SAS 5581-2016

USING PROC TABULATE AND LAG(n) FUNCTION FOR RATES OF CHANGE

Khoi To, Office of Planning and Decision Support, Virginia Commonwealth University

ABSTRACT

For SAS® users, PROC TABULATE and PROC REPORT (and its compute blocks) are probably among the most common procedures for calculating and displaying data. It is, however, pretty difficult to calculate and display changes from one column to another using data from other rows with just these two procedures. Compute blocks in PROC REPORT can calculate additional columns, but it would be challenging to pick up values from other rows as inputs. This presentation shows how PROC TABULATE can work with the lag(n) function to calculate rates of change from one period of time to another. This offers the flexibility of feeding into calculations the data retrieved from other rows of the report. PROC REPORT is then used to produce the desired output. The same approach can also be used in a variety of scenarios to produce customized reports.

INTRODUCTION

First of all, let's have a look at the final report. The lines "Headcount Growth" and "%Growth over prior Fall" (highlighted in yellow) display the year-on-year changes in absolute numbers and percentages, respectively.

Student Enrollment by Major

MCV-Campus Schools												
	Fall 2009			Fall 2010			Fall 2011			Fall 2012		
	In State	Out State	Total	In State	Out State	Total	In State	Out State	Total	In State	Out State	Total
Allied Health Professions												
Baccalaureate	165	12	177	176	10	186	176	8	184	183	12	195
Masters	390	154	544	391	167	558	385	163	548	388	141	529
Doctoral	249	90	339	234	95	329	213	78	291	209	73	282
Grad. Post-Bacc Certs	10	2	12	5	2	7	9	.	9	6	2	8
Grad. Post-Masters Certs	20	4	24	18	2	20	25	3	28	23	7	30
All Degrees and Certs	**834**	**262**	**1,096**	**824**	**276**	**1,100**	**808**	**252**	**1,060**	**809**	**235**	**1,044**
Headcount Growth	.	.	.	-10	14	4	-16	-24	-40	1	-17	-16
% Growth over prior Fall	.	.	.	-1.2	5.3	0.4	-1.9	-8.7	-3.6	0.1	-6.7	-1.5
School of Dentistry												
Baccalaureate	57	3	60	64	2	66	59	3	62	55	1	56
Masters	10	22	32	11	24	35	6	27	33	7	26	33
First Professional	245	152	397	255	153	408	268	153	421	269	147	416
All Degrees and Certs	**312**	**177**	**489**	**330**	**179**	**509**	**333**	**183**	**516**	**331**	**174**	**505**
Headcount Growth	.	.	.	18	2	20	3	4	7	-2	-9	-11
% Growth over prior Fall	.	.	.	5.8	1.1	4.1	0.9	2.2	1.4	-0.6	-4.9	-2.1

…….(More schools/colleges here)…….

ALL DEGREE PROGRAMS												
Baccalaureate	845	45	890	886	40	926	856	36	892	763	35	798
Masters	735	284	1,019	720	302	1,022	711	299	1,010	726	266	992
First Professional	1,111	547	1,658	1,130	548	1,678	1,170	534	1,704	1,198	524	1,722
Doctoral	411	325	736	394	338	732	355	321	676	348	311	659
Grad. Post-Bacc Certs	90	48	138	63	35	98	63	25	88	70	31	101
Grad. Post-Masters Certs	29	4	33	25	2	27	40	4	44	40	8	48
All Degrees and Certs	**3,221**	**1,253**	**4,474**	**3,218**	**1,265**	**4,483**	**3,195**	**1,219**	**4,414**	**3,145**	**1,175**	**4,320**
Headcount Growth	.	.	.	-3	12	9	-23	-46	-69	-50	-44	-94
% Growth over prior Fall	.	.	.	-0.1	1	0.2	-0.7	-3.6	-1.5	-1.6	-3.6	-2.1

The above report can be achieved by the following steps:

1. Using proc tabulate to output aggregate data into data sets
2. Using the lag(n) function to set up the data so that year-on-year changes (in numbers and percentages) can be calculated
3. Using proc report to produce a report in the desired format

Let's look at each step in more detail below.

