

Paper 265-2008

Column Headings and Super-Headings: Using ACROSS Variables in PROC REPORT

Jillian Campbell, Howard M. Proskin and Associates Inc., Rochester, NY

ABSTRACT

Programmers are commonly requested to produce tables that include variable values as column headings. The use of an ACROSS variable in PROC REPORT can easily accommodate this request. In some cases the tables must include multiple variables as column headings. In these cases, two ACROSS variables can be used in PROC REPORT. This technique allows a SAS® programmer to include variables as column headings and super-headings with little to no additional programming.

INTRODUCTION

When a programmer is given a table template and the task of creating a matching data table using SAS, there are many ways the programmer can reach that end goal. The use of across variables can increase program efficiency and decrease program complexity. In this paper I will work through an example of how to generate a table that includes a heading and a super-heading (Table 1). For simplicity I have divided the programming into two steps: (1) formatting the input dataset, and (2) using the proper Proc Report commands.

Table 1

	Children	Clothes & Shoes		Sports		
Statistic	Children Sports	Clothes	Shoes	Swim Sports	Team Sports	Winter Sports
MAX	3161	7332	7850	3704	3674	3462
MEAN	716	1159	3858	2256	1036	1453
MIN	10	25	1759	1117	109	485
N	176	240	48	16	64	32
STD	671	1636	1674	1021	919	847

Source data:sashelp.orsales

FORMATTING THE DATASET

To make use of an across variable, every record in the input dataset must include a variable that contains the value of the across variable. For example, Table 1 involves the use of two separate across variables, product line and product category, thus the input dataset must include a variable for product line and a variable for product category (see Figure 1).

In the current example the input dataset (Figure 1) was created with the following program:

```

**SUMMARIZE THE DATA FOR THE PROC REPORT PROCEDURE;
PROC SUMMARY DATA=SASHELP.ORSALES;
  **TOO MANY PRODUCT CATEGORIES TO FIT ON ONE PAGE, SO SOME ARE ELIMIATED;
  WHERE PRODUCT_CATEGORY NOT IN('ASSORTED SPORTS ARTICLES',
    'GOLF', 'INDOOR SPORTS', 'RACKET SPORTS', 'RUNNING - JOGGING', 'OUTDOORS');
  **PRODUCT LINE AND PRODUCT CATEGORY WILL BE MY ACCROSS VARIABLES;
  CLASS PRODUCT_LINE PRODUCT_CATEGORY;
  TYPES PRODUCT_LINE*PRODUCT_CATEGORY;
  VAR QUANTITY PROFIT;
  OUTPUT OUT=FIGURE1(DROP=_TYPE_ _FREQ_);
RUN;

```

Figure 1

	Product Line	Product Category	_STAT_	Number of Items	Profit in USD
1	Children	Children Sports	N	176	176.00
2	Children	Children Sports	MIN	10	209.80
3	Children	Children Sports	MAX	3161	58258.01
4	Children	Children Sports	MEAN	716	13733.64
5	Children	Children Sports	STD	671	13890.23
6	Clothes & Shoes	Clothes	N	240	240.00
7	Clothes & Shoes	Clothes	MIN	25	1665.75
8	Clothes & Shoes	Clothes	MAX	7332	209707.67
9	Clothes & Shoes	Clothes	MEAN	1159	38368.23
10	Clothes & Shoes	Clothes	STD	1636	45876.88
11	Clothes & Shoes	Shoes	N	48	48.00
12	Clothes & Shoes	Shoes	MIN	1759	90265.95
13	Clothes & Shoes	Shoes	MAX	7850	383969.46
14	Clothes & Shoes	Shoes	MEAN	3858	185198.87
15	Clothes & Shoes	Shoes	STD	1674	84531.85
16	Sports	Swim Sports	N	16	16.00
17	Sports	Swim Sports	MIN	1117	22944.55
18	Sports	Swim Sports	MAX	3704	74303.54
19	Sports	Swim Sports	MEAN	2256	45491.79
20	Sports	Swim Sports	STD	1021	19898.91
21	Sports	Team Sports	N	64	64.00
22	Sports	Team Sports	MIN	109	1349.70
23	Sports	Team Sports	MAX	3674	64959.90
24	Sports	Team Sports	MEAN	1036	15738.11
25	Sports	Team Sports	STD	919	17293.59
26	Sports	Winter Sports	N	32	32.00
27	Sports	Winter Sports	MIN	485	52940.85
28	Sports	Winter Sports	MAX	3462	254706.73
29	Sports	Winter Sports	MEAN	1453	122776.06
30	Sports	Winter Sports	STD	847	62221.13

Notice that in the input dataset each record contains a value for product line and for product category represents. Also, notice that each record corresponds with one data cell in Table 1.

