

# DCSUG Newsletter

Washington, DC SAS Users Group

## Fourth Quarter Meeting

When: Tuesday, December 4, 2007  
Time: 8:30a.m. - Noon  
Place: Bureau of Labor Statistics, Conference Center  
Postal Square Building  
Room G440  
2 Massachusetts Ave., N.E.  
(use the First Street entrance)  
Washington DC



Agenda:	8:30 - 9:00	Registration and refreshments
	9:00 - 9:15	Announcements
	9:15 - 10:15	Andrew Karp, Sierra Information Services <b>Getting the Most from PROC MEANS</b>
	10:15 - 10:30	Break
	10:30 - 10:45	User-to User Questions and Answers Moderators - DCSUG officers
	10:45 - 11:45	Mike Molter, PPD Inc. <b>Tips and Tricks for Creating the Reports Your Clients Need to See</b>
	11:45 - Noon	Door Prize Drawing and Meeting Wrap-up

**\*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 11/27/07

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

**NOTE: If you are not currently a paid member of the DCSUG and are planning to attend the meeting, please email Larry Altmayer at [Lawrence.W.Altmayer@census.gov](mailto:Lawrence.W.Altmayer@census.gov) or call him at (301) 763-2569 by November 28th 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.**

\*\*\*\*\*

## President's Message

Mike Rhoads

Once again, it's been a very busy year for the DC SAS user community. In addition to our four quarterly meetings on Tuesday mornings, we have been fortunate enough to also have two well-attended evening get-togethers. Many DCSUGers also participated in the inaugural SAS Global Forum (OK, the renamed "SUGI" conference) in Orlando, not to mention the November NESUG and SESUG conferences in Baltimore and Hilton Head, respectively. Each of the latter was co-chaired by one of our "local heroes" and featured numerous other DC-area speakers and volunteers. 2007 also saw the launching of the sasCommunity.org web site, as well as continued high activity on the SAS-L listserv.

None of this would be possible without many hours of unpaid effort from countless individuals who prepare and give presentations, manage finances, keep track of membership databases, coordinate with SAS Institute and other user groups, maintain web sites, prepare newsletters, and perform numerous other tasks that keep things running. I would like to express my appreciation to the many persons who have contributed and continue to contribute to DCSUG in these ways, not just in 2007 but throughout its lifetime. While I am retiring from active involvement in the group at the end of the year, I expect that I will still be attending some of the meetings, and I am confident that the group will continue to grow and thrive in 2008 and beyond.

## Speaker Abstracts

### Getting the Most from PROC MEANS

Andrew Karp, Sierra Information Services

This presentation shows you ways to tap in to the power and flexibility of PROC MEANS, one of the most powerful (and probably least-understood) analytic tools in BASE SAS Software. The tips and techniques presented in this paper go beyond the basics of PROC MEANS' capabilities and show you powerful functionalities added in SAS Version 8 and SAS 9 Software that make it even easier (and more efficient) for you to carry out the analyses you need. Among the topics discussed are: 1) easier ways to create multiple output data sets in a single PROC MEANS "run" with the CHARTYPE option; 2) using the PRELOADFMT and COMPLETETYPES options to obtain observations in your output data set where no observations in the input data set have a formatted value of the given classification variable; 3) using the AUTONAME option to have SAS generate the names of variables in the output analysis data set; 4) how the IDGROUP option is used to output extreme observations from in the input data set to variables in the output data set; 5) using ordering options for classification variables; 6) how to apply Multilabel Formats to subgroup and present your results; and, 6) the TYPES and WAYS statements. We will also discuss the implementation of threading technology to this procedure in Version 9 and how it may impact the time it takes to use PROC MEANS against your larger data sets.

*Andrew Karp first started using SAS Software as a graduate student at The George Washington University in 1982 and since 1994 has had his own SAS Software consulting and training firm, Sierra Information Services. Although he now lives in the California Wine Country (where he hopes to someday open a SAS training bed and breakfast operation), Andrew frequently returns to the DC metropolitan area to speak at user group events and give training classes. He is a thirteen-time SUGI/SGF invited speaker and has presented papers at events for SAS users in eleven countries.*

### Tips and Tricks for Creating the Reports Your Clients Need to See

Mike Molter, PPD Inc.

