DCSUG Newsletter

First Quarter Meeting

When: Tuesday, March 7, 2000
Time: 8:30a.m. to noon
Place: Bureau of Labor Statistics
            Postal Square Building
            Room G440
            First Street North East
            Washington DC

Agenda: 8:30 - 9:00   Registration and refreshments
            9:00 - 9:15   Announcements
            9:15 - 10:15   Martin O'Connell, Geico
                               Some Issues to Consider When Implementing SAS Programs in a Production Environment
            10:15 - 10:30   Break
            10:30 - 11:30   Richard Stoll
                               A Different Proc Transpose Application
            11:30 - Noon   User-to User Questions and Answers
                               Moderators - DCSUG officers
                               *__Presentation times subject to change__

Visitors must use the visitor's entrance on First Street NE. Directions for those taking Metro: Take the Red Line to Union Station. Exit the station via the Amtrak exit, but do not go up to the train level. Instead, after going up to and out through the farecard gates, turn to the left where there is an exit to the street (First Street NE). Directly across from the exit are the employee and visitor's entrances to the building. After checking in with security, take the main elevators down to the G level. **BRING PICTURE ID!**

> If you have special needs and are planning to attend this meeting, please contact any DCSUG officer no later than 2/24/2000

SAS® is a registered trademark of the SAS Institute, Cary, NC

F F F F F F F F   **If you are not currently a paid member of the DCSUG and are planning to attend the meeting, please Email Larry Altmayer at laltmaye@census.gov or call 301-457-2581 by Mar 1st and leave your name. Security at the Bureau of Labor Statistics is very tight and having your name on a list of possible attendees will make your entrance to the building easier.**
F F F F F F F F

President's Message
Peter Knapp

I would like to start by thanking everyone who contributed to making sure the Y2K "bug" did not affect most of us. It amazes me that people are grumbling that so much money was spent and nothing went wrong. That, of course, was the point of all the hard work, extra hours, and sacrificed vacations that went into making sure the world as we know it did not end on January 1, 2000. Good job, everyone!

Speaking of hard work, a busy group of volunteers is putting together the 25th SAS Users Group International (SUGI) Conference. It will be held from April 9-12, 2000 in Indianapolis, IN and I am sure it will be a great conference. I hope to see you there.

A little closer to home is the next DCSUG meeting, which will be held on March 7, 2000 at the Bureau of Labor Statistics. Martin O'Connell and Richard Stoll will be giving two exciting presentations. If you can spare a morning from your post Y2K activities, I hope you can make the meeting. I'm certainly looking forward to attending the meeting an hopefully learning something new!

Some Issues to Consider When Implementing SAS Programs in a Production Environment

Martin O'Connell, Geico

So, you've got your favorite SAS program running reliably daily, weekly, monthly, or quarterly and just know the folks in your IT shop will be more than happy to manage it now, eh? Well, chances are, to them and their counterparts in your Production Control and Systems departments, this is not "business as usual" and will require special attention.

This paper discusses some of the major implementation issues you'll want to consider so that your MVS or UNIX SAS program transitions smoothly from your desktop to your batch production cycle. It illustrates useable program structures that enable basic production controls to be built around your program.

Martin J. O'Connell earned an MS in Mathematics from North Carolina State University in 1991. His area of study was Numerical Analysis and Optimization. He has 7 years of SAS experience on mainframes and 4 years in a client-server environment. He is currently working with SAS development on the GEICO Actuarial Data Warehouse. His areas of expertise include Base SAS and client server computing with the SAS system..

A Different Proc Transpose Application

Richard Stoll

Most of the time, merging two files is old hat. Every once in a while though, it turns out that it's not a standard merge. When each of the files contain several hundred variables and the files were not constructed according to a coordinated plan, several of the 'non-linking' variables can have identical names, and that can cause a problem. Are the variables identical for all common values? If not, from which file should a given variable be selected?

Proc Transpose can be used to identify those variables. This paper presents the methodology used in that Proc Transpose, along with the preliminary and follow-up steps that are an integral part of the overall analysis procedure.

Richard Stoll received a BA in Physics from New York University in 1953. He started working as a Physicist for the Navy and then drifted first into Operations Research and then into computer programming in SAS. Professionally, he has been in aircraft weapons systems, arms control and disarmament, mine health and safety, solar energy, energy statistics, and finally health statistics. He retired from the Department of Energy in 1990 and from Social and Scientific Systems, Inc. in March of 1999. Currently, he is a volunteer Programmer/Analyst (using SAS) at the Research Institute of the Hebrew Home of Greater Washington, in Rockville, Maryland.

