

PRE-CONFERENCE WORKSHOPS

Section Chairs: Deb Cassidy
Cardinal Health
Dublin, OH

Adding pre-conference workshops **enhances your SESUG conference experience** by giving you the opportunity to receive in-depth instruction from well-known SAS experts, on a wide variety of topics, at reasonable rates.

In addition, the discounted lodging rate at the conference hotel is available if you need to arrive early in Nashville to attend the workshops. The SESUG conference leadership has selected these instructors based on their subject matter expertise, presentation skills, and enthusiasm for teaching others about SAS. The table below shows the wide array of learning opportunities available by attending these sessions. Descriptions of the workshops and instructor profiles are on the following pages. You can register for these sessions when you register for the SESUG '04 conference by using the Conference Registration Form from the registration book or register online at www.sesug.org. Tuition, which includes a copy of the workshop materials and morning/afternoon refreshments, is only **\$99** for the half-day workshops. Full-day workshops are \$198.

Attendees at the workshops who do **not** also register for SESUG '04 are charged an additional \$25 per person administrative fee.

SESUG and the instructors will hold all workshops with pre-registered attendees unless there is an unexpected event preventing the instructor from attending the conference. If a workshop is cancelled, the sole liability of SESUG and/or the instructors is a refund of fees paid for the cancelled workshop, and they are **NOT** liable for special or consequential damages arising from the cancellation.

Workshop Times: The workshops will start promptly at the hour listed. Refreshments will be available at 7:30 am for the morning workshops. Workshops will also have a mid-morning and mid-afternoon refreshment break. Lunch is on your own. Full day workshops will break from 12:00 noon to 1:00 pm for lunch.

Saturday, October 30th, 2004

<u>Time</u>	<u>Instructor</u>	<u>Title</u>
1:00 pm - 5:00 pm	Patricia Cerrito	Using SAS/Stat to Mine Data
1:00 pm - 5:00 pm	Steven First	SAS Efficiencies – Techniques for Improving Your Code

Sunday, October 31st, 2004

8:00 am - 12:00 noon	Steven First	What's New in SAS 9 - An Overview of a Few Hundred New Things
8:00 am – 5:00 pm	David Cassell	Sampling with the SAS System
8:00 am – 5:00 pm	Frank Dilorio	Building SAS Utilities: Putting Dictionary Tables, SQL and the Macro Language to Work
1:00 pm - 5:00 pm	Mal Foley	Merging, Combining and Subsetting Files(Tricks, Traps and Techniques)

Using SAS/Stat® to Mine Data

Instructor: Patricia B. Cerrito

When: Saturday Oct 30, 1:00pm-5:00 pm

Abstract: Data mining is a process of investigating large datasets so that the investigator can “drill down” into the data to find meaningful patterns and relationships. Data mining relies on the entire distribution of data, instead of just using means and standard deviations.

Although Enterprise Miner was designed to do data mining, the techniques in SAS/Stat can also be used. Data mining includes using a variety of methods to investigate the data to discover patterns, reserving some of the data for the purpose of validation of results. The importance of validation cannot be over-emphasized.

One of the most important methods of data mining is that of visualization. There are many ways to examine the data, including kernel density estimation. Kernel density can be used to compare different populations without reducing the dataset to averages. Enterprise Guide also has some excellent graphics that can be used to examine the data. Simple data transformation can be used to find more meaningful information from the data. Although inferential statistics are used to prove hypotheses, it is the visualization that convinces of the “rightness” of the results.

If the dataset has a large number of variables, or data fields, it is possible to use some type of factor or cluster analysis to reduce the number to a meaningful few. While factor analysis is often used to classify variables into categories, it can be used as a reduction technique so that the total number of variables is more manageable in standard inferential techniques such as regression. For example, a store inventory can have thousands or tens of thousands of categories. The general linear model was never designed to handle so many.

Examples of “drill down” results will be provided, and examined in terms of the statistical methods used. Several demonstrations will show the effectiveness of this “drill down” into the data.

Recommended Background: Although some background in statistics is helpful, it is NOT required.

Intended Audience: Those who want to learn more about the data mining process. This workshop is for a very general audience and not just statisticians.