Advanced reporting tools such as PROC REPORT have progressed by leaps and bounds in terms of options available to programmers for creating tables. Add to that the always-improving Output Delivery System (ODS) and it seems the sky is the limit for creating ready-for-delivery output straight from SAS. Just when we start believing this, reality exposes limitations. Templates, or shells of tables, including titles, footnotes, column headers, content, right down to the formatting, are designed to maximize readability, and to present the data in a way that allows the reader to make well-informed decisions - and without regard to programming challenges.

Unfortunately for programmers, such tables can't always be generated with the basic tools everyone talks about. In this presentation we will journey into some of the lesser known areas of PROC REPORT and ODS by looking at two such tables and identifying features that seem innocent on the surface, but require some innovation when the time comes to program them. Our journey will take us into PROC REPORT features such as ACROSS variables. We'll go behind the scenes to see how a report is constructed, and ultimately, to the lesser known temporary variable. We'll also get into certain ODS features such as styles, in-line formatting, and the use of destination-specific markup. At the end of our journey we will have solved the challenges of these two tables. Along the way, we will have discovered some of the documentation hiding places of these and many other PROC REPORT and ODS facts.

*Mike Molter is a SAS programmer for PPD and makes his home in Cary NC. With a Bachelors and Masters degree in mathematics from Western Michigan University, he began learning SAS in 1999 in the healthcare industry before moving to clinical trials in 2003. Mike's programming interests include ODS, PROC REPORT, and tool building through the macro facility and PROC TEMPLATE. Personal interests include science and history, cycling, triathlon, and the Detroit Red Wings*

## 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.

## DCSUG 2007 Steering Committee

### Officers

#### **CHAIR -- Mike Rhoads**

(301) 251-4308 (301) 294-2040 (fax)  
rhoasm1@westat.com

#### **SECRETARY -- Larry Altmayer**

(301) 763-2569 (301) 763-8804 (fax)  
Lawrence.W.Altmayer@census.gov

#### **TREASURER -- Anastasiya Osborne**

(202) 690-0446  
stefan.osborne@comcast.net

#### **SAS LIAISON -- Roger Goodwin**

(703) 877-8000 ex 120  
roger\_goodwin@nass.usda.gov

#### **OTHER COMMITTEE MEMBERS**

##### **Karen Dennis**

kcdennis@comcast.net

##### **Frank Fry**

(703) 250-9272  
frank.fry@cox.net

##### **Paul Gorrell**

(301) 628-3237 (301) 628-3201 (fax)  
pgorrell@s-3.com

##### **Howard Schreier**

(703) 979-2720

[hs@dc-sug.org](mailto:hs@dc-sug.org)

##### **Arlene Siller**

(301) 458-4498 (301) 458-4032 (fax)  
asiller@cdc.gov

##### **Christopher Zogby**

(703) 380-0973  
info@zogbyenterprises.com

#### **DESKTOP SPECIAL INTEREST GROUP**

##### **CO-CHAIR -- Joe Mirabal**

(703) 265-4620  
jmmirabal@aol.com

##### **CO-CHAIR -- Yesvy Gustasp**

(301) 625-9576 (202) 307-5888 (fax)  
Yesvy\_Gustasp@yahoo.com



### To Our Sponsors

These organizations make it possible for DCSUG to host quarterly meetings, featuring informative presentations about SAS software, while keeping dues affordable



**Bureau of Labor  
Statistics**



**Deloitte Services LLP**



**Sierra Information  
Services**



**Social & Scientific  
Systems, Inc.**



**Westat**



**Zogby Enterprises, Inc.**

# 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 one of the DCSUG officers. There is a DCSUG directory with phone, fax, and E-mail contact information elsewhere in this newsletter.

