

About the presenter

- Khoi D. To is a senior research analyst in the Enterprise Analytics & Advanced Research Group, Office of Planning & Decision Support, Virginia Commonwealth University. His responsibilities include data management, reporting, and conducting analytical research related to students, faculty, and institutional operations.
- He earned a doctoral degree in Higher Education (area of specialty: Quantitative Methods) from the University of Georgia and is a certified SAS programmer.

 SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

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

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

#SASGF 

Student Enrollment by Residency, School, and Degree

	Fall 2009			Fall 2010		
	In State	Out State	Total	In State	Out State	Total
Allied Health Professions						
Baccalaureate	165	12	177	176	10	186
Masters	390	154	544	391	167	558
Doctoral	249	90	339	234	95	329
Grad. Post-Bacc Certs	10	2	12	5	2	7
Grad. Post-Masters Certs	20	4	24	18	2	20
All Degrees and Certs	834	262	1,096	824	276	1,100
Headcount Growth	.	.	.	-10	14	4
% Growth over prior Fall	.	.	.	-1.2	5.3	0.4
..... (More schools/colleges here)						
ALL DEGREE PROGRAMS						
Baccalaureate	845	45	890	886	40	926
Masters	735	284	1,019	720	302	1,022
First Professional	1,111	547	1,658	1,130	548	1,678
Doctoral	411	325	736	394	338	732
Grad. Post-Bacc Certs	90	48	138	63	35	98
Grad. Post-Masters Certs	29	4	33	25	2	27
All Degrees and Certs	3,221	1,253	4,474	3,218	1,265	4,483
Headcount Growth	.	.	.	-3	12	9
% Growth over prior Fall	.	.	.	-0.1	1	0.2

The report can be achieved by the following steps

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

Step 1: Output aggregate data into data sets (proc tabulate)

- Output individual lines into a data set:

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

- Output total lines into another data set:

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

"Total lines" data set from proc tabulate

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

Step 2: Use lag(n) function to set up data and calculate year-on-year 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;
```

Each school/college is treated as one group.

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

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.

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

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

Data set after lag(3) calculation

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

Step 3: Stack the “total lines” data set to the “individual lines” data set, and use proc report for the final output

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

SAS® GLOBAL FORUM 2016
IMAGINE. CREATE. INNOVATE.

Thank You