Questions and Answers

Moderated by DCSUG Officers

Bring your SAS programming or procedure questions to DCSUG and its members. During this open session, you can ask questions of and get answers from experienced SAS users. All are welcomed to participate

Desktop Special Interest Group

The next meeting TBA

Check the DCSUG home page for the next meeting announcement:
http://www.ita.doc.gov/industry/otea/dcsug

Please contact Yesvy Gustasp or Joe Mirabal (see Steering Committee column) with questions, requests, suggestions, and offers.

To get to Westat, take I-270 8 miles from the Capital Beltway to exit #6B - Route 28 (toward Darnestown). Once on 28, go past the traffic light at the Shell station and turn right at the next traffic light, which is research Blvd. The 1650 building is less than 1/4 mile on the left, just over the crest of the. You can park in any non-reserved space. Signs will be posted to direct you to the meeting room.

If using Metro, take the Red line towards Shady Grove and get off at the Rockville station. Take Ride-On bus #54 in the direction of Lakeforest Mall. The bus goes out on 28 past I-270 and the Shell station and then turns onto Research Blvd. It takes 8-10 minutes from the Metro station to Westat. The #54 bus runs every half hour.

Note: Don't be confused by a large Westat sign that is in front of another Westat building on Research Blvd (1500 & 1550).

DCSUG 2000 Steering Committee
OFFICERS

CHAIR -- Peter Knapp
(202) 482-1359   (202)482-1388(fax)
peter_knapp@ita.doc.gov

SECRETARY -- Larry Altmayer
(301) 457-2581   (301) 457-4437 (fax)
laltmaye@census.gov

TREASURER -- Arlene Siller
(301) 458-4498   (301) 458-4032 (fax)
abs2@cdc.gov

SAS LIAISON -- Mike Rhoads
(301) 251-4308   (301) 294-2040 (fax)
rhoadsm1@westat.com

OTHER COMMITTEE MEMBERS

Karen Dennis
(301) 294-3876   (301) 294-2034 (fax)
dennisk1@westat.com

Frank Fry
(202) 452-2666   (202) 452-6433 (fax)
ffry@frb.gov

Michael Raithel
(301) 610-4997   (301) 315-5934 (fax)
raithem@westat.com

Linda Atkinson
(202) 694-5046   (202) 694-5718 (fax)
atkinson@econ.ag.gov

Howard Schreier
(202) 482-4180   (202) 482-4614 (fax)
Howard_Schreier@ita.doc.gov

DESKTOP SPECIAL INTEREST GROUP

CO-CHAIR -- Joe Mirabal
(703) 265-4620

jmmirabal@aol.com

VICE-CHAIR -- Yesvy Gustasp
(301) 589-4530   (202) 307-5888 (fax)
Yesvy.Gustasp@mix.cpcug.org

Membership Application/Renewal

Date:

Name:

Company Name:

Mailing Address:

City,   State,

                Zip

Phone #   Fax #

E-Mail:

Preferred mode of contact (check one):
                9 Fax   9 e-mail

SAS Version(s)

SAS Product(s):

Operating System(s)

Check one:
        9 Individual membership   9 Corporate sponsorship
            for 1 year is $10.00   for 1 year is $50.00

Please indicate a contact person for Corporate Sponsorships!

Check one:
9 New membership   9 Renewal membership

Mail to:

Washington DC SAS Users Group
P.O. Box 44670
Washington, DC 20026-4670

SAS Talk

Ian Whitlock, Westat

Recently on SAS-L, Machiko Hollifield asked how to create 100 samples with one record chosen randomly from each group. Say the data look like

GROUP	VALUE
1	11
1	12
1	13
2	21
2	22
3	31
3	32
3	33
3	34
...	...

Thus one sample might be:

GROUP	VALUE
1	13
2	21
3	32
...	...

Typical answers said to use a macro. For example, one person wrote:

```
%macro sample(n);

  %do i=1 %to &n;

    data sample;
      set frame;
      choose=ranuni(-1);
    run;

    proc sort data=sample;
      by group choose;
    run;

    data s&i ( drop = choose);
      set sample;
      by group;
      if first.group;
    run;

  %end;

%mend sample;

%sample(100)
```

So what's wrong with the answer? It does what was asked for. It produced 100 data sets and it took 300 steps to do it. A simple macro can generate an awful lot of steps.

