Paper CC-13

SAS® TIPS FOR INSTITUTIONAL RESEARCHERS TO TRACK STUDENT OUTCOMES EFFICIENTLY

Vijayalakshmi Sampath, Northern Virginia Community College (NOVA), Annandale, VA

ABSTRACT

Analysts and programmers in the Institutional Research (IR) field at Higher Education Institutes often work on research studies where a cohort of students must to be tracked across semesters (Longitudinal studies) in order to summarize their outcomes such as retention, course success rate, and graduation. Some studies also involve tracking outcomes and summarizing results for different types of student cohorts for a given semester (Cross Sectional studies). Since student information in College databases is constantly changing, researchers use static versions of the data for all official reporting. The static IR files in many colleges are generated after the Census date for every semester, resulting in multiple data files. As a result, tracking students across multiple files or tracking multiple cohorts becomes challenging and tedious for multi-year studies. This paper presents innovative SAS programming logic to track and output results of various student outcomes under the multiple semester and the multiple cohort scenarios. It utilizes Macro programming and Arrays within and outside of the DATA step in a robust manner to achieve this.

INTRODUCTION

Institutional Research (IR) offices at Higher Education Institutes are responsible for leading and conducting diverse research and analytical studies in the areas of student outcomes, program evaluation, institutional planning, and policy formulation. They also support the internal and external reporting requirements of the Institutes which include conducting statistical analysis, tracking student performance, grant writing, and conducting studies on special College initiatives. IR offices often use many data sources in order to efficiently complete the research and reporting processes. These include static SAS data files, Oracle production databases, State and Federal data resources, etc. Longitudinal studies often involve tracking a group of students (Cohort) across multiple time periods in order to assess their outcomes such as retention and graduation. Additionally, Cross Sectional studies examining the same outcomes for cohorts with different characteristics (demographics, enrollment attributes, etc) are also conducted. Developing SAS code to conduct such studies could become tedious and cumbersome since multiple outcomes need to be tracked over a period of time and this involves working with multiple SAS datasets corresponding to the semesters under consideration. This paper provides the programming solution under two scenarios utilizing features such as Macros and Arrays within SAS to efficiently track cohorts of students across semesters.

SINGLE COHORT MULTIPLE TIME PERIODS

Students often need to be tracked semester by semester to see if they are having a positive outcome (retention, graduation, transfer) or a negative one (drops, unsuccessful grades, etc). Another analysis of interest is to observe, over time, how these outcome rates are changing in order to implement strategic plans and target the students at the right time to improve their future outcomes. However, simultaneously tracking more than one outcome over time becomes daunting; especially when one has to 'look up' many different data sets to flag students. The case presented here is as follows:

- Static SAS data files are available for every semester which show student enrollment details. Additionally, for every academic year, SAS data files are available showing graduation details for students who graduated. These files need to be accessed dynamically in order to correctly flag students.

- Starting with a student cohort in the Fall semester, retention and graduation status for five consecutive semesters need to be tracked.
- Outcomes: Retention for every semester is tracked independent of the previous semesters. Hence, we are interested in seeing if the student is still enrolled in a particular semester rather than seeing if the student is continuously enrolling every semester. Also, if a student graduates in an academic year, the student is considered retained for the semesters spanning that year.

