

“Zero My Hero!”
Getting SAS® to put those Zeros
onto your PROC REPORT
or PROC TABULATE

Anna Vincent
Center for Health Statistics (CHS),
Texas Department of State Health Services (DSHS)

Abstract

This paper will demonstrate how to put zeros, null, or blank values into a PROC REPORT or a PROC TABULATE.

When you use a plain PROC REPORT or a PROC TABULATE for some crosstabs, if a value is zero or blank, that value will not show by default. What if you need it to show?

Introduction

In my past work, there were many times when we needed to tabulate telephone survey codes by the month they were recorded. However, not all codes happened in every month. At times, these codes don't happen at all. We still needed them on the report to see when and where they did show up.

I wanted to use either the PROC REPORT or PROC TABULATE to get my data into a crosstab table display. When I first tried them, I did not get the tables I wanted. The tables did not include zeros, null, and blank values. Below is how I got SAS to include these values.

Using PROC REPORT:

If one runs the codes by month on just the plain PROC REPORT, by default, SAS will not produce any outcome that has all blanks. This is because SAS does not by default produce a value in a cross tab that is not in the data set.

Just by adding a few simple keywords, we can get PROC REPORT to provide the values with zeros and in addition, we can make SAS put the variables in any order that we want.

The first thing to do is add values and order using the PROC FORMAT procedure:

```
proc format;  
value monf      1 = 'JAN'
```

```

2 = 'FEB'
3 = 'MAR'
4 = 'APR'
5 = 'MAY'
6 = 'JUN'
7 = 'JUL'
8 = 'AUG'
9 = 'SEP'
10 = 'OCT'
11 = 'NOV'
12 = 'DEC' ;

value dispf
1100 = '1100, Complete'
1200 = '1200, Partial Complete'
2111 = '2111, HH level refusal (LL only)'
2112 = '2112, Known respondent refusal'
2120 = '2120, Break off/ termination within questionnaire'
2210 = '2210, Respondent never available'
2220 = '2220, HH answering device (LL only)'
2330 = '2330, Language barrier, selected respondent'
~~more code~~
4700 = '4700, HH, no eligible respondent'
4900 = '4900, Miscellaneous, non-eligible' ;

run;

```

Then by adding these three options, COMPLETEROWS, PRELOADFMT, and FORMAT= into a PROC REPORT statement, we can get SAS to produce a report complete with zeros!

```

proc report data=present.landmain split = "\" nowd spacing=3
      headline COMPLETEROWS;

Columns studymonth dispcode N;

define studymonth / group order=data 'Month' right PRELOADFMT FORMAT=monf. ;
define dispcode / group order=data 'Disp. Code ' left PRELOADFMT FORMAT=dispf.
;
define N / 'Monthly\Totals' right;

rbreak after / summarize;

title1 j=c "DispCode by Month for &cy, Example Two";
title2 j=c "Last Run on this date: &fdate";
run;

```

In the above example, there are two group variables, studymonth and dispcode. In order to have all crosstab values appear, we use the COMPLETEROWS option to display all possible combinations of the values of the group variables, even if one or more of the combinations do not occur in the input data set.

PRELOADMT specifies to preload the format for that particular variable. This applies to group and across variables. It has no effect unless you use either the EXCLUSIVE or ORDER=DATA command and you must assign a format to that variable.

- EXCLUSIVE excludes all combinations of values that are not found in the preloaded range of user-defined formats. You might want all the values even if they are not in the formats. It's a good check to see if you formatted your variables correctly.
- ORDER=DATA merely orders the values as they are in the input data set. So sometimes you might want to do a PROC SORT before you run your report.

And here are the results:

The “before” output, with the codes of d4460, d4470, and d4900 missing:

DispCode by Month for 2013, Example One
Last Run on this date: July 31, 2013

Month	DISPCODE	Monthly Totals
JAN	1100, Complete	592
	1200, Partial Complete	43
	2111, HH level refusal (LL only)	69
	2112, Known respondent refusal	69
	2120, Break off/ termination within questionnaire	54
	2210, Respondent never available	110
	2320, Respondent physically or mentally unable to complete interview	46
	2330, Language barrier, selected respondent	3
	3100, Unknown if housing unit	1471
	3130, No answer	1
	3150, Telecommunication barrier	150
	3200, HH, not known if respondent eligible	20
	3322, Physical or mental impairment (HH level)	12
	3330, Language barrier, (HH level)	11
FEB	4100, Out of sample	12
	4200, Fax/data/modem	372
	4300, Nonworking number/disconnected	5849
	4400, Special technological circumstances	613
	4450, Cell Phone (LL only)	18
	4500, Non-residence	783
	4510, Group home	19
	4700, HH, no eligible respondent	4
	1100, Complete	537
	1200, Partial Complete	35
	2111, HH level refusal (LL only)	69
	2112, Known respondent refusal	72

And the “after” output. Notice that the desired codes with the zeros are included!

