

DCSUG Newsletter

Washington, DC SAS Users Group

First Quarter Meeting

When: Tuesday, March 4, 2003
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	Roger Goodwin, U.S. Bureau of the Census Developing SAS® Software for Generating a Complete Set of Ratio Edits
	10:15 - 10:30	User-to User Questions and Answers Moderators - DCSUG officers
	10:30 - 10:45	Break
	10:45 - 11:15	David Chapman, U.S. Bureau of the Census Using PROC REPORT To Produce Tables With Cumulative Totals and Row Differences
	11:15 - 11:45	Christopher Zogby, Zogby Enterprises, Inc. Proc SQL: Conditional Logic, Macro Variables and Subqueries
	11:45 - Noon	Wrapup

*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 02/20/03

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 kdennis@enterprise.foundation.org or call (410) 772-2780 by Feb 26th 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

Peter Knapp

In the next two months, DCSUG will be hosting two meetings. The first is a DCSUG Desktop SIG that will be held on Wednesday, February 26, and feature a discussion on exploratory data analysis given by David L. DesJardins. Please see the article about this meeting elsewhere in this newsletter. The second meeting, the regular quarterly DCSUG meeting, will be held on March 4. We have three speakers lined up; David D. Chapman who will be presenting his PROC REPORT talk from last year's SUGI, Christopher Zogby who will be finishing his SQL talk from the December DCSUG meeting, and Roger Goodwin, who will be talking about software developed for survey data edits.

Also, for those lucky enough to be able to make the trip, this year's SUGI will be held in Seattle, WA from March 30 – April 2. I hope to see you at one of these gatherings!

Speaker Abstracts

Using PROC REPORT To Produce Tables With Cumulative Totals and Row Differences

David Chapman, U.S. Bureau of the Census

PROC REPORT is a tool for tabulating and reporting the contents of SAS® data sets. PROC REPORT's strength is its ability to create table columns, modify individual cell values, and use information from different rows of the report. This flexibility allows users to create tables that show cumulative totals and row differences. Cumulative totals are used to report values such as the number of responses year-to-date, percent of a total for all periods to date, and cumulative distribution functions. Row differences show changes between rows. Two PROC REPORT features that allow this are the COMPUTE block, and the distinction that PROC REPORT makes between report and data variables. This presentation discusses how PROC REPORT builds a report, COMPUTE blocks, and what report and data variables are and how they work. The code to produce one and two level tables containing both cumulative row totals and row differences are shown with explanatory comments.

David Chapman is a Statistician and SAS programmer in the Economic Statistical Methods and Programming Division of the US Census Bureau. This paper is based on work he is currently doing to create SAS data sets that provide users access the Census Bureau's Business Register. He is a SAS Certified Professional - Business Intelligence V6

Developing SAS® Software for Generating a Complete Set of Ratio Edits

Maria Garcia and Roger Goodwin, U.S. Bureau of the Census

Survey data editing using the Fellegi-Holt model requires the complete set of explicit and implicit edits. This report presents new SAS® software that generates the complete set of implicit (implied) edits from a given set of explicit ratio edits. We describe the corresponding methodology, along with detailed discussions on the considered alternatives. The new software implements a shortest path algorithm and borrows ideas from the Generate Edits portion (subroutine) currently used in the U.S. Census Bureau's Plain Vanilla Ratio Module.

Roger Goodwin started working at the Census Bureau in 1999. He received his BS in Computer Science in 1988 from Old Dominion University and an MS in Applied Statistics in 1998 from Old Dominion University. He has taught computer software courses at a proprietary school, and has taught a math course at Prince George's Community College for the Fall 1999 - Spring 2000 academic year. At the Bureau, Mr Goodwin has worked in the Decennial Directorate, and the Economic Directorate.

Proc SQL Tips: Conditional Logic, Macro Variables and Subqueries.

Christopher Zogby, Zogby Enterprises, Inc.

Proc SQL offers a number of features that make it a significant coding asset. For a variety of coding tasks, proc SQL can offer the most compact and novel solution. In this paper I will discuss three features that have helped increase my SQL mileage.

Chris Zogby is a Washington, DC area data mining consultant and proprietor of Zogby Enterprises, Inc.