Figure 1. SAS Logic for Single Cohort Scenario

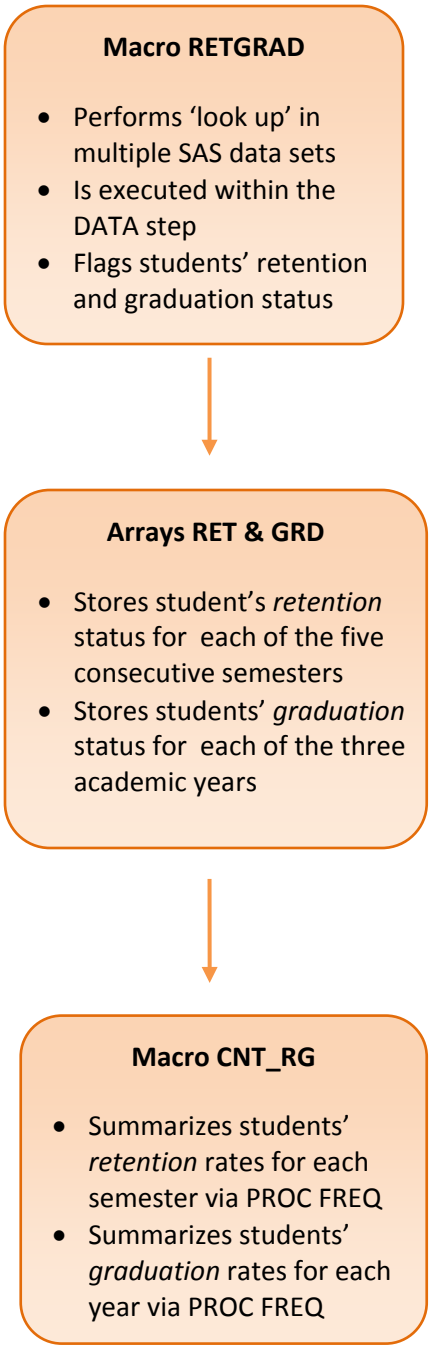


Figure 1 shows the basic flow of SAS programming logic to achieve this. The code employs a combination of macros and arrays to track a single cohort, starting college in Fall 2004, across five semesters till Spring 2007. In SAS® Code 1 the %LET statement defines the global parameter A which stores the starting semester of the cohort which, in this case, is Fall 2004.
The macro RETGRAD performs a file look up operation within the DATA step for each row of data read during the MERGE process. Macro RETGRAD is defined with two parameters to store the file name endings. It is to be noted that the SAS data file extensions are indicative of the semester/year for which student information is stored. Parameter B is dynamically assigned the file names pertaining to enrollment information in subsequent semesters and, likewise, C is assigned the file names pertaining to graduation information.

Two arrays namely RET and GRD are defined in the DATA step which store the retention and graduation indicators for every student over the course of the five semesters and three years, respectively. Storing the students' outcomes status in this manner is useful to examine the rates for individual semesters over a period of time.

In the MERGE statement the file extensions, indicating the semester and year, are used to define the IN= DATA step option for each of the files contributing to the merge. Then all the retention and graduation indicator fields (appearing in the ARRAY statement) are initialized to "NO" using DO loops. Next, the macro RETGRAD is executed multiple times for each row of student data that is read for the MERGE process. As the macro executes and finds the retention and graduation information in the files being merged, the indicators are updated in the ARRAY to "YES" for the respective semesters.

In the end macro CNT_RG employs PROC FREQ to obtain the retention rates for each semester and the graduation rates for each year. Since the program, in general, has to look up many SAS data files the SASFILE command is used in the beginning to load all the files to be used; this reduces the processing time as the data sets are held in the memory to be accessed recursively during macro execution within the DATA step.