## December

4     **DCSUG**     DC SAS Users Group, General Meeting  
Time:       8:30 am - 12:00 noon  
Site:        Bureau of Labor Statistics  
              2 Massachusetts Ave. NE  
Program:    See front page announcement  
Contacts:    Roger Goodwin at (703) 877-8000 ext 120  
              *Roger\_Goodwin@nass.usda.gov*  
              Mike Rhoads at (301) 251-4308  
              *MikeRhoads@Westat.com*

## March 2008

11     **DCSUG**     DC SAS Users Group, General Meeting  
Time:       8:30 am - 12:00 noon  
Site:        TBA  
Program:    TBA

16-19 **SAS Global Forum 2008**  
Chair:       Warren Stinson  
Site:        San Antonio  
Further Info:  
<http://support.sas.com/events/sasglobalforum/2008/index.html>

## June

3     **DCSUG**     DC SAS Users Group, General Meeting  
Time:       8:30 am - 12:00 noon  
Site:        TBA  
Program:    TBA

## Other Groups

**CENSUG**     Census SAS Users Group  
Contact:     Andrea Potter at (301) 763-1687

**MDSUG**     Maryland SAS Users Group  
Contact:     Jayne May Miller, (410) 455-6849  
              *jmiller@chpdm.umbc.edu*  
              <http://www.chpdm.org/mdsug.htm>

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

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

**SIGSTAT**    Special Interest Group in Statistics, CPCUG and WINFORMS  
Contact:     Charlie Hallahan, 202-694-5051  
              *hallahan@ers.usda.gov*

**VASUG**     Virginia SAS Users Group (Richmond area)  
Contact:     Carrie Mariner  
              *liaison@vasug.org*  
              <http://www.vasug.org>

# SASTalk - Unique SAS® Names in a Macro Context

by Ian Whitlock

A recent question on SAS-L asked how to guarantee unique temporary data set names in macros where the caller may have important data sets stored in the work library.

The suggestion to use the SAS default names DATA1, DATA2, etc. is not perfect. For example, the user first creates his important data set and calls it DATA2. Then he calls a macro that creates two data sets using default names. By convention the names will be DATA1 and DATA2 on the first call. Hence the user's important temporary data set is overwritten with the new data from the macro. SAS keeps track of its names and guarantees that the number always increases, but it does not check data set names created explicitly. Another problem with this suggestion is that it becomes hard to follow the data set names in a complex program.

The problem is really one of names in general. How can you guarantee that clashes will not occur? Many modern languages use the concept of name spaces to alleviate the problem. In this case the same name can be used twice, as long as each name is in a different name space. For example, in SAS macro you can declare macro variables as local. In this case the same name can be used in different macros or even the same macro as long as they are in different executions. However, SAS does not use this concept in a more general setting. In particular, there is no concept of local data set names.

I tend to create artificial name spaces by choosing a three letter abbreviation for each macro and then naming the objects created by using that three letter abbreviation followed by an underscore as a prefix to each name when it is not the name of a local macro variable. This too has problems when the system is large and/or more than one person is working on the code.

Long ago Paul Dorfman introduced a solution, still not perfect, but one in which clashes would be extremely rare. His suggestion was to generate seven random digits and use that as a suffix to names. (The suggestion was made back in the days when SAS names could be no longer than 8 characters.) Since many names can now be 32 character long, we can extend the idea to 14 random digits with little trouble. Here is a macro to generate up to 14 random digits.

```
%macro ranpart (len=14,seed=0) ;
  *** return random integer of &tlen digits
      where tlen is minimum of &len and 14
  ***;
  %local r x tlen ;
  %let r = %sysevalf(10**14 * %sysfunc(ranuni(&seed))). ;
  %let tlen = %sysfunc(min(&len,14)) ;
  %let x = %sysfunc(min(&len,%index(&r,.)-1)) ;
  *** value to return ***;
  %sysfunc(putn(%substr(&r,1,&x),z,&tlen))
%mend ranpart;
```

In macro, the function RANUNI keeps track of the seed from the first call across all macros. Hence two calls to %RANPART will generate digits from the same sequence of random numbers. This means that the values will be different with a probability close to one when the length 14 is used and the number of names is small, say less than a million. The sequence will not uniquely correspond to objects in a program, where the program does not always generate the same specific objects in the same order, or when there are other arbitrary calls to RANUNI via %SYSFUNC.

The only time the uniqueness of names is important is when one wants to compare two different runs. In this case the numbers should be prepared ahead of time and put into a file or a sequence of global macro variables, say G\_MV1, G\_MV2, ... Of course, the file solution is preferable to avoid pollution of the global macro environment.

To create a specific important work data set, you might have as part of a macro:

```
%local my_set ;
%let my_set = w%ranpart() ;
data &my_set ;
    ...
run ;
```

Note that the particular random number used is not important to the code since it is referred to by one constant reference, MY\_SET. Of course, you have to take appropriate steps to protect this value. Usually declaring the macro variable local is sufficient, since macros should be relatively modularized and tend to be written and maintained by one person.

