style=[just=R cellwidth=90 borderrightstyle=hidden]]
     pctn=''*[style=[just=L cellwidth=110
                     borderleftstyle=hidden]f=paren.]);
run;
```

Female n (%)	Male n (%)
9 (47.4)	10 (52.6)

Output 2. PROC TABULATE – Number and percent of students by Gender (ODS RTF)

While this presentation focuses on PROC TABULATE, the concepts presented here can also be applied to other procedures as well. In the example below (Output 3), the mean and standard deviation for age is presented for each of the gender groups. Note that the digital selectors for the *paren.* format have been modified to accommodate for two decimal places in the standard deviation.

Code

```
proc format;
  picture paren (round)
    low-high = '(009.99)'
    ( prefix = '(' );
run;

proc report data=sashelp.class nowd split='|';
 columns ('Gender' sex) ('Age|Mean (SD)' age=agem age=agesd);
   define sex / ' ' group style=[cellwidth=200 just=C];
   define agem / ' ' analysis mean style=[cellwidth=100 just=R
                                  borderrightsyle=hidden] f=8.1;
   define agesd / ' ' analysis std style=[cellwidth=100 just=L
                               borderleftstyle=hidden] f=paren.;
run;
```

ODS RTF Output

Gender	Age Mean (SD)
Female	13.2 (1.39)
Male	13.4 (1.65)

Output 3. PROC REPORT – Students' Age (Mean and Standard Deviation) by Gender

Examples

Before we apply this approach further, let's work around the need to hide the borders between table cells that we would like to see "merged" within the PROC TABULATE output. We can do this by defining a simple style template that removes all vertical lines, similar to presentations seen in scientific journals. (Note: This can also be done by calling one of the predefined templates such as STYLES.JOURNAL.)

```
proc template;
  define style nolines;
     parent = styles.minimal;
     replace table from document /
             frame = hsides /* borders at top and bottom */
             rules = groups /* line separating table headers */
       cellpadding = 2pt    /* minimize cell padding as needed */
       cellspacing = 0pt    /* remove cell spacing to avoid white "line" between cells */
       borderwidth = 1.5pt; /* set borders to desired width */
   end;
run;
```

We will now generate a listing displaying mean height along with a 95% confidence interval that appears to be within a single cell (Output 4). As before, our first step is to define our formats followed by our TABULATE statements.

Code

```
proc format;
  picture lcl (round) /* Lower limit requires left parenthesis and comma */
    low-high = '009.99,'
    ( prefix = '(' );

  picture ucl (round)
    low-high = '009.99)'; /* Upper limit requires only right parenthesis */
run;

proc tabulate data=sashelp.class alpha=0.05;
  class sex;
  var height;
  keyword lclm / style=[just=R]; /* Create illusion of merged cells */
  keyword uclm / style=[just=L];
  table sex='', height*(mean = 'Mean' * [style=[just=C cellwidth=125]       ]
                        lclm = '95%'  * [style=[just=R cellwidth=75 ] f=lcl.]
                        uclm = 'C.I.' * [style=[just=L cellwidth=75 ] f=ucl.]);
run;
```

ODS HTML Output

	Height	
	Mean	95% C.I.
Female	60.59	(56.73, 64.45)
Male	63.91	(60.38, 67.44)

Output 4. PROC TABULATE – Students' Height (Mean and 95% Confidence Intervals) by Gender

We can now explore more complicated tables that require multiple combined cells. In the example below, we calculate the descriptive statistics for height for males and females (Output 5). We also introduce the *enddash.* format to accommodate the column for Min/Max. Notice the combined pairs that are coupled below.

Code

```
proc format;
  picture enddash (round)
    low-high = '00009.9'
    ( prefix = '- ' );
run;

proc tabulate data=sashelp.class;
  class sex;
  var height weight;
  keywords mean median min / style=[just=R];
  keywords std qrange max  / style=[just=L];
  table (sex='' all='Total')*(height weight),
    (n      = 'N'      * [style=[just=C cellwidth=75]]
     Mean   = 'Mean'   * [style=[just=R cellwidth=75] f=8.2]
     Std    = '(SD)'   * [style=[just=L cellwidth=75] f=paren.]
     Median = 'Med'    * [style=[just=R cellwidth=75] f=8.2]
     Qrange = '(IQR)'  * [style=[just=L cellwidth=75] f=paren.]
     Min    = 'Min'    * [style=[just=R cellwidth=75] f=8.1]
     Max    = '- Max'  * [style=[just=L cellwidth=75] f=enddash.]) / box=' ';
run;
```

ODS HTML Output

		N	Mean (SD)	Med (IQR)	Min - Max
Female	Height	9	60.59 (5.018)	62.50 (7.800)	51.3 - 66.5
	Weight	9	90.11 (19.384)	90.00 (18.500)	50.5 - 112.5
Male	Height	10	63.91 (4.938)	64.15 (8.000)	57.3 - 72.0
	Weight	10	108.95 (22.727)	107.25 (43.000)	83.0 - 150.0
Total	Height	19	62.34 (5.127)	62.80 (9.000)	51.3 - 72.0
	Weight	19	100.03 (22.774)	99.50 (28.500)	50.5 - 150.0

Output 5. PROC TABULATE – Descriptive Statistics for Students' Height and Weight by Gender and Overall

Conclusion

PROC FORMAT's PICTURE statement, when coupled with tweaks within a reporting procedure, is a powerful tool that allows for the development of more concise code while still adhering to the nuances of statistical presentations. The methods described in this paper can also be extended for use in any type of presentation that requires concatenation of data within a single cell. Eliminating the use of DATA steps increases computing efficiency (especially when working with large datasets) and more concise code allows for easier program debugging.

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

Heli Ghandehari
Baxter Healthcare Corporation
1 Baxter Way
Westlake Village, CA 91362
(805) 372-3548
helieh_ghandehari@baxter.com

Victor A Lopez
Baxter Healthcare Corporation
1 Baxter Way
Westlake Village, CA 91362
(805) 372-4048
victor_lopez1@baxter.com

SAS® GLOBAL FORUM