The input dataset is now complete. The next step is using PROC REPORT to produce the data table.

USING PROC REPORT

BACKGROUND

PROC REPORT includes seven possible statements: BREAK, BY, COLUMN, DEFINE, FREQ, RBREAK, and WEIGHT. In this paper, only the COLUMN statement and DEFINE statement are used.

The COLUMN statement in PROC REPORT is used to list the variables that will be presented in the report. The COLUMN statement is used by PROC REPORT to determine the order and arrangement that the variables will be displayed.

The DEFINE statement in PROC REPORT is used to define what type of variable for each variable listed in the COLUMN statement. The PROC REPORT procedure includes five different types of variables that can be defined: ACROSS, ANALYSIS, COMPUTED, DISPLAY, GROUP, and ORDER. All of these, with the exception of ACROSS variables, will be presented vertically. (Table 1 utilizes GROUP, ANALYSIS, and ACROSS variables.)

GROUP variables are used to display multiple records from the input dataset in a single row of the report. In the example presented in Table 1, the first column of the report includes a GROUP variable, '_STAT_' (Statistic). (Notice that each record in Figure 1 that contains _STAT_='MAX' is displayed in the first row of the report.)

ANALYSIS variables can be used to present numerical variables. All numeric variables are ANALYSIS by default. The variable 'QUANTITY' (Number of Items) is as an ANALYSIS variable.

ACROSS variables are used to define columns in PROC REPORT, each value of an ACROSS variable is presented as a column in the report. In this example, there are two ACROSS variables: PRODUCT_LINE ('Product Line') and PRODUCT_CATEGORY ('Product Category').

THE COLUMN STATEMENT

The COLUMN statement must carefully layout the order in which the columns and across variables will occur. The first variable mentioned in the COLUMN statement will be the variable that occurs in the far left of the report.

The syntax for where to list ACROSS variables in the COLUMN statement is a little more difficult to understand. The variable that will be used as the top most heading must appear before the second heading, which must appear before the third heading, and so on. All ACROSS variables in the COLUMN statement must be separated by a comma. After all ACROSS variables have been listed (separated by commas), the name of the variable that includes the data table cell values is listed. The COLUMN statement for Table 1 is:

```
COLUMN ('Statistic' _STAT_) PRODUCT_LINE,PRODUCT_CATEGORY,QUANTITY;
```

The above COLUMN statement indicates that _STAT_ will be presented in a column to the far left (with the word 'Statistic' as a column label), PRODUCT_LINE will define the column super-headings, PRODUCT_CATEGORY will define column headings, and QUANTITY contains the data to be presented in the report.

THE DEFINE STATEMENTS

Once the COLUMN statement has been defined, each variable should be defined using a DEFINE statement. The PROC REPORT code used to produce Table 1 is given below:

```
PROC REPORT DATA=FIGURE1 NOWINDOWS;
  COLUMN ('STATISTIC' _STAT_) PRODUCT_LINE,PRODUCT_CATEGORY,QUANTITY;
  DEFINE _STAT_ / ' ' GROUP;
  **THE NO ZERO COMMAND IS USED SINCE ALL PRODUCT CATEGORIES WILL NOT BE
  REPRESENTED FOR EACH PRODUCT LINE IN THE TABLE;
  DEFINE PRODUCT_LINE/ ' ' ACROSS NOZERO ORDER=INTERNAL;
  DEFINE PRODUCT_CATEGORY / ' ' ACROSS ORDER=INTERNAL;
  DEFINE QUANTITY / ' ' ;
  RUN;
```

In the completed PROC REPORT there are two additional commands that are necessary that were not previously discussed. The first is that if two or more ACROSS variables are used in the DEFINE statement the NOZERO command is often necessary. The second is that in order for the column label for GROUP variables (*i.e.*, _STAT_) to align horizontally with the ACROSS variables the column label must be presented in the COLUMN statement as opposed to in the DEFINE STATEMENT.

