

Paper PO-4

Getting Out of the PROC PRINT Comfort Zone to Start Using PROC REPORT

Imelda C. Go, Office of Assessment, South Carolina Department of Education
 Abbas S. Tavakoli, College of Nursing, University of South Carolina–Columbia

ABSTRACT

PROC PRINT is one of the first things taught to a beginner SAS programmer because it provides an easy and simple way to view the records in a data set. The procedure is fast, simple, and straightforward. As one continues to learn about SAS, one finds out about other procedures, such as PROC REPORT. This paper is written for the PROC PRINT user who has not, for whatever reason, ventured into PROC REPORT territory. The paper provides examples of PROC PRINT code and the corresponding PROC REPORT code that produces the same results. Examples of what PROC REPORT can produce that PROC PRINT cannot are also provided.

INTRODUCTION

The paper will approach the transition from PROC PRINT to PROC REPORT by using a question-and-answer format. Each question will be answered using each PROC with the results presented side-by-side for easy comparison. The paper is written for SAS programmers already familiar with PROC PRINT.

The emphasis will be on using PROC REPORT to produce what is generated by PROC PRINT. There are indeed a variety of ways to get the same things accomplished, but the focus is to provide a simple example that will accomplish the task.

A few examples of PROC REPORT output that are not directly available through PROC PRINT are also provided. The simple examples will hopefully whet the PROC REPORT beginner's appetite to learn more about PROC REPORT's syntax and powerful capabilities.

The sample code in this paper was tested using SAS Version 9.2. The PROC REPORT examples were generated in a batch environment and the interactive/windowing environment is not discussed. Therefore, the NOWD option was used in all the PROC REPORT examples in order to suppress the windowing environment.

DATA SET FOR EXAMPLES

The following data set will be used for the examples:

obsnum	points	grade	lname	fname	gender
1	100	3	Griffin	Angela	F
2	200	3	James	Ellen	F
3	400	4	Peeler	Lane	M
4	600	6	Whitlock	Nancy	F

PROC PRINT VS. PROC REPORT

Q1: How do I get a listing of all the variables in a data set?

You can use the simplest code (one statement) as shown below. All variables will be listed when no variable list is provided. The OBS column, which can be suppressed by the NOBS option, is generated automatically by PROC PRINT. For the PROC PRINT examples, the OBS column will appear in order to remind readers they are looking at PROC PRINT output on the left side of the side-by-side comparisons. PROC REPORT does not generate such a column automatically. Note that by default there is no space between the PROC REPORT variable names and the first record in the data set.

proc print;							proc report nowd;					
Obs	obsnum	points	grade	lname	fname	gender	obsnum	points	grade	lname	fname	gender
1	1	100	3	Griffin	Angela	F	1	100	3	Griffin	Angela	F
2	2	200	3	James	Ellen	F	2	200	3	James	Ellen	F
3	3	400	4	Peeler	Lane	M	3	400	4	Peeler	Lane	M
4	4	600	6	Whitlock	Nancy	F	4	600	6	Whitlock	Nancy	F

Q2: How do I put a space after the variable name row in PROC REPORT so that the output will look more like PROC PRINT output, which by default has a blank line after the variable names?

Use the HEADSKIP option in PROC REPORT

proc print;							proc report nowd headskip;					
Obs	obsnum	points	grade	lname	fname	gender	obsnum	points	grade	lname	fname	gender
1	1	100	3	Griffin	Angela	F	1	100	3	Griffin	Angela	F
2	2	200	3	James	Ellen	F	2	200	3	James	Ellen	F
3	3	400	4	Peeler	Lane	M	3	400	4	Peeler	Lane	M
4	4	600	6	Whitlock	Nancy	F	4	600	6	Whitlock	Nancy	F

Q3: How do I get a listing of selected variables in a data set?

List the variables using the VAR statement with PROC PRINT and the COLUMN statement with PROC REPORT.