SAS® CODE 1

```
OPTIONS SYMBOLGEN MPRINT;

SASFILE IRIS2000.S20052 LOAD; SASFILE IRIS2000.S20054 LOAD; SASFILE IRIS2000.S20062 LOAD;
SASFILE IRIS2000.S20064 LOAD; SASFILE IRIS2000.S20072 LOAD; SASFILE ALUMGRAD.GRAD2005 LOAD;
SASFILE ALUMGRAD.GRAD2006 LOAD; SASFILE ALUMGRAD.GRAD2007 LOAD;

%LET A=FA04;
   /** CREATING MACRO RETGRAD FOR FILE LOOK UP **/
%MACRO RETGRAD(B=, C=);
        IF &A.=1 AND &B.=1 THEN DO;
                RET_&B.="YES";
                        IF &C.=1 THEN GRD_%SUBSTR(&C.,2)= "YES";
        END;
        ELSE IF &A.=1 AND &B.=0 THEN DO;
                IF &C.=1 THEN DO;
                        RET_&B.= "YES";
                        GRD_%SUBSTR(&C.,2)= "YES";
                END;
        END;
%MEND RETGRAD;

DATA RETGRAD_0407 (DROP=I J K);
ARRAY RET[*] $ RET_SP05 RET_FA05 RET_SP06 RET_FA06 RET_SP07;
ARRAY GRD[*] $ GRD_0405 GRD_0506 GRD_0607;
LENGTH GRD_STATUS $3. ;
MERGE IRIS2000.S20044(IN=FA04 KEEP=SSN TYPE) IRIS2000.S20052(IN=SP05 KEEP=SSN)
      IRIS2000.S20054(IN=FA05 KEEP=SSN)
      IRIS2000.S20062(IN=SP06 KEEP=SSN) IRIS2000.S20064(IN=FA06 KEEP=SSN)
      IRIS2000.S20072(IN=SP07 KEEP=SSN) ALUMGRAD.GRAD2005(IN=G0405 KEEP=SSN)
      ALUMGRAD.GRAD2006(IN=G0506 KEEP=SSN) ALUMGRAD.GRAD2007(IN=G0607 KEEP=SSN);
BY SSN;
IF FA04=1 AND TYPE IN ('1','3');
GRD_STATUS= "NO"; /* INITIALIZING OVERALL GRADUATION STATUS */
  DO I=1 TO DIM(RET); /* INITIALIZING SEMESTER RETENTION AND GRADUATION STATUS */
        RET[I] ="NO";
  END;
  DO J=1 TO DIM(GRD);
        GRD[J] = "NO";
  END;
     /** EXECUTING MACRO RETGRAD WITHIN DATA STEP **/
     %RETGRAD(B=SP05, C=G0405)
     %RETGRAD(B=FA05, C=G0506)
     %RETGRAD(B=SP06, C=G0506)
     %RETGRAD(B=FA06, C=G0607)
     %RETGRAD(B=SP07, C=G0607)
  DO K=1 TO DIM(GRD);
     IF GRD[K] = "YES" THEN DO;
        GRD_STATUS = "YES";  /** OVERALL 3 YR GRADUATION STATUS **/
     END;
  END;
RUN;

/** MACRO TO OBTAIN RETENTION AND GRADUATION NUMBERS ***/
%MACRO CNT_RG(V=, T=);
PROC FREQ DATA=RETGRAD_0407;
TITLE "&V.";
TABLES &T./CROSSLIST NOCOL ;
RUN;
%MEND CNT_RG;

   /** EXECUTING MACRO CNT_RG **/
ODS RTF FILE = "DESKTOP\RETENTION AND GRADUATION FOR FALL 2004 COHORT.RTF" BODYTITLE;
%CNT_RG(V= COHORT SEMESTERWISE RETENTION: FALL 2004-SPRING 2007 (INCLUDES GRADUATES),
        T= RET_SP05 RET_FA05 RET_SP06 RET_FA06 RET_SP07);
%CNT_RG(V= COHORT YEARWISE GRADUATION: FALL 2004-SPRING 2007,
        T= GRD_0405 GRD_0506 GRD_0607);
ODS RTF CLOSE;
```

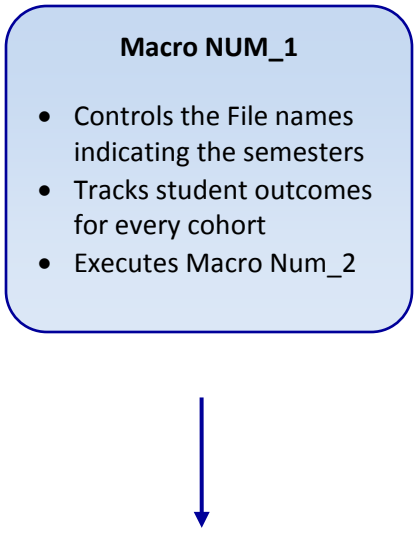
MULTIPLE COHORTS SINGLE TIME PERIOD

This section discusses the case where student outcomes need to be analyzed for cohorts with different characteristics but the analysis is restricted to a single semester. The resourcefulness of using macros within macros is demonstrated using specific cohort characteristics, but the logic may be extended to more general cases. The multiple macros facilitate easy progression from one cohort to the next while simultaneously calculating the outcome related variables. The additional advantage to this method is being able to efficiently output the results to the appropriate destination without having to rewrite code to enumerate the outcome statistics for each individual cohort.