DispCode by Month for 2013, Example Two
Last Run on this date: July 31, 2013

Month	Disp. Code	Monthly Totals
JAN	1100, Complete	592
	1200, Partial Complete	43
	2111, HH level refusal (LL only)	69
	2112, Known respondent refusal	69
	2120, Break off/ termination within questionnaire	54
	2210, Respondent never available	110
	2220, HH answering device (LL only)	0
	2320, Respondent physically or mentally unable to complete interview	46
	2330, Language barrier, selected respondent	3
	3100, Unknown if housing unit	1471
	3130, No answer	1
	3140, Answering device, unk. if residence or respondent eligible	0
	3150, Telecommunication barrier	
	3200, HH, not known if respondent eligible	
	3322, Physical or mental impairment	
FEB	2230, Language barrier, selected respondent	372
	3100, Unknown if housing unit/disconnected	5849
	4400, Special technological circumstances	613
	4430, Call forwarding/pager	0
	4450, Cell Phone (LL only)	18
	4460, Landline (CP only)	0
	4470, Cell Phone respondent with landline	0
	4500, Non-residence	783
	4510, Group home	19
	4700, HH, no eligible respondent	4
	4900, Miscellaneous, non-eligible	0
	1100, Complete	537

Using PROC TABULATE:

Now with PROC TABULATE we can have a little more fun! You will again need to run the PROC FORMAT syntax before, you can use this SAS code.

```
proc tabulate data=present.landmain format=8.2 CLASSDATA=present.landmain
EXCLUSIVE;
  class studymonth dispcode / PRELOADFMT EXCLUSIVE;
  table dispcode all="Monthly Totals",
        studymonth='Months' all="Up to Date Disp. Code for &cy" /
        PRINTMISS BOX='Disp. Codes' MISSTEXT='Hand over the
Chocolate!';
  format studymonth monf. dispcode dispf.;
  title "DispCode by Month for &cy, Example Two";
  title2 "Last Run on this date: &fdate";
```

```
run;
```

PROC TABULATE is a little more complicated and some things are better done in PROC REPORT. They both have their pros and cons, so you might want to run both just to see which is best for you.

The CLASSDATA= specifies that a data set that contains the **combinations of values of the class variables** that must be present in the output. Any combinations of values of the class variables that occur in the CLASSDATA= data set but not in the input data set, appear in each table or output data set and have a frequency of zero.

The CLASSDATA= data set must contain all class variables and their data type and format must match the corresponding class variables in the input data set. If you use the EXCLUSIVE option, then PROC TABULATE excludes any observations in the input data set whose combinations of values of class variables are not in the CLASSDATA= data set. You can use the CLASSDATA= data set to filter or supplement the input data set. The CLASSDATA= does not have to be your input data set. They can be different, but the variables must match in type and data attributes between the two data sets.

The EXCLUSIVE keyword excludes from the table and output data sets all combinations of the class variables values that are NOT found in the CLASSDATA= data set.

PRINTMISS in the TABLE section just tells SAS to print all the values that occur for a CLASS variable every time, even if there are no data for some of the cells that are created. However, if an entire page has only missing values, that page will not be printed whether you use PRINTMISS or not.

BOX= will give you a title for the top left hand cell of the table. You should use single quotes unless you're wanting to use it in a macro. In a macro, use double quotes.

Finally, MISSTEXT= will let you put whatever you want for all the cells in the table that are missing a value.

The following are the results from running the same cross tabulation as we did with PROC REPORT, using PROC TABULATE but with only default options:

DispCode by Month for 2013, Example One
Last Run on this date: July 31, 2013

		Months					Up to Date Disp. Code for 2013
		JAN	FEB	MAR	APR	MAY	
DISPCODE	Sum	39809291	39472985	33583990	39521912	36188451	1.8858E8
Monthly Totals	N	10321	10312	8854	10367	9549	49403

This is not at all what we wanted! By adding the above options in red, the results will include this additional information. Here is what it looks like:

DispCode by Month for 2013, Example Two
Last Run on this date: July 31, 2013

Disp. Codes	Months												Up to Date Disp. Code for 2013
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	
	N	N	N	N	N	N	N	N	N	N	N	N	
DISPCODE													
1100, Complete	592	537	480	534	548	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	2691
1200, Partial Complete	43	35	30	51	51	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	210
2111, HH level refusal (LL only)	69	69	56	50	47	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	291
2112, Known respondent refusal	69	72	89	108	91	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	429
2120, Break off/ termination within questionnaire	54	38	31	41	37	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	201
2210, Respondent never available	110	115	80	107	115	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	527
2220, HH answering device (LL only)	Hand over the Chocolate!	421	550	564	530	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	2065
2320, Respondent physically or mentally unable to complete interview	46	57	15	19	27	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	Hand over the Chocolate!	
2330, Language barrier, selected respondent	3	7	9	6		Hand over the Chocolate!							

What was a blank upper left hand cell now reads “Disp. Codes” with the use of “BOX=”. Anything that is missing in the table now reads “Hand over the Chocolate!”, which was done nicely with the “MISSTEXT=” option.

Conclusion

By including a few additional options in PROC REPORT and PROC TABULATE, you can ensure SAS shows values in reports by showing zeroes when there are no records with that value in the data file. With this new knowledge, you can keep all your values that you need to show in your reports, show the places that are missing and even add a little more description to your tables.

Play around with SAS and see what you can discover!

Contact Information

Anna Vincent
Program Specialist
Center for Health Statistics, MC-1898
Texas Department of State Health Services
PO Box 149317
Austin, TX 78714-9347
512-776-2724 work
512-458-3255 fax

Emails:

Anna.vincent@dshs.texas.gov

For Texas Vital Events Data:

VSTAT@dshs.texas.gov

For Website:

<http://www.dshs.texas.gov/chs/chs/default.shtm>



The authors of the paper/presentation have prepared these works in the scope of their employment with the Texas Department of State Health Services and the copyrights to these works are held by the Center for Health Statistics.

Therefore, Anna Vincent hereby grants to SCSUG Inc a non-exclusive right in the copyright of the work to the SCSUG Inc to publish the work in the Publication, in all media, effective if and when the work is accepted for publication by SCSUG Inc.

This the 1st day of September, 2017.