Paper 085-2011

Custom Sequence of Rows and Columns in PROC TABULATE

Erik W. Tilanus, Driebergen, the Netherlands

ABSTRACT

PROC TABULATE can order class variables in rows and columns in several ways. Internal order or formatted order are commonly used. Internal orders according the internal value of the variable, numeric or alphabetic and formatted order does the same on the formatted values. But what if you want to order your columns in a specific way, which is neither internal nor formatted? A simple trick using PROC FORMAT can help you out.

INTRODUCTION

You can specify the sequence of columns and rows for CLASS variable values in various ways in PROC TABULATE. You specify that with the ORDER= option in the PROC statement as in:

```
PROC TABULATE DATA= ...  ORDER=INTERNAL;
```

The order option can take four values: INTERNAL (or UNFORMATTED, the default value) which orders the class variable values according to their internal value. Next is FORMATTED, which, you guess is ordering according to formatted values. Then there is FREQ, which orders in descending frequency and finally DATA which uses the order in which the data appear in the input data set.

When you are working with date values, for instance in a report covering several month, the INTERNAL order is most appropriate since it will order date values correctly.

Let us turn to the airline world for an example. The input data set contains passengers booked on flights during a certain period. For each flight there are several different fares. Each fare is booked in a different booking- or fareclass. There is only limited standardization in the use of these fare classes. With a few exceptions for first class (F) and full fare economy (Y), each airline uses it's own letters to designate them. A typical series of class codes would be: F C Y B M Q K, from first class, via business class to deep discounted economy fares.

Output 1 shows PROC PRINT output of some observations in the input data set. Program 1 shows a basic PROC TABULATE step to turn this data into a monthly report like in Output 2.

Output 1: PROC PRINT output of the input data set (few observations)

Obs	flight	flight_ date	class	Seats_ sold
...	...	...	.	.
2000	5023	28FEB09	Y	.
2001	5023	01MAR09	B	8
2002	5023	01MAR09	C	.
2003	5023	01MAR09	C	.
2004	5023	01MAR09	C	.
2005	5023	01MAR09	K	19
2006	5023	01MAR09	M	11
2007	5023	01MAR09	Q	46
2008	5023	01MAR09	F	.
2009	5023	01MAR09	Y	1
2010	5023	01MAR09	Y	.
2011	5023	01MAR09	Y	.
2012	5023	02MAR09	B	.
2013	5023	02MAR09	C	2
2014	5023	02MAR09	C	.
2015	5023	02MAR09	C	.
2016	5023	02MAR09	K	19
2017	5023	02MAR09	M	8
2018	5023	02MAR09	Q	19
2019	5023	02MAR09	F	.
2020	5023	02MAR09	Y	.
...	...	...	.	.

Program 1: Basic PROC TABULATE, dates in sequence, booking classes alphabethical

```
PROC TABULATE DATA=booking F=6.;
CLASS flight_date class;
VAR Seats_Sold;
FORMAT flight_date monyy.;
TABLE flight_date,class*Seats_Sold*sum=' ';
LABEL class='Booking class' flight_date='Month';
RUN;
```

Output 2: Plain PROC TABULATE output from Program 1. Note alphabetical order

	Booking class						
	B	C	F	K	M	Q	Y
	Seats-sold	Seats_ sold	Seats_ sold	Seats_ sold	Seats_ sold	Seats_ sold	Seats_ sold
Month							
JUN08	974	781	365	15958	2559	16139	150
JUL08	2361	1558	1451	30121	4881	28455	733
AUG08	2206	1060	1376	23880	4136	16533	607
SEP08	443	551	180	13138	1035	7181	57
OCT08	3580	414	2461	14194	5135	13153	404
NOV08	495	478	256	11724	1301	6987	194
DEC08	1620	1448	892	21189	3029	15699	298
JAN09	2088	800	866	30239	5645	22476	213
FEB09	4697	758	2899	32019	9711	26207	1002
MAR09	2162	372	1145	18391	6277	15671	480

As you can see the months are ordered correctly since we are using the default INTERNAL order. Using FORMATTED would put AUG08 in the first row, followed by DEC08 and FEB08. So internal it will be!

But we also see that the booking classes are now showing up in alphabetical order: B C F K M Q Y It is obvious that in your report you want to maintain the original hierarchical order. How do we get the rows or columns in this order?

THE KEY TO THE SOLUTION: PROC FORMAT.

A practical solution is to create a new numeric variable, derived from class code with an informat that specifies the ranking order. Then we use that variable as the class variable in the reporting and apply a format to translate back to the original values. With that combination we can use the INTERNAL sequence and still have the sequence as we prefer.