The analysis may include summarizing the details for cohorts originating from different years and different semesters. As mentioned earlier, this requires scanning through individual SAS files for each of these cohorts. The case analyzed here is as follows:

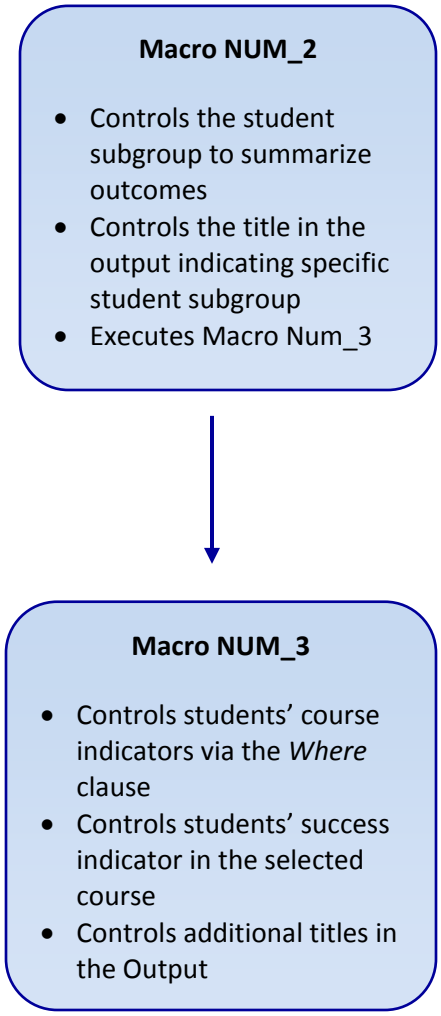
- There are three student cohorts originating from three consecutive Fall semesters. For each semester, the student attributes are tracked in one SAS data file and the student course enrollment information is tracked in another SAS data file for the corresponding semester. The semester designations are a part of the SAS data file name extensions.
- Outcomes: Student retention from the Fall semester to the Spring semester and graduation in the corresponding academic year need to be tracked for all three cohorts. Additionally, student enrollment and performance in a specific course needs to be analyzed as well.
- The analysis needs to be broken down by specific student characteristics. These include whether or not they are fresh from high school (first time students) or are returning students.
- Individual outputs corresponding to the outcomes mentioned above for each cohort subgroup need to be generated. Additionally, appropriate titles and output destinations need to be included.

Figure 2. SAS Logic for Multiple Cohorts Scenario



The analysis is carried out using three dedicated macro programs where the macros are executed within each other. In SAS® Code 2, there are three macros namely Num_1, Num_2, and Num_3. During the Execution of Num_1, Num_2 and Num_3 are also executed; Figure 2 summarizes the steps that these three macros control.

In Macro NUM_1, the parameter Y1 is used to reference the SAS data files for three Fall semesters and Y2 is used to reference the subsequent Spring semesters, respectively. During each execution of NUM _1 indicators for students in the cohort under consideration are created (via PROC SQL) to track students who were retained from the Fall to the Spring semesters. Their enrollment and grades in a specific course (prefixed SDV) in the Fall semester is tracked, and, their graduation status spanning the Fall and Spring semesters is tracked. Since graduation SAS data files are created only once a year, the file extensions are different but are referenced efficiently by using the %SUBSTR macro function.



Next, macro NUM _2 has two parameters namely T and T1 to identify exclusive sections of the cohort to run the Macro NUM_3. Some students are attending college for the very first time in the Fall semester while the others are returning from the previous semesters; the field *Type* within the SAS Data set identifies them.

The summary for the outcomes retention and graduation is obtained within Macro NUM_3 via the PROC FREQ procedure for different cohort sub groups. The parameter W controls how the *Where* clauses are phrased within PROC FREQ while S1 and S2 facilitate the outcomes summarizations for student cohorts who are enrolled /not enrolled in an SDV course and for those who did/did not complete the course successfully.

Additionally, the %If and %Then statements within macro NUM _2 controls (via Macro NUM _3) whether or not to place a *Where* clause at all in the PROC FREQ procedure, depending on whether outcomes need to be obtained for all students (*Type* specification not required) or only first time students (*Type* =1).

