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

Heli Ghandehari¹, Victor A Lopez¹
¹Baxter 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 stay 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 required 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 envelop 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 statistical 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 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.

**Proposed approach**

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

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.)

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.

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.

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

Baxter Healthcare Corporation

Baxter Healthcare Corporation

(800) 825-9000

(800) 825-9000

Baxter Healthcare Corporation

Baxter Healthcare Corporation

(800) 825-9000

(800) 825-9000

SAS and all other SAS Institute Inc. product or service names are registered trademarks or service marks of SAS Institute Inc. in the USA and other countries. ©2023 SAS Institute Inc. All rights reserved. Other brand and product names are trademarks of their respective companies.