<code>proc print;</code> <code>var points lname fname;</code>				<code>proc report nowd headskip;</code> <code>column points lname fname;</code>		
Obs	points	lname	fname	points	lname	fname
1	100	Griffin	Angela	100	Griffin	Angela
2	200	James	Ellen	200	James	Ellen
3	400	Peeler	Lane	400	Peeler	Lane
4	600	Whitlock	Nancy	600	Whitlock	Nancy

Q4: How do I label a variable?

Use the LABEL option and statement in PROC PRINT. Use a DEFINE statement for each variable that needs a label in PROC REPORT.

<pre>proc print label; var points lname fname; label lname='Last Name' fname='First Name';</pre>				<pre>proc report nowd headskip; column points lname fname; define lname/'Last Name'; define fname/'First Name';</pre>		
Obs	points	Last Name	First Name	points	Last Name	First Name
1	100	Griffin	Angela	100	Griffin	Angela
2	200	James	Ellen	200	James	Ellen
3	400	Peeler	Lane	400	Peeler	Lane
4	600	Whitlock	Nancy	600	Whitlock	Nancy

Q5: How do I control the splitting of variable labels?

Use the SPLIT option for each of the two PROCs.

<pre>proc print label split='*'; var points lname fname; label lname='Sur*-name' fname='First Name';</pre>				<pre>proc report nowd headskip split='*'; column points lname fname; define lname/'Sur*-name'; define fname/'First Name';</pre>		
Obs	points	Sur- name	First Name	points	Sur- name	First Name
1	100	Griffin	Angela	100	Griffin	Angela
2	200	James	Ellen	200	James	Ellen
3	400	Peeler	Lane	400	Peeler	Lane
4	600	Whitlock	Nancy	600	Whitlock	Nancy

Q6: How do I change the width of a column of data?

Use a format with PROC PRINT and the WIDTH option with PROC REPORT's DEFINE statement. Note a variable name in PROC REPORT will be displayed up to the same width of the corresponding variable.

<pre>proc print; var lname; format lname \$4.;</pre>		<pre>proc report nowd headskip; column fname; define lname/width=4;</pre>	
Obs	lname	lname	
1	Grif	Grif	
2	Jame	Jame	
3	Peel	Peel	
4	Whit	Whit	

The reduction of the variable name display width in PROC REPORT is even more apparent in the next example where variable labels are used.

<pre>proc print data=demo label; var lname; format lname \$4.; label lname='Student Last Name';</pre>	<pre>proc report nowd headskip; column lname; define lname/width=4 'Student Last Name';</pre>															
<table><tr><th>Obs</th><th>Student Last Name</th></tr><tr><td>1</td><td>Grif</td></tr><tr><td>2</td><td>Jame</td></tr><tr><td>3</td><td>Peel</td></tr><tr><td>4</td><td>Whit</td></tr></table>	Obs	Student Last Name	1	Grif	2	Jame	3	Peel	4	Whit	<table><tr><th>Stud ent Last Name</th></tr><tr><td>Grif</td></tr><tr><td>Jame</td></tr><tr><td>Peel</td></tr><tr><td>Whit</td></tr></table>	Stud ent Last Name	Grif	Jame	Peel	Whit
Obs	Student Last Name															
1	Grif															
2	Jame															
3	Peel															
4	Whit															
Stud ent Last Name																
Grif																
Jame																
Peel																
Whit																

Q7: How do I produce a list with sorted values?

The PROC PRINT example requires sorting by `lname` and `fname` in order to get a list of names in alphabetical order. The data need to be sorted using PROC SORT prior to using PROC PRINT with a BY statement. The data do not need to be sorted using PROC SORT prior to PROC REPORT data processing. Each variable in the sorting criteria will be in a DEFINE statement with an ORDER option.