SAS® CODE 2

```
OPTIONS SYMBOLGEN MPRINT;

%MACRO NUM_1(Y1=, Y2=);

DATA S&Y1.(KEEP=SSN TYPE) ;
 SET IRIS2000.S&Y1.;
RUN;

PROC SQL;
CREATE TABLE S&Y1._SDV AS
SELECT A.*,
 (CASE WHEN EXISTS (SELECT "X"
                    FROM IRIS2000.S&Y2. B
                    WHERE A.SSN=B.SSN)
       THEN "YES" ELSE "NO" END) AS RETAINED,
 (CASE WHEN EXISTS (SELECT "X"
                    FROM IRIS2000.C&Y1. C
                    WHERE A.SSN=C.SSN AND
                          C.DISC IN ("STD","SDV"))
       THEN "YES" ELSE "NO" END) AS SDV_IND,
 (CASE WHEN EXISTS (SELECT "X"
                    FROM IRIS2000.C&Y1. D
                    WHERE A.SSN=D.SSN AND
                          D.DISC IN ("STD", "SDV") AND
                          D.GRADE IN ("A","B","C","P","S"))
       THEN "YES" ELSE "NO" END) AS SDVSUCC_IND,
 (CASE WHEN EXISTS (SELECT "X"
                    FROM ALUMGRAD.GRAD%SUBSTR(&Y2.,1,4) E
                    WHERE A.SSN=E.SSN)
       THEN "YES" ELSE "NO" END) AS GRADUATED

FROM S&Y1. A ;
QUIT;

 %MACRO NUM_2(T=, TL=);

   %MACRO NUM_3(W=, S1=, S2=, N=);
     TITLE "FALL TO SPRING RETENTION FOR FALL %SUBSTR(&Y1.,1,4): &TL. STUDENTS # &N.";
     TITLE2 "SUBGROUPS: STUDENT TYPE= &TL., SDV ENROLLED= &S1., SDV SUCCESS= &S2.";
      PROC FREQ DATA=S&Y1._SDV;
      TABLE RETAINED*GRADUATED/NOCOL;
      &W.
      RUN;
   %MEND NUM_3;

   %IF &T. =  %THEN %DO;
     %NUM_3(W=, S1=, S2=, N=1)
     %NUM_3(W=WHERE SDV_IND= "YES"; , S1=YES, S2=, N=2)
     %NUM_3(W=WHERE SDV_IND= "YES" AND SDVSUCC_IND= "YES"; , S1=YES, S2=YES, N=3)
   %END;
   %ELSE %DO;
     %NUM_3(W=WHERE TYPE="1";, S1=, S2=, N=1)
     %NUM_3(W=WHERE TYPE="1" AND SDV_IND= "YES";, S1=Y, S2=, N=2)
     %NUM_3(W=WHERE TYPE= "1" AND SDV_IND= "YES" AND SDVSUCC_IND= "YES";, S1=YES, S2=YES, N=3)
   %END;
 %MEND NUM_2;

 %NUM_2(T=, TL=ALL)
 %NUM_2(T="1", TL=FIRST TIME)
%MEND NUM_1;

ODS RTF FILE = "DESKTOP\RETENTION FOR FALL 2004 TO FALL 2006 COHORTS.RTF" BODYTITLE;
%NUM_1(Y1=20044, Y2=20052)
%NUM_1(Y1=20054, Y2=20062)
%NUM_1(Y1=20064, Y2=20072)
ODS RTF CLOSE;
```

CONCLUSIONS

Analyzing students' enrollment and performance over time is a constant area of research at IR offices in Higher Education Institutes. With the judicious use of many SAS features such as Macros, Arrays, and ODS in SAS programs IR analysts may conduct data analyses more efficiently by reducing the programming time, maintaining the code logic in a concise manner, and controlling output display. The tips presented here may be extended to other situations with simple modifications.

ACKNOWLEDGEMENTS

I would like to thank Dr. George Gabriel for encouraging and supporting my paper contribution to the SESUG 2010 Conference.

CONTACT INFORMATION

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

Vijayalakshmi Sampath
Office of Institutional Research, Planning, and Assessment
Northern Virginia Community College
4001 Wakefield Chapel Rd.
Annandale, VA 22003
E-mail: vsampath@nvcc.edu or vibha_atm75@yahoo.com
Ph: (703) 323-3129