In response to a plea on SAS-L for the ability to specify the output data set at run time, Howard Schreier wrote that he rarely needed such an ability because the only use would be to split a data set up into many little data sets; and it was far better to keep the data together and use by group processing to distinguish sets. The same principle applies here.

Suppose we ask instead how to produce one file with 100 samples, where each sample has one randomly chosen record from each group. If we knew the range of observation numbers for a group then we could use the POINT option to pick out the chosen record. Thus we need a step in front to find out the size of each group. Here is the complete program in three steps.

```
proc summary data = test nway ;
   class group ;
   output out = freq ( keep = group _freq_ ) ;
run ;

data samples ( keep = sample group value ) ;
   retain start end ;
   set freq ;
   end + _freq_ ;
   start = end - _freq_ + 1 ;

   do sample = 1 to 100 ;
      ptr = start + int ((end-start+1) * ranuni (684946631)) ;
      set test point = ptr ;
      output samples ;
   end;
run ;

proc sort data = samples ;
   by sample group ;
run ;
```

No macro was needed. We can repeat the run. We used three steps instead of three hundred, and we don't have to worry about handling 100 different data sets. Instead of more macro code to manage the analysis of each sample, we simply need by-group processing in one program. To see the advantage imagine being asked to identify the sample with the largest mean VALUE. In our case the code is simply:

```
proc summary data = samples nway ;
   var value ;
   class sample ;
   output out = means mean=meanval ;
run ;

proc summary data = means ;
   id sample ;
   var meanval ;
   output out = maxavg max=maxmean ;
run ;
```

I leave it to you to plan how to obtain the same information from the 100 samples asked for.

When SAS Misbehaves Using The IN= Data Set Option
By Peter Knapp, Department of Commerce

The intent of this program is to add exchange rates to a dataset of U.S. sales. The two datasets look like this:

U.S. SALES

Obs	DATE	CONNUMU	PRICE
1	07JAN1999	A	$24
2	08JAN1999	B	$52
3	09JAN1999	C	$33

EXCHANGE RATES

Obs	DATE	RATE
1	07JAN1999	0.563
2	08JAN1999	0.248
3	09JAN1999	0.123

The LOG of the program (using Version 6.12) looks like this:

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software Release 6.12 TS055

```
1 DATA USSALES;
2   MERGE USSALES (IN = US)
3       EXRATES (IN = RATE);
4   BY DATE;
5   IF US AND RATE;
6 RUN;
```

NOTE: The data set WORK.USSALES has 3 observations and 3 variables.
NOTE: The DATA statement used 0.08 seconds.

The OUTPUT looks like this:

U.S. SALES WITH EXCHANGE RATES ADDED

Obs	DATE	CONNUMU	PRICE
1	07JAN1999	A	$24
2	08JAN1999	B	$52
3	09JAN1999	C	$33

The Problem: The exchange rate variable RATE is not in the output data set. No NOTES, WARNINGS, or ERROR messages warn the user that something has gone wrong.

The Reason: RATE is the variable name created by the IN= option of the EXRATES dataset in line 3 of the LOG. It is a temporary variable that exists only for the duration of the data step. RATE is also the name of a permanent variable in the EXRATES dataset. Because two variables with the same name cannot exist in the same dataset, SAS chooses to keep the temporary IN= variable RATE over the permanent variable RATE that is read in from the EXRATES dataset. Being temporary, RATE is not written to the new combined USSALES dataset.

The Resolution: Make sure to use unique names when using the IN= dataset option. It's interesting to note that both Version 7 and Version 8 Developer's Release produce a NOTE in the LOG that informs the user that the program produces unexpected results. The logs generated by Version 7 and Version 8 Developer's Release follow:

NOTE: Copyright (c) 1998 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software Version 7 (TS P1)

```
1 DATA USSALES;
2   MERGE USSALES (IN = US)
3       EXRATES (IN = RATE);
4   BY DATE;
5   IF US AND RATE;
6 RUN;
```

NOTE: The variable RATE exists on an input data set, but was also specified in an I/O statement option. The variable will not be included on any output data set.
NOTE: The data set WORK.USSALES has 3 observations and 3 variables.
NOTE: DATA statement used:
     real time   0.10 seconds
     cpu time   0.04 seconds