STEP 1: USING PROC TABULATE TO OUTPUT AGGREGATE DATA INTO DATA SETS

```
* pull out raw data and recode degree programs:

proc sql;
create table bio1 as
select distinct academic_period, academic_period_desc
              , person_uid, college1, student_level1
              , student_classification1, residency
 from odsstu.registration_bio
where schev_hrs > 0
  and census_period = '2'
  and substr(academic_period,5,2) = '10'
  and academic_period between '201010' and '201310'
  and college1 in ('AH','DN','MD','NR','PH') /* MCV campus only */
; quit;

data bio2; set bio1;
     if student_level1          = 'UG' then cat = 1; /* bacc */
else if student_level1          = 'PR' then cat = 3; /* first prof */
else if student_classification1 = 'GM' then cat = 2; /* masters */
else if student_classification1 = 'GC' then cat = 5; /* grad post-bacc */
else if student_classification1 = 'GP' then cat = 6; /* grad post-masters */
else cat = 4; /* doctoral */
run;

* output individual rows;

proc tabulate data=bio2 missing out=bio3;
class college1 cat academic_period residency;
table college1*cat all*cat, academic_period*(residency all);
run;
```

```
data bio3; set bio3;
if college1  = ' ' then college1  = 'UZ'; /* university total rows */
if residency = ' ' then residency = 'Z';  /* residency total columns */
run;

* output total rows;
* (this data set is used to calculate rates of change);

proc tabulate data=bio2 missing out=bio_total;
class college1 /* cat */ academic_period residency;
table college1 all, academic_period*(residency all);
run;

data bio_total; set bio_total;
if college1  = ' ' then college1  = 'UZ'; /* university total rows */
if residency = ' ' then residency = 'Z';  /* residency total columns */
run;
```

Below is part of the data set “bio_total” we get from proc tabulate above.

COLLEGE1	ACADEMIC_PERIOD	RESIDENCY	_TYPE_	_PAGE_	_TABLE_	N
AH	201010	N	111	1	1	262
AH	201010	R	111	1	1	834
AH	201010	Z	110	1	1	1096
AH	201110	N	111	1	1	276
AH	201110	R	111	1	1	824
AH	201110	Z	110	1	1	1100
AH	201210	N	111	1	1	252
AH	201210	R	111	1	1	808
AH	201210	Z	110	1	1	1060
AH	201310	N	111	1	1	235
AH	201310	R	111	1	1	809
AH	201310	Z	110	1	1	1044
DN	201010	N	111	1	1	177
DN	201010	R	111	1	1	312
DN	201010	Z	110	1	1	489
DN	201110	N	111	1	1	179
DN	201110	R	111	1	1	330
DN	201110	Z	110	1	1	509
DN	201210	N	111	1	1	183
DN	201210	R	111	1	1	333
DN	201210	Z	110	1	1	516
DN	201310	N	111	1	1	174
DN	201310	R	111	1	1	331
DN	201310	Z	110	1	1	505

STEP 2: USING THE LAG(N) FUNCTION TO SET UP THE DATA SO THAT YEAR-ON-YEAR CHANGES (IN NUMBERS AND PERCENTAGES) CAN BE CALCULATED

```
* calculate total row number changes and total row percentage changes;

proc sort data=bio_total; by college1 academic_period residency; run;

data bio_total_b; set bio_total;
by college1;
count + 1;
if first.college1 then count = 1;

n_3 = lag3(n);
change = n - n_3;
change_pct = (change/n_3)*100;

if count in (1,2,3) then do;
   n_3 = .;
   change = .;
   change_pct = .;
end;
run;
```

Each school/college is treated as one group.