Consequently, Paul's suggestion gives a practical solution to the problem of name clashes. However, one can simplify in the case of naming temporary data sets. Let each macro create its own subdirectory of the WORK directory and eliminate it when finished with it. Here is an example.

```
options xsync noxwait ;
%let my_dir = %ranpart() ;
%let path = %sysfunc(pathname(work))\&my_dir ;
%put path=&path ;
%let rc = %sysfunc(system( md "&path" )) ;
%put rc for subdirectory is &rc ;
%let my_lib = q%ranpart(len=7) ;
libname &my_lib "&path" ;

data &my_lib..w ;
    x = 1 ;
run ;
%let rc = %sysfunc(system( del /Q "&path\*.*" )) ;
%put rc for deletion of subdirectory is &rc ;
```

If you do not delete the subdirectory, then the work directory is not automatically deleted when the session ends since it contains non-empty data sets. If the deletion is done in a DATA step, then one benefit is that if the job aborts, the work directory will remain intact and can be inspected after the job finishes.



## Nominations for 2008 DC SAS Users Group Officers

It's that time again -- we need to elect DCSUG officers for next year and we need *your* help.

The election will be conducted at the 4th quarter DCSUG meeting on Dec. 4 at the BLS.

Please contact any of the current DCSUG officers (please see page 3) if you have questions or are interested in running for one of these positions:

*President*  
*Secretary*



*Treasurer*  
*SAS Liaison*

## SAS Software Training Opportunities in the Washington, DC Area

Please visit the SAS Institute customer support site, [support.sas.com](http://support.sas.com), to access the current schedule of public training seminars offered at their training centers in Arlington and Richmond, VA, Washington, DC and Rockville, MD.

In addition, third-party vendors periodically offer SAS training in the Washington DC area.

Sierra Information Services will present a range of full- and half-day SAS training sessions at the Johns Hopkins Univ. Washington, DC Center near Dupont Circle in January 2008. For more information, including seminar descriptions and on-line registration, go to [www.SierraInformation.com](http://www.SierraInformation.com). Their seminar line-up is:

- Jan. 9 **SAS 101: An Introduction to the SAS System**
- Jan. 10 **Summarizing and Reporting Data Using the SAS System**
- Jan. 11 **SAS 201: Advanced SAS Programming Techniques**
- Jan. 22 **Advanced SAS Reporting Techniques**
- Jan. 23 **Success with the SAS Output Delivery System**
- Jan. 24 **Exchanging Data Between SAS and Microsoft Products (AM)**
- Jan. 24 **SAS Data Step Efficiencies (PM)**
- *Note: Two half-day seminars on 1/24. Attend both for the price of one full-day registration.*
- Jan. 25 **Getting the Most from SAS 9 Software**



*“Data Mining Using SAS @ Enterprise Miner” introduces the reader to a wide variety of data mining techniques in SAS® Enterprise Miner v4.3. This first-of-a-kind book explains every node that is a part of Enterprise Miner with regard to SEMMA design and data mining analysis. Each chapter starts with a short introduction to the assortment of statistics that are generated from the various nodes, followed by detailed explanations of the option settings and results that are generated within each node.*

Highlights of the book are the following:

- An overview to the Sampling nodes in SAS® Enterprise Miner to read, randomly sample, and partition the data.
- An overview to the Exploring nodes that are usually used as a preliminary step in data mining to graphically view the data.
- An overview to the Modifying nodes that are designed to perform preliminary modifications to the SAS data set.
- An overview to the various Modeling nodes and modeling designs that can be performed within SAS® Enterprise Miner.
- An overview to the Assessment node that evaluates and assesses the results that are generated from the various models.

**To order call 877-762-2974 or visit <http://www.Wiley.com>**

## 2008 Meetings

Plan ahead and mark your calendars now. The DCSUG meetings in 2008 are scheduled for Mar 11, June 3, Sept 9, Dec 9. 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 are  
invited*

**To visit the DCSUG home page**

The URL for the DCSUG home page on the  
Web:

<http://dcsug.org>

## Remember to Join DCSUG!!!



Don't forget to renew your DCSUG membership for 2008. 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 membership is \$50 and corporate sponsorship is \$250