Using SAS® to Produce Report-Ready Summaries of Likert-Type Survey Data: PROC TABULATE, Output Delivery System, PROC TEMPLATE

Imelda C. Go, South Carolina Department of Education, Columbia, SC

ABSTRACT

Beginners in SAS can use simple methods, such as the DATA Step, PROC FREQ, and PROC MEANS, to compute the summary statistics for their survey data. However, the form in which default SAS output appears is typically not compatible with the appearance of formal reports. In the worst-case scenario, one can get the required statistics from the SAS output and type them manually into the report. This paper shows how, with data set manipulation, a programmer can use PROC TABULATE, the Output Delivery System, and PROC TEMPLATE to produce customized report-ready output in rich text format (RTF). It is possible to produce RTF output that can be included in formal reports with little or no revision.

Note: The sample code in this paper was tested using SAS Version 9.1.2.

Programmers who are new to SAS often analyze, summarize, and report survey data using the DATA Step, PROC FREQ, and PROC MEANS. Let us consider the following three-item Likert-type survey as the example for this paper. In addition to the survey responses, the gender of the respondents was also collected.

Survey Item	SD	D	N	A	SA
1. I know about the issues surrounding the *Great Pacific Garbage Patch*.	☐	☐	☐	☐	☐
2. I am concerned about the amount of floating plastic debris in our oceans.	☐	☐	☐	☐	☐
3. The poor condition of our oceans will have an adverse effect on the entire planet.	☐	☐	☐	☐	☐

Note: SD = *strongly disagree* coded as 1, D = *disagree* coded as 2, N = *neutral* coded as 3, A = *agree* coded as 4, SA = *strongly disagree* coded as 5

Let us use the following as our sample SAS data set named `surveydata`.

ID	Gender	Item1	Item2	Item3
1	M	4	3	4
2	M	3	3	5
3	M	2	5	5
4	F	1	2	3
5	F	2	4	4
6	F	5	5	5

Let us suppose your task is to create summary tables that look like this:

Table 1. Summary of Survey Responses from All Respondents (N=6)

	Responses					Negative Response (Disagree or Strongly Disagree)	Positive Response (Agree or Strongly Agree)	Response	
	1 SD	2 D	3 NO	4 A	5 SA				
	%	%	%	%	%	%	%	Mean	SD
Survey Item									
1. I know about the issues surrounding the Great Pacific Garbage Patch.	16.7	33.3	16.7	16.7	16.7	50.0	33.3	2.8	1.5
2. I am concerned about the amount of floating plastic debris in our oceans.	0.0	16.7	33.3	16.7	33.3	16.7	50.0	3.7	1.2
3. The poor condition of our oceans will have an adverse effect on the entire planet.	0.0	0.0	16.7	33.3	50.0	0.0	83.3	4.3	0.8

Table 2. Summary of Survey Responses from Males (N=3)

	Responses					Negative Response (Disagree or Strongly Disagree)	Positive Response (Agree or Strongly Agree)	Response	
	1 SD	2 D	3 NO	4 A	5 SA				
	%	%	%	%	%	%	%	Mean	SD
Survey Item									
1. I know about the issues surrounding the Great Pacific Garbage Patch.	0.0	33.3	33.3	33.3	0.0	33.3	33.3	3.0	1.0
2. I am concerned about the amount of floating plastic debris in our oceans.	0.0	0.0	66.7	0.0	33.3	0.0	33.3	3.7	1.2
3. The poor condition of our oceans will have an adverse effect on the entire planet.	0.0	0.0	0.0	33.3	66.7	0.0	100.0	4.7	0.6

Table 3. Summary of Survey Responses from Females (N=3)

	Responses					Negative Response (Disagree or Strongly Disagree)	Positive Response (Agree or Strongly Agree)	Response	
	1 SD	2 D	3 NO	4 A	5 SA				
	%	%	%	%	%	%	%	Mean	SD
Survey Item									
1. I know about the issues surrounding the Great Pacific Garbage Patch.	33.3	33.3	0.0	0.0	33.3	66.7	33.3	2.7	2.1
2. I am concerned about the amount of floating plastic debris in our oceans.	0.0	33.3	0.0	33.3	33.3	33.3	66.7	3.7	1.5
3. The poor condition of our oceans will have an adverse effect on the entire planet.	0.0	0.0	33.3	33.3	33.3	0.0	66.7	4.0	1.0

SIMPLE METHOD

We can use the following simple code to produce the numbers needed for the three tables:

SAS CODE	COMMENTS

```proc format;```
```value likert 1='1 SD'```
```   2='2 D'```
```   3='3 N'```
```   4='4 A'```
```   5='5 SA';```
```value grouped 1,2='A. Disagreed'```
```   3='C. Neutral'```
```   4,5='B. Agreed';```
	The labels defined in PROC FORMAT are used to make it easier to read the SAS output. The `likert.` format shows the numeric code assigned to each type of response. The `grouping.` format groups the D and SD responses together and groups the A and SA responses together.