<pre>proc sort; by lname fname; proc print; var grade lname fname;</pre>	<pre>proc report nowd headskip; column grade lname fname; define lname/order; define fname/order;</pre>																																			
<table><tr><th>Obs</th><th>grade</th><th>lname</th><th>fname</th></tr><tr><td>1</td><td>3</td><td>Griffin</td><td>Angela</td></tr><tr><td>2</td><td>3</td><td>James</td><td>Ellen</td></tr><tr><td>3</td><td>4</td><td>Peeler</td><td>Lane</td></tr><tr><td>4</td><td>6</td><td>Whitlock</td><td>Nancy</td></tr></table>	Obs	grade	lname	fname	1	3	Griffin	Angela	2	3	James	Ellen	3	4	Peeler	Lane	4	6	Whitlock	Nancy	<table><tr><th>grade</th><th>lname</th><th>fname</th></tr><tr><td>3</td><td>Griffin</td><td>Angela</td></tr><tr><td>3</td><td>James</td><td>Ellen</td></tr><tr><td>4</td><td>Peeler</td><td>Lane</td></tr><tr><td>6</td><td>Whitlock</td><td>Nancy</td></tr></table>	grade	lname	fname	3	Griffin	Angela	3	James	Ellen	4	Peeler	Lane	6	Whitlock	Nancy
Obs	grade	lname	fname																																	
1	3	Griffin	Angela																																	
2	3	James	Ellen																																	
3	4	Peeler	Lane																																	
4	6	Whitlock	Nancy																																	
grade	lname	fname																																		
3	Griffin	Angela																																		
3	James	Ellen																																		
4	Peeler	Lane																																		
6	Whitlock	Nancy																																		

The second example sorts the list alphabetically per `grade`. The PROC PRINT example uses the BY and ID statements. The SKIP option in the BREAK AFTER statement puts a line after each cluster of records per `grade` in PROC REPORT.

<pre>proc sort; by grade; proc print; var points lname fname; by grade; id grade;</pre>	<pre>proc report nowd headskip; column grade points lname fname; define grade/order; break after grade/skip;</pre>																																						
<table><tr><th>grade</th><th>points</th><th>lname</th><th>fname</th></tr><tr><td rowspan="2">3</td><td>100</td><td>Griffin</td><td>Angela</td></tr><tr><td>200</td><td>James</td><td>Ellen</td></tr><tr><td>4</td><td>400</td><td>Peeler</td><td>Lane</td></tr><tr><td>6</td><td>600</td><td>Whitlock</td><td>Nancy</td></tr></table>	grade	points	lname	fname	3	100	Griffin	Angela	200	James	Ellen	4	400	Peeler	Lane	6	600	Whitlock	Nancy	<table><tr><th>grade</th><th>points</th><th>lname</th><th>fname</th></tr><tr><td rowspan="2">3</td><td>100</td><td>Griffin</td><td>Angela</td></tr><tr><td>200</td><td>James</td><td>Ellen</td></tr><tr><td>4</td><td>400</td><td>Peeler</td><td>Lane</td></tr><tr><td>6</td><td>600</td><td>Whitlock</td><td>Nancy</td></tr></table>	grade	points	lname	fname	3	100	Griffin	Angela	200	James	Ellen	4	400	Peeler	Lane	6	600	Whitlock	Nancy
grade	points	lname	fname																																				
3	100	Griffin	Angela																																				
	200	James	Ellen																																				
4	400	Peeler	Lane																																				
6	600	Whitlock	Nancy																																				
grade	points	lname	fname																																				
3	100	Griffin	Angela																																				
	200	James	Ellen																																				
4	400	Peeler	Lane																																				
6	600	Whitlock	Nancy																																				

Q8: How do I get observation numbers in the listing?

The observation number is automatically displayed by PROC PRINT unless it is suppressed using the NOBS option. To create the observation number, use PROC REPORT's COMPUTE block to calculate the OBS column. The ENDCOMP statement marks the end of the COMPUTE block.