DCSUG 2003 Steering Committee Officers

CHAIR -- Peter Knapp

(202) 482-1359 (202) 482-1388(fax)

peter_knapp@ita.doc.gov

SECRETARY -- Paul Gorrell

(301) 628-3237 (301) 628-3201 (fax)

pgorrell@s-3.com

TREASURER -- Arlene Siller

(301) 458-4498 (301) 458-4032 (fax)

asiller@cdc.gov

SAS LIAISON -- Mike Rhoads

(301) 251-4308 (301) 294-2040 (fax)

rhoadsm1@westat.com

OTHER COMMITTEE MEMBERS

Larry Altmayer

(301) 763-2569 (301) 457-4437 (fax)

Lawrence.W.Altmayer@census.gov

Karen Dennis

(410) 772-2780 (410) 294-715-2310 (fax)

kdennis@enterprisefoundation.org

Frank Fry

(202) 452-2666 (202) 452-6433 (fax)

frank.fry@frb.gov

Christopher Zogby

(703) 671-6330 (703) 671-6360 (fax)

czogby@erols.com

Linda Atkinson

(202) 694-5046 (202) 694-5718 (fax)

Atkinson@ers.usda.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

CO-CHAIR -- Yesvy Gustasp

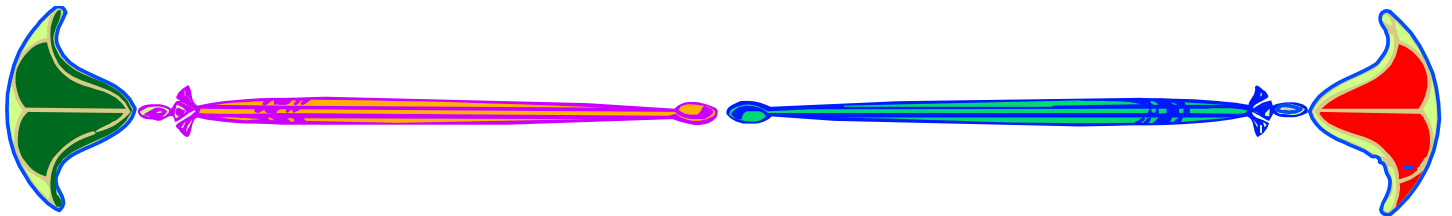
(301) 625-9576 (202) 307-5888 (fax)

Yesvy_Gustasp@yahoo.com



We at DCSUG are proud to announce a new URL for our web site:

<http://dc-sug.org>



DCSUG Desktop Special Interest Group

The next meeting of the DCSUG Desktop SIG will be held on Wednesday, February 26, 2003, from 12:30 to 1:30 pm. See below for location and directions.

The program will revolve around a presentation on EDA (exploratory data analysis) using JMP® software, and, in particular, using easily generated "live" graphs to effectively edit and analyze statistical data. The speaker will be David L. DesJardins; he will cover material from his NESUG & SUGI papers on the subject. Questions regarding corresponding use of JMP's point-and-click "sister" package, SAS/INSIGHT software, will also be answered.

This is a joint meeting with SIGSTAT, which is the joint Special Interest Group in Statistics for CPCUG (Capital PC User Group) and WINFORMS (Washington Institute for Operations Research and the Management Sciences).

Location: Room S3031, 1800 M St. NW, Washington DC. Directions: <http://www.cpcug.org/user/sigstat/directns.htm>

If this is your first SIGSTAT meeting, contact Charlie Hallahan either by email (hallahan@ers.usda.gov) or phone (202-694-5051) and leave your name in order to enter the building. The SIGSTAT "Admit" list is not the same as DCSUG's, so if you have attended DCSUG meetings in the past, but not SIGSTAT meetings, you should contact Mr. Hallahan.

One Easy Way to Modify The Numeric Format of SAS Output

Paul Gorrell
Social & Scientific Systems, Inc.

I've always been a bit frustrated that I couldn't choose a format for the output statistics for PROCs such as FREQ and MEANS. One reason for this is that I never like to see scientific notation for numbers that exceed the SAS default format. Consider PROC FREQ output with the LIST option. If you look at the first table on the next page you'll see that, after the second row, the cumulative frequency is given in scientific notation. Let's see why this happens. The ODS output object associated with the LIST option is called BASE.FREQ.LIST (you can find this out by running PROC FREQ inside of ODS TRACE ON/OFF statements (see *The Complete Guide to the SAS Output Delivery System, Version 8*, pp. 71-72). The default format for frequency output with BASE.FREQ.LIST is BEST8., and any value too large for this format is given in scientific notation (Note that the V5FMT option can't help here).