```proc freq data=surveydata order=formatted;```
```   tables gender*(item1-item3)/nocol;```
```   format item1-item3 likert.;```
 	
The FREQ Procedure
 	
Table of Gender by Item1
 	
Gender   Item1
 	
Frequency / Percent / Row Pct
 	
		1 SD	2 D	3 N	4 A	5 SA	Total

F	1 / 16.67 / 33.33	1 / 16.67 / 33.33	0 / 0.00 / 0.00	0 / 0.00 / 0.00	1 / 16.67 / 33.33	3 / 50.00
M	0 / 0.00 / 0.00	1 / 16.67 / 33.33	1 / 16.67 / 33.33	1 / 16.67 / 33.33	0 / 0.00 / 0.00	3 / 50.00
Total	1 / 16.67	2 / 33.33	1 / 16.67	1 / 16.67	1 / 16.67	6 / 100.00 | Corresponding tables for item2 and item3 are similar to the above table.	The ORDER= option makes sure the responses appear in the output according to the alphabetical order of the formatted responses. The NOCOL option suppresses the column percentages, which are not needed. The output provides the frequency distribution of the responses per item and by gender. The `likert.` format was applied to the item responses so that it is clear to the reader what the numeric codes mean.

```proc freq data=surveydata order=formatted;```
```   tables gender*(item1-item3)/nocol;```
```   format item1-item3 grouped.;```
 	
The FREQ Procedure
 	
Table of Gender by Item1
 	
Gender   Item1
 	
Frequency / Percent / Row Pct
 	
		A. Disagreed	B. Agreed	C. Neutral	Total

F	2 / 33.33 / 66.67	1 / 16.67 / 33.33	0 / 0.00 / 0.00	3 / 50.00
M	1 / 16.67 / 33.33	1 / 16.67 / 33.33	1 / 16.67 / 33.33	3 / 50.00
Total	3 / 50.00	2 / 33.33	1 / 16.67	6 / 100.00 | Corresponding tables for item2 and item3 are similar to the above table.	This is similar to the preceding PROC FREQ statement except that the `grouped.` format was applied to the item responses to get the percentages.

SAS CODE	COMMENTS
```proc means data=surveydata maxdec=1 n mean std; var item1-item3;``` The MEANS Procedure 	Variable	N	Mean	Std Dev
---	---	---	---
Item1	6	2.8	1.5
Item2	6	3.7	1.2
Item3	6	4.3	0.8		The N, MEAN, and STD options limit the statistics to the three specified. The MAXDEC=1 option makes the means and standard deviations appear with one decimal place.
```proc means data=surveydata maxdec=1 n mean std; class gender; var item1-item3;``` The MEANS Procedure 	Gender	N Obs	Variable	N	Mean	Std Dev
---	---	---	---	---	---
F	3	Item1	3	2.7	2.1
		Item2	3	3.7	1.5
		Item3	3	4.0	1.0
M	3	Item1	3	3.0	1.0
		Item2	3	3.7	1.2
		Item3	3	4.7	0.6		This is similar to the preceding PROC MEANS except that it computes the statistics by gender. The N, MEAN, and STD options limit the statistics to the three specified. The MAXDEC=1 option makes the means and standard deviations appear with one decimal place.

The simple code produces the desired statistics, but it is obvious that it can be a cumbersome task to manually put these numbers into the desired report format. This type of manual work creates the opportunity for human error. Fortunately, there is an alternative. With the combined use of data set manipulation, PROC TABULATE, and the Output Delivery, it is possible to produce rich text format (RTF) output that can be included in formal reports with little or no revision.

ALTERNATIVE METHOD

The first step is to manipulate the data. Since there are three survey items, there will be three data sets created with certain properties. The first one will only have survey responses from the first item, and so on.

SAS CODE	COMMENTS
``%macro dsets(dsname,item,stem);
data &dsname;
    set all (rename=(&item=response));
    item="&stem";
    if response in (1,2) then disagreed=100;
                          else disagreed=0;
    if response in (4,5) then agreed=100;
                          else agreed=0;
    response2=response;
    keep disagreed agreed response2 response item gender;
%mend dsets;

%dsets(dsname=I1,item=ITEM1,stem=%str(1. I know about the issues surrounding the Great Pacific Garbage Patch.));

%dsets(dsname=I2,item=ITEM2,stem=%str(2. I am concerned about the amount of floating plastic debris in our oceans.));

%dsets(dsname=I3,item=ITEM3,stem=%str(3. The poor condition of our oceans will have an adverse effect on the entire planet.));

data manipulated;
    set i1 i2 i3;``` | The macro creates a data set that corresponds to each survey item. The dsname macrovariable is the data set name. The item macrovariable is the survey item variable that requires a corresponding manipulated data set. The stem macrovariable is the survey item's text.

The item variable contains the stem or text for the item. The disagreed and agreed variables are 100/0 indicators of whether a response was negative or positive respectively. When averaged, they produce the percentage of responses that were negative or positive. The response2 variable contains the same value as the response variable. For the sake of illustrating the minimum data required by PROC TABULATE to produce the desired result, only the five variables listed were kept for each data set.

