Automatically Obtaining and Using Data Creation Dates
Varsha C. Shah, Dept. of Biostatistics, UNC-CH, Chapel Hill, NC
Kathy Roggenkamp, Dept. of Biostatistics, UNC-CH, Chapel Hill, NC

ABSTRACT
This paper describes two ways to take automatic advantage of data creation dates, as well as two easy ways to obtain them (PROC SQL and PROC CONTENTS). The first example illustrates making a date part of the title of a data management report. The second shows incorporation of the date into data set names.

INTRODUCTION
This presentation illustrates two instances when the value of a data creation date is useful, followed by a description of two procedures – PROC SQL and PROC CONTENTS – for automatically obtaining the data creation date. The former offers a more compact method, but those not comfortable with SQL have the latter procedure available as an alternative. Illustrations and code are included in the appendix.

EXAMPLE 1: Date in the Title
Suppose that data for demographics is generated for the first time on April 15, 2002, and the program for validation checks is submitted. The same process is repeated after 23 days on May 8th, and another time after 16 days on May 24th. The validation results for these three retrievals could look like Illustrations 1 through 3 in the appendix, with the data creation date included in the data management report title.

As the data checking reports are reviewed over time, the date in the title will help the data manager track progress in cleaning up the data from one retrieval to the next.

The results in Illustration 3 seem suspicious in the sense that subject 157 has a missing race value, while her race was not missing in the previous two retrievals. For a greater quantity of output, tables could be saved as output data sets and PROC COMPARE used for comparisons. This brings us to the next issue.

EXAMPLE 2: Date as Part of the Data Set Name
Suppose the process is repeated 10-15 times. Let us say it is desirable to compare the error list of March with the error list of October. If the resulting files were named with suffixes relating to the retrieval date as in Illustration 4, this task would be easier.

There are not hard and fast rules about designing the name. Here are some personal preferences.

- If the data retrieval takes place every month for exactly one calendar year starting in January, then the suffixes will range from 1 to 12.
- If there are multiple retrievals within a month, it is desirable to have values of the month and date both as part of the name.
- If the data retrievals continue over years, then it is also desirable to have the value of the year added to the suffix.

DERIVING THE RETRIEVAL DATE
The steps for the two methods are described side-by-side in Illustration 5 of the appendix for a library "v1" with a member data set "demo".

Macro variable &prtdate yields the date value in the form DDMMMYY as used in the titles for Illustrations 1-3. The following steps show how to separate each of the date parts from this value. Here a macro variable &dsnum is created in YYMMDD form to be used as a suffix for an output data set.

```plaintext
data _null_;
    length tmpdate $7 y m d $2 dsnum $6;
    tmpdate="&prtdate"d;
    date=input(tmpdate,7.);
    y = substr(put(year(date),4.),3,2);
    m = left(trim(put(month(date),2.)));
    if length(m) = 1 then m = '0' || m;
    d = left(trim(put(day(date),2.)));
    if length(d) = 1 then d = '0' || d;
    dsnum = y || m || d;
    call symput('dsnum',dsnum);
run;
*** check/print the results ***;
%put "the retrieval date YYMMDD resolved to" &dsnum;
```

Finally, the following statements will create data set names as shown in Illustration 4.

```plaintext
data error_demo&dsnum;
    set MyWorkData;
    other data step statements  etc.
run;
```
REFERENCES

CONTACT INFORMATION
Varsha C. Shah
CSCC, Department of Biostatistics, CB #8030
University of North Carolina
Chapel Hill, NC 27514
Phone: 919-966-5160
Fax: 919-966-6335
Email: uccvcs@mail.cscc.unc.edu

Kathy J. Roggenkamp
CSCC, Department of Biostatistics, CB #8030
University of North Carolina
Chapel Hill, NC 27514
Phone: 919-966-5304
Fax: 919-966-3265
Email: ucckjr@mail.cscc.unc.edu

APPENDIX

Illustration 1:
Table 1: Missing Demographic Data
Based on Data Retrieved on 15APR02

<table>
<thead>
<tr>
<th>Subj</th>
<th>Age</th>
<th>Race</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>67</td>
<td>B</td>
<td>.</td>
</tr>
<tr>
<td>106</td>
<td>53</td>
<td>F</td>
<td>.</td>
</tr>
<tr>
<td>123</td>
<td>.</td>
<td>M</td>
<td>.</td>
</tr>
<tr>
<td>157</td>
<td>.</td>
<td>W</td>
<td>F</td>
</tr>
</tbody>
</table>

Illustration 2:
Table 1: Missing Demographic Data
Based on Data Retrieved on 08MAY02

<table>
<thead>
<tr>
<th>Subj</th>
<th>Age</th>
<th>Race</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>67</td>
<td>B</td>
<td>.</td>
</tr>
<tr>
<td>123</td>
<td>.</td>
<td>.</td>
<td>M</td>
</tr>
<tr>
<td>157</td>
<td>.</td>
<td>W</td>
<td>F</td>
</tr>
</tbody>
</table>

Illustration 3:
Table 1: Missing Demographic Data
Based on Data Retrieved on 24MAY02

<table>
<thead>
<tr>
<th>Subj</th>
<th>Age</th>
<th>Race</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>67</td>
<td>B</td>
<td>.</td>
</tr>
<tr>
<td>123</td>
<td>.</td>
<td>.</td>
<td>M</td>
</tr>
<tr>
<td>157</td>
<td>.</td>
<td>W</td>
<td>F</td>
</tr>
</tbody>
</table>

Illustration 4:

<table>
<thead>
<tr>
<th>Name</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>error_demo020415</td>
<td>4/15/02 2:57 PM</td>
</tr>
<tr>
<td>error_demo020508</td>
<td>5/08/02 10:02 AM</td>
</tr>
<tr>
<td>error_demo020524</td>
<td>5/24/02 9:47 PM</td>
</tr>
</tbody>
</table>

Illustration 5:

```sql
proc sql noprint;
  select crdate
  into :crdate
  from dictionary.tables
  where upcase(libname)='V1' and
    upcase(memname)='DEMO';
quit;
run;

%let crdate=%substr(&crdate,1,7);
%put &crdate "result of SQL";
%let prdate=%substr(&crdate,1,7);
%put &prdate;
run;
```

```sql
proc contents data=V1.DEMO noprint out=conts;
run;
data conts;
  set conts;
  if _n_=1;
  keep crdate;
run;
data _null_; set conts;
  call symput('crdate',put(crdate,datetime16.));
run;
%put &crdate "results of contents";
%let prdate=%substr(&crdate,1,7);
%put &prdate;
run;
```