Fortunately the EDIT statement in PROC TEMPLATE offers an easy way to modify numeric formats for procedure output. SAS stores system template information in SASHELP.TMPLMST. By default, user created and modified templates are stored in SASUSER.TEMPLAT. The ODS PATH statement in line '01' below is probably your system default. That is, system templates are read-only (READ), but user templates can be created, modified or deleted (UPDATE).

The following program changes the default format for frequency list output from BEST8. to COMMA11., and the percent format from 6.2 to 5.1. Lines '02' – '17' are a PROC TEMPLATE step with nested EDIT statements. Like DO, each EDIT statement must be closed with an END. The outer EDIT statement has the output object as its argument. The four inner EDIT statements take particular output statistics as their arguments (to find out the output names and default formats you would use the SOURCE statement in PROC TEMPLATE (see p. 232 of *The Complete Guide...*). Once the template has been modified, the PROC FREQ output will use the specified numeric formats, as shown in the second table.

Lines '24' – '26' delete the user-defined template that was just created (this is an option you have).

```
01. ODS PATH SASUSER.TEMPLAT(UPDATE) SASHELP.TMPLMST(READ) ;
02. PROC TEMPLATE;
03.   EDIT BASE.FREQ.LIST;
04.     EDIT CUMPERCENT;
05.       FORMAT=5.1;
06.     END;
07.     EDIT CUMFREQUENCY;
08.       FORMAT=COMMA11.;
09.     END;
10.     EDIT PERCENT;
11.       FORMAT=5.1;
12.     END;
13.     EDIT FREQUENCY;
14.       FORMAT=COMMA11.;
15.     END;
16.   END;
17. RUN;

18. PROC FREQ DATA= MEPS1998;
19.   TABLES SEX*CENSUS_REGION
20.   / LIST ;
21.   FORMAT SEX SEX. CENSUS_REGION REGION.;
22.   WEIGHT WTDPER98;
23. RUN;

24. PROC TEMPLATE;
25.   DELETE BASE.FREQ.LIST;
26. RUN;
```

The output in the second table is much easier to read than the first table. Of course, if you want to keep two decimal places for the percent columns, you could just use the defaults. Also, the format modifications in lines '02' – '17' apply to all subsequent PROC FREQ output when the LIST option is used. For example, suppose, in addition to LISTING output (as shown here), you also want RTF output with user-specified formats. Then you would only need to add the appropriate ODS RTF statements.

You can modify the output properties of other PROCs as well. For example, if you want to modify the output of PROC MEANS/SUMMARY, you would edit BASE.SUMMARY instead of BASE.FREQ.LIST. In general, the first thing you need to do is use the ODS TRACE and the PROC TEMPLATE SOURCE statements to find out the names and keywords of whatever you want to modify. Of course you can't do everything, and finding out what can and cannot be done takes some patience (e.g. you can only modify PROC FREQ crosstab output for the LIST option). SAS programmers have long wished they could have more control over procedure output. Sometimes PROC TEMPLATE feels like a reminder that you should be careful what you wish for. There's a lot to learn. In fact, PROC TEMPLATE (along with SQL and TABULATE) is one of the few SAS procedures that could easily have its own *SAS Guide* (the chapter on PROC TEMPLATE in *The Complete Guide...* is over 100 pages). Fortunately, the EDIT statement is a simple and easy way to get started, and to make a useful improvement with minimal time and effort.