The three data sets are then combined to form one data set that will be processed using PROC TABULATE. |

Data Set I1

Obs	Gender	response	item	disagreed	agreed	response2
1	M	4	1. I know about the issues surrounding the Great Pacific Garbage Patch.	0	100	4
2	M	3	1. I know about the issues surrounding the Great Pacific Garbage Patch.	0	0	3
3	M	2	1. I know about the issues surrounding the Great Pacific Garbage Patch.	100	0	2
4	F	1	1. I know about the issues surrounding the Great Pacific Garbage Patch.	100	0	1
5	F	2	1. I know about the issues surrounding the Great Pacific Garbage Patch.	100	0	2
6	F	5	1. I know about the issues surrounding the Great Pacific Garbage Patch.	0	100	5

Data Set I2

Obs	Gender	response	item	disagreed	agreed	response2
1	M	3	2. I am concerned about the amount of floating plastic debris in our oceans.	0	0	3
2	M	3	2. I am concerned about the amount of floating plastic debris in our oceans.	0	0	3
3	M	5	2. I am concerned about the amount of floating plastic debris in our oceans.	0	100	5
4	F	2	2. I am concerned about the amount of floating plastic debris in our oceans.	100	0	2
5	F	4	2. I am concerned about the amount of floating plastic debris in our oceans.	0	100	4
6	F	5	2. I am concerned about the amount of floating plastic debris in our oceans.	0	100	5

Data Set I3

Obs	Gender	response	item	disagreed	agreed	response2
1	M	4	3. The poor condition of our oceans will have an adverse effect on the entire planet.	0	100	4
2	M	5	3. The poor condition of our oceans will have an adverse effect on the entire planet.	0	100	5
3	M	5	3. The poor condition of our oceans will have an adverse effect on the entire planet.	0	100	5
4	F	3	3. The poor condition of our oceans will have an adverse effect on the entire planet.	0	0	3
5	F	4	3. The poor condition of our oceans will have an adverse effect on the entire planet.	0	100	4
6	F	5	3. The poor condition of our oceans will have an adverse effect on the entire planet.	0	100	5

The second step is to make the report-appearance specifications using PROC TEMPLATE.

SAS CODE	COMMENTS

``
proc template;
define style temporary;
PARENT=styles.default;
replace color_list/
'bgA4' = cxFFFFFF
'bgA3' = cxFFFFFF
'bgA2' = cxFFFFFF
'bgA1' = cxFFFFFF
'bgA' = cxFFFFFF
'fgB2' = cx0066AA
'fgB1' = cx004488
'fgA4' = cxAAFFAA
'fgA2' = cx0033AA
'fgA1' = cx000000
'fgA' = cx002288;
replace font
"Fonts used in the default style" /
'TitleFont3' = ("Arial",12pt,Italic)
'TitleFont' = ("Arial",12pt)
'StrongFont' = ("Arial",7pt)
'EmphasisFont' = ("Arial",7pt)
'FixedEmphasisFont' = ("Arial,Arial",7pt)
'FixedStrongFont' = ("Arial,Arial",7pt)
'FixedHeadingFont' = ("Arial,Arial",7pt)
'BatchFixedFont' = ("Arial, Arial",7pt)
'FixedFont' = ("Arial,Arial",7pt)
'headingEmphasisFont' = ("Arial",7pt)
'headingFont' = ("Arial",7pt)
'docFont' = ("Arial",7pt);
replace Body from Document
"Controls the Body file." /
bottommargin = 1.0in
topmargin = 1.0in
rightmargin = 1.0in
leftmargin = 1.5in;
end;
``
	The style name is temporary. This style is based on the default parent style, which has table cell shading in a number of cells. The shading was removed by using the specified color list. The default fonts and font sizes were replaced by using the values listed. The page margins were altered to the values listed.
	
	Note: Only a few of the many customizable report features using PROC TEMPLATE are mentioned in this example.

The third step is to create the tables using PROC TABULATE and channel the output to ODS RTF.

SAS CODE	COMMENTS
| ```
proc format;
value likert 1='1 SD' 2='2 D' 3='3 N' 4='4 A' 5='5 SA';

%macro maketable(wherecondition,title);
proc tabulate data=manipulated order=formatted;

class item;
class response/preloadfmt;

var agreed disagreed response2;

table item="Survey Item",
response="Responses"*(rowpctn="%"*f=6.1)
disagreed="Negative Response (Disagree or  Strongly
Disagree)"*(mean="%"*f=6.1)
agreed="Positive Response (Agree or Strongly
Agree)"*(mean="%"*f=6.1)
response2="Response"*(mean="Mean"*f=3.1 std="SD"*f=3.1) /
printmiss rts=50;
format response likert.;
&WHERECONDITION
title "&title";

%mend maketable;

ods rtf file="c:\three summary tables.rtf"
        style=temporary bodytitle;

%maketable(wherecondition=%str(),title=%str(Table 1.
Summary of Survey Responses from All Respondents (N=6)));

%maketable(wherecondition=%str(where
gender='M';),title=%str(Table 2. Summary of Survey
Responses from Males (N=3)));

%maketable(wherecondition=%str(where
gender='F';),title=%str(Table 3. Summary of Survey
Responses from Females (N=3)));
run;
ods rtf close;

proc template;
delete temporary/store=sasuser.templat;
``` | The user-defined format `likert.` is defined using PROC FORMAT.