Program 2 shows how. The PROC FORMAT step defines an informat ① and a format ② that are each others complement. In the subsequent DATA step we use this informat to create a new variable: Classrank. ③ Now we have a data set, which looks like the one in Output 3. Next follows the PROC TABULATE step. It specifies the Classrank variable as the CLASS variable ④ rather than the Class variable and in the format statement we include the classcode format for Classrank. ⑤ Using the default order (internal) you get exactly what you want.

Note that the Flight_Date variable contains actual flight dates. Using this variable as a class variable would result in a row for every day. However PROC TABULATE groups by the formatted values of the class variables. So by applying the MONYY. format on Flight_Date we produce total figures per month.

Program 2: Creating customized sequence of rows/columns in PROC TABULATE

```
        PROC FORMAT;
①       INVALUE classordinal
        'F' = 1
        'C' = 2
        'Y' = 3
        'B' = 4
        'M' = 5
        'Q' = 6
        'K' = 7';
②       VALUE classcode
        1 = 'F'
        2 = 'C'
        3 = 'Y'
        4 = 'B'
        5 = 'M'
        6 = 'Q'
        7 = 'K';
        RUN;

        DATA booking;
        SET booking;
③       Classrank = INPUT(class,classordinal.);
        RUN;

        PROC TABULATE DATA=booking F=6.;
④       CLASS flight_date classrank;
        VAR Seats_Sold;
        FORMAT flight_date monyy.
⑤              classrank classcode.;
        TABLE flight_date,classrank*Seats_Sold='seats sold'*sum=' ';
        LABEL classrank='Booking class' flight_date='Month';
        RUN;
```

Output 3: Input data set after including the classrank variable

Obs	flight	Flight_ date	class	Seats_ sold	Classrank
...	...	...	.	.	.
2000	5023	28FEB09	Y	.	3
2001	5023	01MAR09	B	8	4
2002	5023	01MAR09	C	.	2
2003	5023	01MAR09	C	.	2
2004	5023	01MAR09	C	.	2
2005	5023	01MAR09	K	19	7
2006	5023	01MAR09	M	11	5
2007	5023	01MAR09	Q	46	6
2008	5023	01MAR09	F	.	1
2009	5023	01MAR09	Y	1	3
2010	5023	01MAR09	Y	.	3
2011	5023	01MAR09	Y	.	3
2012	5023	02MAR09	B	.	4
2013	5023	02MAR09	C	2	2
2014	5023	02MAR09	C	.	2
2015	5023	02MAR09	C	.	2
2016	5023	02MAR09	K	19	7
2017	5023	02MAR09	M	8	5
2018	5023	02MAR09	Q	19	6
2019	5023	02MAR09	F	.	1
2020	5023	02MAR09	Y	.	3
...	...	...	.	.	.

Output 4: PROC TABULATE output now in custom order

	Booking class						
	F	C	Y	B	M	Q	K
	seats sold	seats sold	seats sold	seats sold	seats sold	seats sold	seats sold
Month							
JUN08	365	781	150	974	2559	16139	15958
JUL08	1451	1558	733	2361	4881	28455	30121
AUG08	1376	1060	607	2206	4136	16533	23880
SEP08	180	551	57	443	1035	7181	13138
OCT08	2461	414	404	3580	5135	13153	14194
NOV08	256	478	194	495	1301	6987	11724
DEC08	892	1448	298	1620	3029	15699	21189
JAN09	866	800	213	2088	5645	22476	30239
FEB09	2899	758	1002	4697	9711	26207	32019
MAR09	1145	372	480	2162	6277	15671	18391

CONCLUSION

By defining an informat, specifying the desired ranking of rows or columns and a format to translate that ranking back into the original values, you can define any desired sequence of rows and columns in PROC TABULATE.

WHERE TO GO FROM HERE - RECOMMENDED READING

Books:

Base SAS 9.2 Procedures Guide

Lauren Haworth: PROC TABULATE by Example (SAS Press)

Papers:

086-2011 Christina Morris: Using Multi-label Formats to Create Subtotals in PROC TABULATE

087-2011 Peter Cerussi: Force Missing Rows and Columns with CLASSDATA in Proc Tabulate

260-2011 Art Carpenter: PROC TABULATE: getting started

173-2011 Art Carpenter: PROC TABULATE: doing more

CONTACT INFORMATION

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

Name	Erik Tilanus	Work Phone:	+31 343 517007
Address	Horstlaan 51	Fax:	+31 343 517007
Postal code, City	3971 LB Driebergen	E-mail:	erik.tilanus@planet.nl
Country	the Netherlands	Web:	www.synchrona.nl

At the website you can also find other presentations by the author, held at previous SUGI and SAS Global Forum meetings.