1998 MEDICAL EXPENDITURE PANEL SURVEY POPULATION BY SEX, CENSUS REGION
(WHERE CENSUS REGION IS KNOWN)

USING THE DEFAULT NUMERIC FORMATS

SEX	CENSUS_REGION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
MALE	NORTHEAST	25020048	9.23	25020048	9.23
MALE	MIDWEST	31143573	11.49	56163621	20.73
MALE	SOUTH	46003717	16.98	1.0217E8	37.70
MALE	WEST	30237621	11.16	1.324E8	48.86
FEMALE	NORTHEAST	26866252	9.91	1.5927E8	58.77
FEMALE	MIDWEST	32100516	11.85	1.9137E8	70.62
FEMALE	SOUTH	48641677	17.95	2.4001E8	88.57
FEMALE	WEST	30979548	11.43	2.7099E8	100.00

USING SPECIFIED NUMERIC FORMATS (PROC TEMPLATE)

SEX	CENSUS_REGION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
MALE	NORTHEAST	25,020,048	9.2	25,020,048	9.2
MALE	MIDWEST	31,143,573	11.5	56,163,621	20.7
MALE	SOUTH	46,003,717	17.0	102,167,337	37.7
MALE	WEST	30,237,621	11.2	132,404,958	48.9
FEMALE	NORTHEAST	26,866,252	9.9	159,271,210	58.8
FEMALE	MIDWEST	32,100,516	11.8	191,371,727	70.6
FEMALE	SOUTH	48,641,677	17.9	240,013,404	88.6
FEMALE	WEST	30,979,548	11.4	270,992,952	100.0

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

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: Mike Rhoads at (301) 251-4308
RhoadsM1@Westat.com
Peter Knapp at (202) 482-1359
peter_knapp@ita.doc.gov

March/April

30-2 **SUGI 28** 28th Annual SAS Users Group International Conference
Site: Seattle
Chair: Art Carpenter
Further Info: support.sas.com/usergroups/sugi/sugi28/index.html

June

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

September

7-10 **NESUG 2003** NorthEast SAS Users Group Conference
Site: Arlington VA (Crystal Gateway Marriott)
Co-Chairs: Larry Altmayer, Ralph Leighton
Further Info: www.nesug.org

22-24 **SESUG '03** SouthEast SAS Users Group Conference
Site: St. Pete Beach FL
Co-Chairs: Steve Noga, Gary Schlegelmilch
Further Info: www.sesug.org

Other Groups

CENSUG Census SAS Users Group
Contact: Andrea Potter at (301) 457-1687

MDSUG Maryland SAS Users Group
Contact: Mr. R.H. Miller, (410) 740-4229
millerrh@juno.com
www.umbc.edu/chpdm/mdsug/

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

NOVASUG Northern Virginia SAS Users Group
Contact: Gary McQuown at (703) 628-5681
gmcquown@dasconsultants.com
www.novasug.org

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: Mark Jordan at president@vasug.org
www.vasug.org



The next NESUG Conference will be held Sept. 7 through Sept. 10 in the Washington, DC locale. The NESUG 2003 Conference hotel will be the Crystal Gateway Marriott across the Potomac in Arlington. The location of the conference hotel permits easy access to the mall area in DC (where the museums are): only four stops on the DC Metro. Bring your family!

Highlights of the Conference will include the NESUG (SAS) Exhibit and Demo Room, presentations by SAS and by fellow SAS users, a vendor pavilion, where you can meet service providers and solution providers, pre-conference training, the opportunity to meet and mingle with fellow SAS users, and special events, such as the attendee party and possibly a Potomac River dinner cruise.

The core of the conference are the presentations, by SAS technical staff and by a host of fellow SAS users: six tracks spanning two and a half days, from the ever popular Beginning and Advanced Tutorials, to Statistics and Econometrics, SAS Applications and the Emerging Technologies section.

This year, all paper abstracts should be submitted via the NESUG 16 website, either directly (<http://www.nesug.org/NESUG162003.html>), or accessible via the link from the NESUG website (www.nesug.org). The deadline for abstract submission is March 27. You may also participate in such ways as a registration desk volunteer, code clinician or session coordinator, by indicating interest on the on-line registration form, beginning in mid-May.

For more information, contact one of your NESUG 2003 Conference Co-chairs:

Larry Altmayer
(301) 763-2569
lawrence.w.altmayer@census.gov

Ralph Leighton
(860) 547-3014
rleighton@thehartford.com



Call for Papers and Participation

11th Annual Conference of the SouthEast SAS Users Group

September 22-24, 2003

Trade Winds Island

Grand Resort

St. Pete Beach, FL

visit www.sesug.org for more information

Co-chairs

Steve Noga – snoga@rhoworld.com

Gary Schlegelmilch – sesug03@juno.com