The macro processes the data set `manipulated` with PROC TABULATE. The `wherecondition` value is the WHERE statement that isolates the subset of interest (i.e., all students, all males, and all females in this example). The `title` macrovariable is the table's title.

The ORDER=FORMATTED option makes the responses appear in the output according to the alphabetical order of the formatted responses.

The two CLASS variables are `item` and `response`. The PRELOADFMT was used with the `response` class variable. This specifies that the order in which the values appear in the output will be the alphabetical order they have in the user-defined format `likert.` The MISSING option was excluded from both CLASS statements to exclude observations with a missing value for any class variable.

The analysis variables are `agreed`, `disagreed`, and `response2`. Within the same PROC TABULATE specification, a variable cannot be listed both as a CLASS and as an analysis (VAR) variable. This is why `response2` has the same value as `response`. `Response` is used for the frequency distribution and `response2` is used to compute the response mean and standard deviation.

Each component in the TABLE statement is given a custom label and given a numeric format.

The macro is invoked within ODS RTF three times (first for all students, second for all males, and third for all females) to create the tables. The style used is the `temporary` style defined using PROC TEMPLATE. The BODYTITLE puts the titles into the body of the document and not the header.

After ODS RTF processing is closed, the style is deleted. |

CONCLUSION

The paper shows an example of a solution that uses a combination of techniques to reduce human error by producing custom reports automatically. The data set was manipulated to take advantage of PROC TABULATE processing. The PROC TABULATE output was routed through ODS and its appearance customized with PROC TEMPLATE to produce reports in RTF format that most resemble the desired final product. It may not always be possible to produce reports exactly as desired, but SAS provides several tools that enable the programmer to customize the appearance of the report. Since the data processing was repetitive, the code was also repetitive. Therefore, SAS macros were incorporated to make the code wieldier.

REFERENCES

Haworth, Lauren E. 2001. *Output Delivery System: The Basics.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2002. *Base SAS® 9 Procedures Guide.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2002. *SAS® 9 Output Delivery System User's Guide.* Cary, NC: SAS Institute Inc.

TRADEMARK NOTICE