

Paper 78-27

Let Summary Sum and Tabulate Format

Marianne Whitlock, Westat, Rockville, MD

Abstract

Occasionally you want a report formatted the way SAS® PROC TABULATE formats data, but you want special rows or columns and a special layout, which are difficult or impossible to do directly with TABULATE. A solution is to use PROC SUMMARY, followed by a data step to prepare the data for TABULATE. This paper uses a simple example, to demonstrate a method for doing all the calculating before using TABULATE, and using the resulting data set as input to TABULATE. You can use the method in many variations.

Introduction

TABULATE produces clear and readable output when you want counts by class variables, with margin totals, grand total, and percentages. However, in the example in this paper, all the numbers are pre-calculated for TABULATE, including the cell percentages of one analysis variable of another. Then TABULATE is used for formatting.

The Method

The method consists of four steps:

- 1) Create variables with values of 1 and 0, meaning True and False respectively.
- 2) Use PROC SUMMARY to sum them at all levels (0 to NWAY).
- 3) Use a Data Step to calculate the percentages wanted at all levels.
- 4) Use PROC TABULATE to format the data set.

Step 1: Create 0/1 variables

The example shown in this paper uses a data set, PondData. It has class variables: Plant (giving the name of a plant) and Pond (giving the name of a pond); and analysis variables: Requested, Received, and Photo. These three variables have values of 1(True) or 0(False). There are no records with Received =1 (True) unless Requested = 1, and no records with Photo=1 unless Received and Requested are both 1. Thus, for example, the sum of Received divided by the sum of Requested, when multiplied by 100 is the percentage that Received is of Requested. Some records in PondData are:

Plant	Pond	Req.	Rec.	Pho.
Gentian	Flax	1	1	0
Gentian	Flax	1	1	1
Gentian	Higgins	1	0	0
Gentian	Higgins	1	1	1
Gentian	Ruth	1	0	0
Gentian	Ruth	1	0	0

Laurel	Cliff	1	1	0
Laurel	Cliff	1	0	0
Laurel	Flax	1	1	0
Laurel	Higgins	1	1	1
Sweet Pepper	Higgins	1	1	1
Sweet Pepper	Higgins	1	1	1

Here is the code for steps 2 to 4.

```

/*Step 2: Sum 0/1 variables*/
proc summary data = Tab.PondData ;
  class Plant Pond ;
  var Requested Received Photo ;
  output out = Sumry sum= ;
run;

/*Step 3: Create Percentages*/
data TabData ;
  set Sumry ;
  if Plant = "" then Plant = "Total";
  if Pond = "" then Pond = "Total";
  PCT_R_Rq =
    (Received/Requested) * 100 ;
  PCT_P_Rq =
    (Photo/Requested) * 100 ;
  if Received > 0
    then PCT_P_Re =
      (Photo/Received) * 100 ;
  else PCT_P_Re = 0 ;
run;

/*Step 4: Format report*/
title 'Requests for Plant Specimens';
proc tabulate data = Tabdata;
  class Plant Pond ;
  var Received Requested Pct_R_Rq
    Photo Pct_P_Rq Pct_P_Re ;
  tables
    (Requested*f=8. Received*f=8.
    Pct_R_Rq*f=8.2 Photo*f=8.
    Pct_P_Re*f=8.2 Pct_P_Rq*f=8.2)
  * Plant * sum="", Pond
  / row=float rts=40
  box="Counts and Percents";
label
Requested=
  "Number of Specimens Requested"
Received=
  "Number of Specimens Received"
Photo=
  "Number of Specimens Photographed"

```

```

Pct_R_Rq =
  "Percent Received of Requested"
Pct_P_Re =
  "Percent Photographed of Received"
Pct_P_Rq=
  "Percent Photographed of Requested"
;
run;

```

The data set produced by PROC SUMMARY and Data Step is shown below, and, on the last page the PROC TABULATE output is shown.