<pre>proc print; var points lname fname;</pre>				<pre>proc report nowd headskip; column obs points lname fname; compute obs; count+1; obs=count; endcomp;</pre>			
Obs	points	lname	fname	Obs	points	lname	fname
1	100	Griffin	Angela	1	100	Griffin	Angela
2	200	James	Ellen	2	200	James	Ellen
3	400	Peeler	Lane	3	400	Peeler	Lane
4	600	Whitlock	Nancy	4	600	Whitlock	Nancy

Caution: If you use the WHERE statement in PROC PRINT and PROC REPORT, results will likely vary. The OBS number in PROC PRINT is automatically generated and the records will be numbered according to the order they appear in the data set. When a WHERE statement is applied, the original numbering with the complete data set appears in the OBS column in PROC PRINT. In PROC REPORT, the OBS column is created according to the number of rows that are processed. Since the WHERE condition resulted in just one observation, the OBS column value for the PROC REPORT example is 1. In contrast, the OBS column value is 2 in the PROC PRINT example.

<pre>proc print; var points lname fname; where lname='James';</pre>				<pre>proc report nowd headskip; column obs points lname fname; compute obs; count+1; obs=count; endcomp; where lname='James';</pre>			
Obs	points	lname	fname	obs	points	lname	fname
2	200	James	Ellen	1	200	James	Ellen

In order to get the same results in PROC REPORT, the observation number should be determined in the data set prior to using PROC REPORT. The data set has the observation number in the obsnum variable.

<pre>proc print; var obsnum points lname fname; where lname='Griffin';</pre>				<pre>proc report nowd headskip; column obsnum points lname fname; define obsnum/'Obs'; where lname='James';</pre>			
Obs	points	lname	fname	Obs	points	lname	fname
2	200	James	Ellen	2	200	James	Ellen

Q9: How do I get the sum of variables in a listing?

Use the SUM statement with PROC PRINT. Use the RBREAK AFTER statement with the SUMMARIZE option in PROC REPORT. The DOL option indicates a double line should be placed after all the records that are summarized.

<pre>proc print; var points lname fname; sum points;</pre>				<pre>proc report nowd; column points lname fname; rbreak after /summarize dol;</pre>			
Obs	points	lname	fname	points	lname	fname	
1	100	Griffin	Angela	100	Griffin	Angela	
2	200	James	Ellen	200	James	Ellen	
3	400	Peeler	Lane	400	Peeler	Lane	
4	600	Whitlock	Nancy	600	Whitlock	Nancy	
	=====			=====			
	1300			1300			

Q10: How do I get the sum of variables according to values of other variables?

Specify the variables of interest in the SUM statement with PROC PRINT. Use the BREAK AFTER statement with PROC REPORT to produce the sum of the points per grade value. The OL option means one line should be placed prior to the sum of the points per grade. The DOL option means a double line should be placed at the end of the report prior to the sum of the points over all records.

<pre>proc print; by grade; id grade; var points lname fname; sumby grade; sum points;</pre>				<pre>proc report nowd; column grade points lname fname; define grade/order; break after grade/skip summarize ol; rbreak after/summarize dol;</pre>			
grade	points	lname	fname	grade	points	lname	fname
3	100	Griffin	Angela	3	100	Griffin	Angela
	200	James	Ellen		200	James	Ellen
-----	-----			-----	-----		
3	300			3	300		
4	400	Peeler	Lane	4	400	Peeler	Lane
6	600	Whitlock	Nancy	6	400		
	=====			6	600	Whitlock	Nancy
	1300			-----	-----		
				6	600		
					=====		
					1300		

PROC REPORT FEATURES WITH NO PROC PRINT COUNTERPART

In the following examples, PROC REPORT is used to calculate descriptive statistics, perform user-defined calculations, and to display the results. PROC PRINT has a very limited capability for producing descriptive statistics. In order for PROC PRINT to produce output with other types of descriptive statistics, the statistics would be produced by another PROC and then rendered for output using PROC PRINT.

Q11: How do I generate columns for output display using PROC REPORT?

In this PROC REPORT example, two columns in the output appear as a result of using COMPUTE blocks in PROC REPORT to calculate the values. In contrast, PROC PRINT does not have a similar capability.