The command NOZERO indicates that empty columns will not be presented. Without this command the same column headings will be presented for all values of ACROSS variables (Table 2 is produced when the NOZERO command is excluded from the PROC REPORT code).

Table 2

	Children					
Statistic	Children Sports	Clothes	Shoes	Swim Sports	Team Sports	Winter Sports
MAX	3161
MEAN	716
MIN	10
N	176
STD	671

Clothes & Shoes					
Children Sports	Clothes	Shoes	Swim Sports	Team Sports	Winter Sports
.	7332	7850	.	.	.
.	1159	3858	.	.	.
.	25	1759	.	.	.
.	240	48	.	.	.
.	1636	1674	.	.	.

Sports					
Children Sports	Clothes	Shoes	Swim Sports	Team Sports	Winter Sports
.	.	.	3704	3674	3462
.	.	.	2256	1036	1453
.	.	.	1117	109	485
.	.	.	16	64	32
.	.	.	1021	919	847

If the column labels for GROUP variables are presented in the DEFINE statement, as opposed to in the COLUMN statement, the data table will include an extra row for the column label (see Table 3).

Table 3

	Children	Clothes & Shoes		Sports		
	Children Sports	Clothes	Shoes	Swim Sports	Team Sports	Winter Sports
Statistic						
MAX	3161	7332	7850	3704	3674	3462
MEAN	716	1159	3858	2256	1036	1453
MIN	10	25	1759	1117	109	485
N	176	240	48	16	64	32
STD	671	1636	1674	1021	919	847

DISCUSSION

ACROSS variables can be easily used to produce tables with multiple column headings; however, there are a few issues that are worthy of mention.

The CELLWIDTH style option in the DEFINE statement is not easily implemented. If the CELLWIDTH option is specified in the DEFINE statement for the ANALYSIS variable then all ACROSS variable columns will be of identical size.

For the display of character variables, as opposed to numeric ANALYSIS variables, I have had the best results when these variables are specified as GROUP variables. I usually use character variables as opposed to numeric when creating reports, since all values of an ANALYSIS variable will have the same format in the data table. (Note that a character variable could contain the number '4.5' for the Mean record and '12' for the N record.)

Parenthesis can be used to specify the display of multiple values beneath a column heading. The code listed below will produce a report that includes a column for QUANTITY and a column for PROFIT beneath a single column heading (see Table 4):

```
PROC REPORT DATA=FIGURE1 NOWINDOWS;
  COLUMN ('STATISTIC' _STAT_) PRODUCT_LINE,PRODUCT_CATEGORY,(QUANTITY PROFIT);
  DEFINE _STAT_ / '' GROUP;
  **THE NO ZERO COMMAND IS USED SINCE ALL PRODUCT CATEGORIES WILL NOT BE
  REPRESENTED FOR EACH PRODUCT LINE IN THE TABLE;
  DEFINE PRODUCT_LINE/ '' ACROSS NOZERO ORDER=INTERNAL;
  DEFINE PRODUCT_CATEGORY / '' ACROSS ORDER=INTERNAL;
  DEFINE QUANTITY / 'Q' ;
  DEFINE PROFIT / 'P' FORMAT=6.;
  RUN;
```

Table 4

	Children		Clothes & Shoes				Sports					
Statistic	Children Sports		Clothes		Shoes		Swim Sports		Team Sports		Winter Sports	
	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P
MAX	3161	58258	7332	209708	7850	383969	3704	74304	3674	64960	3462	254707
MEAN	716	13734	1159	38368	3858	185199	2256	45492	1036	15738	1453	122776
MIN	10	210	25	1666	1759	90266	1117	22945	109	1350	485	52941
N	176	176	240	240	48	48	16	16	64	64	32	32
STD	671	13890	1636	45877	1674	84532	1021	19899	919	17294	847	62221

CONCLUSION

Defining ACROSS variables in PROC REPORT is an extremely useful tool for producing stylized data tables.

REFERENCES

More information about PROC REPORT can be found in:

SAS Publishing: Base SAS 9.1 Procedures Guide, Volumes 1-4, 2004

SAS Publishing: Base SAS 9.1.3 Procedures Guide, Second Edition, Volumes 1-4, 2006

All information can also be found online in SAS OnlineDoc® 9.1.3

All tables were produced using the sashelp.orsales dataset and the science style

CONTACT INFORMATION

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

Jillian Campbell

Howard M. Proskin and Associates

585-359-2420

jcampbell@hproskin.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.