You will more quickly gain insight into how the method works, and be able to adapt it for your own use, if you remember that TABULATE does not calculate the "Total" rows and columns. SUMMARY does that, enabling you to calculate percentages in the data step, on totals, with the same assignment statement that calculates individual cell percentages. TABULATE sees "Total" as just another value of Plant or Pond.

Formats may be needed

The example in this paper had all other values of class variables start with letters before "T" to force "Total" to the last column and to each of the last rows of the concatenated variables. In the real world, you might have to use formatted numeric variables, assigning a number higher than all values in the data set to the variables with missing values output by SUMMARY. These missing values indicate which class variables the data are summed over, so changing a missing value to "Total" is entirely appropriate.

The author can be contacted by mail:

Marianne Whitlock
 Westat
 1650 Research Boulevard
 Rockville, MD 20850

or e-mail:

MarianneWhitlock@Westat.com

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Data Set from PROC SUMMARY and Data Step (After Step 3):

Plant	Pond	TYPE	FREQ	Requested	Received	Photo	PCT_R_Rq	PCT_P_Rq	PCT_P_Re
Total	Total	0	287	287	103	75	35.89	26.13	72.82
Total	Cliff	1	50	50	14	2	28	4	14.29
Total	Flax	1	100	100	15	8	15	8	53.33
Total	Higgins	1	100	100	63	63	63	63	100
Total	Ruth	1	37	37	11	2	29.73	5.41	18.18
Gentian	Total	2	164	164	58	45	35.37	27.44	77.59
Laurel	Total	2	80	80	30	17	37.5	21.25	56.67
Sweet Pepper	Total	2	43	43	15	13	34.88	30.23	86.67
Gentian	Cliff	3	19	19	7	2	36.84	10.53	28.57
Gentian	Flax	3	73	73	11	6	15.07	8.22	54.55
Gentian	Higgins	3	60	60	36	36	60	60	100
Gentian	Ruth	3	12	12	4	1	33.33	8.33	25
Laurel	Cliff	3	24	24	7	0	29.17	0	0
Laurel	Flax	3	19	19	4	2	21.05	10.53	50
Laurel	Higgins	3	26	26	15	15	57.69	57.69	100
Laurel	Ruth	3	11	11	4	0	36.36	0	0
Sweet Pepper	Cliff	3	7	7	0	0	0	0	0
Sweet Pepper	Flax	3	8	8	0	0	0	0	0
Sweet Pepper	Higgins	3	14	14	12	12	85.71	85.71	100
Sweet Pepper	Ruth	3	14	14	3	1	21.43	7.14	33.33

Requests for Plant Specimens

Counts and Percents		Pond				
		Cliff	Flax	Higgins	Ruth	Total
Number of Specimens Requested	Plant					
	Gentian	19	73	60	12	164
	Laurel	24	19	26	11	80
	Sweet Pepper	7	8	14	14	43
	Total	50	100	100	37	287
Number of Specimens Received	Gentian	7	11	36	4	58
	Laurel	7	4	15	4	30
	Sweet Pepper	0	0	12	3	15
	Total	14	15	63	11	103
Percent Received of Requested	Gentian	36.84	15.07	60.00	33.33	35.37
	Laurel	29.17	21.05	57.69	36.36	37.50
	Sweet Pepper	0.00	0.00	85.71	21.43	34.88
	Total	28.00	15.00	63.00	29.73	35.89
Number of Specimens Photographed	Gentian	2	6	36	1	45
	Laurel	0	2	15	0	17
	Sweet Pepper	0	0	12	1	13
	Total	2	8	63	2	75
Percent Photographed of Received	Gentian	28.57	54.55	100.00	25.00	77.59
	Laurel	0.00	50.00	100.00	0.00	56.67
	Sweet Pepper	0.00	0.00	100.00	33.33	86.67
	Total	14.29	53.33	100.00	18.18	72.82
Percent Photographed of Requested	Gentian	10.53	8.22	60.00	8.33	27.44
	Laurel	0.00	10.53	57.69	0.00	21.25
	Sweet Pepper	0.00	0.00	85.71	7.14	30.23
	Total	4.00	8.00	63.00	5.41	26.13