The `ptsXgrade` variable is simply the product of the `points` and `grade` values. The `wholename` variable is the concatenation of the `lname` and `fname` variables with a comma separating the two.

```
proc report data=demo nowd;
column points grade ptsXgrade lname fname wholename;
define points/display;
define grade/display;
define ptsXgrade/computed 'points X grade';
define wholename/computed;
compute ptsXgrade;
    ptsXgrade=points*grade;
endcomp;
compute wholename/character length=16;
    wholename=trim(lname)||', '||trim(fname);
endcomp;
```

points	grade	points X grade	lname	fname	wholename
100	3	300	Griffin	Angela	Griffin, Angela
200	3	600	James	Ellen	James, Ellen
400	4	1600	Peeler	Lane	Peeler, Lane
600	6	3600	Whitlock	Nancy	Whitlock, Nancy

Q12: How do I generate a simple frequency distribution using PROC REPORT?

The following example is a frequency distribution for the variable `grade`. The percentages are displayed in `percent7.1` format by specifying the `FORMAT` option with the `DEFINE` statement. The `BOX` option was used to display the table gridlines.

```
proc report data=demo nowd box;
column grade n pctn;
define grade/group;
define pctn/format=percent7.1;
```

grade	n	pctn
3	2	50.0%
4	1	25.0%
6	1	25.0%

Q13: How do I generate a frequency distribution with cumulative frequencies and percentages using PROC REPORT?

The following example is a cumulative frequency distribution for the variable `grade`.

```
proc report nowd box;
column grade n pctn cumn cumpctn;
define grade/group;
define pctn/format=percent7.1;
define cumn/computed format=7.;
define cumpctn/computed format=percent7.1;
compute cumn;
    tempn+n;
    cumn=tempn;
endcomp;
compute cumpctn;
    tempn+pctn;
    cumpctn=tempn;
endcomp;
```

grade	n	pctn	cumn	cumpctn
3	2	50.0%	2	50.0%
4	1	25.0%	3	75.0%
6	1	25.0%	4	100%

Q14: How do I generate a cross-tabulation using PROC REPORT?

The following example produces the cross-tabulation between the `grade` and `points` variables. The `GROUP` and `ACROSS` options are used with the corresponding `DEFINE` statements.

```
proc report data=demo nowd box;
column grade points;
define grade/group;
define points/across;
```

grade	points			
	100	200	400	600
3	1	1		
4			1	
6				1

Q15: How do I generate a table of summary statistics using PROC REPORT?

The following example illustrates that PROC REPORT is not limited to displaying record-level detail. It can start mimicking output generated by PROC TABULATE. In this example the sum of the points (`totalpoints` variable) is displayed by `grade`.

```
proc report data=demo nowd box;
column grade points=totalpoints;
define grade/group width=5;
define points/n;
define totalpoints/sum 'Total Points';
```

grade	Total Points
3	300
4	400
6	600

In the next example, more descriptive statistics are computed for the `points` variable by `grade`.

```
proc report nowd box;
column grade points, (n mean min max);
define grade/group;
define points/'Points';
define n/'N' width=3;
define mean/'Mean';
define std/'SD';
define min/'Min';
define max/'Max';
```

grade	N	Points		
		Mean	Min	Max
3	2	150	100	200
4	1	400	400	400
6	1	600	600	600

The next example produces the total points per combination of `grade` and `gender` values.

```
proc report data=demo nowd split='*' box;
column grade gender, points;
define grade/group;
define gender/across center;
define points/analysis sum 'Total*Points';
```

grade	gender	
	F	M
Total Points	Total Points	Total Points
3	300	
4		400
6	600	

REFERENCES

Cranford, Keith. 2008. Learning PROC REPORT by Comparison. *Proceeding of the SAS Global Forum 2008 Annual Conference*. SGF, San Antonio, TX. Paper 170-2008.

SAS Institute Inc. 2012. *Base SAS® 9.3 Procedures Guide, Second Edition*. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Imelda C. Go, PhD
Office of Assessment
Department of Education
1429 Senate Street
Columbia, SC 29201

Abbas S. Tavakoli, DrPH, MPH, ME
College of Nursing
University of South Carolina
1601 Greene Street
Columbia, SC 29208-4001
Tel: (803) 777-2978
Fax: (803) 777-5561
E-mail: abbas.tavakoli@sc.edu

TRADEMARK NOTICE

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.