Selectively Processing SAS Datasets - Technical Tip

Sunil Kumar Gupta
Gupta Programming
SunilKGupta@GuptaProgramming.com
http://www.GuptaProgramming.com

How do you automate the process of generating the contents and a sample printing of all the sas datasets in a given directory without hard coding the names of each dataset?

The following technical tip accomplishes this by reading a table of contents file and storing the sas dataset names into a field. Then, macro variables are created for each dataset name and are utilized by another macro to print the contents and a sample listing.

First, go to the directory containing all your sas datasets. Execute the directory command to save the names of all your datasets in the directory to a file to be read in by the sas program:

```
C:\SASDATA>DIR *.SSD > TOC.LST
```

Once you have this file, create your sas program. Define a libname statement where the sas datasets are located.

```
LIBNAME SDATA 'C:\SASDATA';
```

Create a dataset containing all of the dataset names in the dname field;

```
DATA OUTLST (KEEP=DNAME);

LENGTH NAME $14. DNAME $8.;
```

Access the table of contents file;

```
INFILE 'C:\SASDATA\TOC.LST' END=FINAL;
INPUT NAME;
```

Save only the dataset name to the dname field with the scan function;

```
DNAME=SCAN(NAME, 1, '.');
```

Print the values of name and dname to verify the scan function;

```
PUT NAME = DNAME=;
```

Create the macro variable total containing the total number of sas datasets;

```
IF FINAL THEN CALL SYMPUT('TOTAL', N);

RUN;
```

Verify the contents of the dataset generated;

```
PROC CONTENTS DATA=OUTLST; RUN;
PROC PRINT DATA=OUTLST; RUN;
DATA NULL;

SET OUTLST;
```

Create a macro variable for each dataset name. The root name is 'AN' with the suffix number incrementing with each dataset. The value of the macro variable is the name of the sas dataset contained in the dname field;

```
CALL SYMPUT('AN'||LEFT(_N_), DNAME)

RUN;
```

Verify the contents of the macro variables.

```
%PUT AN1 = &AN1;
```

```
%PUT AN2 = &AN2;
%PUT TOTAL = &TOTAL;

Define a macro to generate the contents and a sample of the dataset;
%MACRO INFO(PDATA) ;
PROC CONTENTS DATA=&PDATA;
RUN;
PROC PRINT DATA=&PDATA (obs%) label;
RUN;
%MEND INFO;
```

Define a macro to select each of the macro variables just created containing the dataset name and execute the info macro for each dataset selected.

```
%MACRO PROCESS;

%LET I=1;
        Process each macro variable one at a time;
%DO %WHILE (&I <= &TOTAL);
    Apply the trim function to remove any trailing blanks in the field;

    %LET ITRIM=%TRIM(&AN&&I);

    Execute the info macro for each dataset in the directory;

    %INFO(SDATA.&ITRIM);

    Increment I by 1;
    %LET I = %EVAL(&I + 1);
%END;
%MEND PROCESS;

Execute the process macro;
%PROCESS;
```