Using lag(3) function to set up the data as we have 3 rows for each school/college in a year: In-state, Out-state, and Total. Then year-on-year rates of change are calculated.

The rates of change for first three rows of each school/college are deleted as they are the first year of the time series.

Below is part of the data set "bio_total_b" after the rates of change are calculated.

COLLEGE1	ACADEMIC_PERIOD	RESIDENCY	N	count	n_3	change	change_pct
AH	201010	N	262	1			
AH	201010	R	834	2			
AH	201010	Z	1096	3			
AH	201110	N	276	4	262	14	5.3435
AH	201110	R	824	5	834	-10	-1.1990
AH	201110	Z	1100	6	1096	4	0.3650
AH	201210	N	252	7	276	-24	-8.6957
AH	201210	R	808	8	824	-16	-1.9417
AH	201210	Z	1060	9	1100	-40	-3.6364
AH	201310	N	235	10	252	-17	-6.7460
AH	201310	R	809	11	808	1	0.1238
AH	201310	Z	1044	12	1060	-16	-1.5094
DN	201010	N	177	1			
DN	201010	R	312	2			
DN	201010	Z	489	3			
DN	201110	N	179	4	177	2	1.1299
DN	201110	R	330	5	312	18	5.7692
DN	201110	Z	509	6	489	20	4.0900
DN	201210	N	183	7	179	4	2.2346

DN	201210	R	333	8	330	3	0.9091
DN	201210	Z	516	9	509	7	1.3752
DN	201310	N	174	10	183	-9	-4.9180
DN	201310	R	331	11	333	-2	-0.6006
DN	201310	Z	505	12	516	-11	-2.1318

STEP 3: USING PROC REPORT TO PRODUCE A REPORT IN THE DESIRED FORMAT

```
* some extra steps before stacking data sets together;

data bio_total; set bio_total;
cat = 98; /* total */
run;

data change_num; set bio_total_b(drop=count n n_3 change_pct);
rename change = n;
cat = 99; /* total number changes */
run;

data change_pct; set bio_total_b(drop=count n n_3 change);
rename change_pct = n;
cat = 100; /* total percentage changes */
run;

* stack all data sets together;

data all; set bio3 bio_total change_num change_pct; run;

* use proc report to produce the final report;

title1'Student Enrollment by Major';
title2'(Census 2)';
proc report data=all missing;
format college1 $college. cat cat. academic_period $term.
residency $residency.;
column ('MCV-Campus Schools' college1 cat academic_period, residency, n);
define college1        / group        '' order=data preloadfmt noprint;
define cat             / group        '' order=data preloadfmt;
define academic_period / across       '' order=data preloadfmt;
define residency       / across       '' order=data preloadfmt;
define n               / analysis sum '' f=comma15.;

compute before college1/style={just=l background=lightblue font_weight=bold};
line college1 $college.;
endcomp;
```

```
compute cat;
if cat = 98 then  call define(_row_,'style','style={font_weight=bold}');
if cat = 99 then  call define(_col_,'style','style={font_style =italic}');
if cat = 100 then call define(_col_,'style','style={font_style =italic}');
endcomp;

compute n;
if cat = 100 then call define(_col_,'format','comma10.1');
endcomp;

compute after college1;
line' ';
endcomp;
run;
```

Define font style and weight for the three rows: total, number changes, and percentage changes

Define the number format for the percentage changes

CONCLUSION

For reporting purposes, procedure tabulate and procedure report are among the most commonly used. Compute blocks in procedure report even offer the capability of calculating additional columns. It is, however, difficult to feed into calculations data from other rows by using only those two procedures.

By utilizing out= option in procedure tabulate and the lag(n) function, rates of change with inputs from other rows of a report can be easily calculated. This same approach can also be used in a variety of scenarios to produce customized reports.

CONTACT INFORMATION

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

Khoi Dinh To, PhD
Office of Planning and Decision Support, Virginia Commonwealth University
todinhkhoi@yahoo.com, kdto@vcu.edu