NOTE: Copyright (c) 1999 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software Version 8 Developer's Release (TS D1)

```
1 DATA USSALES;
2   MERGE USSALES (IN = US)
3   EXRATES (IN = RATE);
4   BY DATE;
5   IF US AND RATE;
6 RUN;
```

NOTE: The variable RATE exists on an input data set, but was also specified in an I/O statement option. The variable will not be included on any output data set.
NOTE: There were 3 observations read from the dataset WORK.USSALES.
NOTE: There were 3 observations read from the   dataset WORK.EXRATES.
NOTE: The data set WORK.USSALES has 3 observations and 3 variables.
NOTE: DATA statement used:
     real time   0.02 seconds
     cpu time   0.02 seconds
_____

Conclusion: It can be very frustrating to try to debug a program that syntactically has no problems. I'm glad to see that SAS Institute Inc. has added a note to help users understand why use of the IN= dataset option sometimes produces unexpected results.

Calendar of Meetings and Events

This feature is meant to give users an idea of some of the activities of area users groups and special interest groups, as well as regional and international groups. If your group would like to be regularly included in this feature, please call Mike Rhoads or one of the DCSUG officers. There is a DCSUG directory with phone, fax, and E-mail contact information elsewhere in this newsletter.

March

7   **DCSUG**   DC SAS Users Group,
              General Meeting
        Time: 8:30 am - 12:00 noon
        Site: Bureau of Labor Statistics
                  2 Massachusetts Ave. N.E.
        Program: S e e   f r o n t   p a g e
                  announcement
        Contacts:
          Mike Rhoads at (301) 251-4308
            RhoadsM1@Westat.com
          Peter Knapp at (202) 482-1359
            peter_knapp@ita.doc.gov

April

9-12   **SUGI** SAS Users Group International
              Conference
        Site: Indiana Convention Center
                  Indianapolis IN
        Contacts:
          Nancy Patton at (703) 892-4472
          Nancy.Patton@GECapital.com

June

13   **DCSUG**   DC SAS Users Group,
              General Meeting
        Time: 8:30 am - 12:00 noon
        Site: Bureau of Labor Statistics
                  2 Massachusetts Ave. N.E.
        Program: TBA

Other Groups

CENSUG   Census SAS Users Group
    Contact: Debbie Gattuso at (301)457-1687

MDSUG   Maryland SAS Users Group
    Contact: Mr. R.H. Miller, (410) 740-4229
      refa09a@prodigy.com

NCHSSUG   National Center for Health Statistics
                  SAS Users Group
    Contacts: Arlene Siller at (301) 458-4498
                        abs2@cdc.gov
                  Linda Tompkins at (301) 458-4533
                        lit3@cdc.gov

NIHSUG   National Institutes of Health SAS
                  Users Group
    Contact: Ray Danner at (301) 496-6037

NOVASUG   Northern Virginia SAS Users Group
    Contact: Joe Mirabal at (703) 265-4620
     j   mmirabal@aol.com

VASUG   Virginia SAS Users Group
                  (Richmond area)
    Contact: Beth Wilson at (804) 967-1000
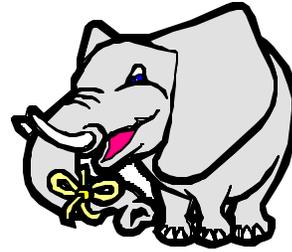   beth.wilson@capitalone.com

Plan ahead and mark your calendars now. The DCSUG will meet on the following dates in 2000: March 7, June 13, September 12, December 5. Meeting times and locations to be announced. If at any time you have a question about an upcoming meeting or other DCSUG information, call one of the members of the Steering Committee.

You're Invited

To visit the DCSUG home page

The URL for the DCSUG home page on the Web:
http://www.ita.doc.gov/industry/otea/dcsug

Washington DC SAS Users Group
P.O. Box 44670
Washington, DC 20026-4670

Remember to Join DCSUG!!!

Don't forget to renew your DCSUG membership for 2000. Dues are used to defray the costs of producing and mailing the newsletter and provide refreshments at our meetings. You'll make sure that you keep receiving the newsletter and keep DCSUG going. If you have not joined DCSUG, now is a great time to do so. Joining DCSUG is as easy as completing the membership form included in this newsletter. Individual memberships are only $10; corporate memberships are $50.