Materials Provided Attendees: Hard-copy of slides

SAS Efficiencies-Techniques for Improving Your Code

Instructor: Steven First

When: Saturday Oct 30, 1:00pm-5:00 pm

Abstract: This workshop will cover how SAS processes large data files and will show several techniques for improving a SAS program. Emphasis will be on the DATA step, with a brief look at improving the use of several SAS PROCs. The objectives of this workshop are to:

- understand how to measure “efficiency”
- understand how the data step handles data, both input and output
- learn how to alter SAS column attributes as needed for correctness and efficiency
- practice techniques for reducing the amount of data SAS has to process
- learn to use SAS programming logic in an efficient manner, and understand why some programs are less efficient than others
- review using FORMAT statements effectively
- understand when, how, and why to sort or index SAS data files
- discuss storage of compiled DATA step source code
- utilize WHERE statements to maximize coding efficiency
- understand when to use SAS VIEWS and PROC SQL
- know how and when to use data compression

Recommended Background: A working knowledge of the fundamentals of SAS is required.

Intended Audience: This course is designed for students who have experience with SAS and want to enrich their SAS skills and improve the efficiency of their SAS programs.

Materials Provided Attendees: Hard-copy of slides

What’s New in SAS 9-An Overview of a Few Hundred New Things

Instructor: Steven First

When: Sunday Oct 31, 8:00am-12:00 noon

Abstract: What? A few hundred new things? Don’t get overwhelmed. There are over 200 new things in just the SAS functions section. This workshop is designed to familiarize students with SAS Version 9. Topics to be covered included:

- Transitioning from version 8 to 9
- Overview of enhancements in version 9
- SAS Explorer Window
- SAS Management Console
- Revised Procedures in version 9
- The Migrate Procedure
- SAS System Options
- SAS Functions and Perl Regular Expressions
- SAS Formats and Informats
- New Libname Engines including Access and Excel
- Importing Data in version 9
- Exporting Data in version 9
- Multi-threaded parallel processing

Recommended Background: A good working knowledge of the programming capabilities of SAS in previous versions.

Intended Audience: Those who already use an earlier version of SAS and want to learn about the new features found in SAS 9.

Materials Provided Attendees: Hard-copy of slides

Sampling with the SAS® System

Instructor: David L. Cassell

When: Sunday, Oct 31, 8:00am-5:00 pm

Abstract: This workshop is designed to provide you with the skills and insights necessary to apply the SAS System's survey sampling procedures to your projects. This workshop is tailored for the knowledge worker who is confronted with standard business and scientific problems, such as:

- Developing cost-effective survey samples that can properly address business needs
- Avoiding the often unintended consequences of poorly-designed survey samples
- Preventing the biases and errors that are incurred when incorrectly analyzing survey samples using standard tools such as PROC REG and PROC MEANS
- Coping with missing data in survey samples.

Through instructor's presentation and the examination of example SAS programs provided in the course materials, you will learn about the basic concepts of survey sampling, as well how to apply these concepts using the SAS System.

Recommended Background: Prior experience with PROC MEANS and PROC REG, as well as familiarity with statistical concepts such as confidence intervals, p-values, population parameters and their estimates, slope and intercept, dependent and independent variables in regression, etc. will enhance the value you will receive from attending this workshop.

Intended Audience: SAS users who currently need to or will want to analyze data that are from sampling designs or develop cost-effective sampling designs. Participants **do not need a deep statistical background**, nor do they need an expert grounding in SAS DATA steps and basic SAS procedures.

Materials Provided Attendees: Hard-copy and diskette of slides and SAS programs, diskette of SAS datasets

Building SAS Utilities: Putting Dictionary Tables, SQL, and the Macro Language to Work

Instructor: Frank Dilorio

When: Sunday, Oct 31, 8:00am-5:00 pm

Abstract: Let's start by acknowledging a simple truth – programmers are, by nature, a lazy lot. We want data to be normalized so we don't have to repeat values more than necessary. We want functions to do routine tasks. We want ODS to figure out how to render our procedure output in HTML.

We also want to avoid writing the same type of program more than once. Given an eye for repetition, a little imagination, and the right tools, it's possible to create elegant, generalized, reliable utility programs that can greatly improve program development time and reliability.

This workshop addresses three components of Base SAS that are vital to developing utilities: dictionary tables, the SAS macro language, and SQL. We'll look at some of the more popular tables in detail, then review the macro language and SQL, emphasizing the aspects of these "languages within a language" that are most relevant to building general-purpose utility programs. With these tools in hand, we then review a set of guidelines for building utilities, using a series of real-world examples as we go.

The workshop emphasizes Windows platforms using Base SAS Version 9.0. We note incompatibilities with earlier Versions when appropriate. Concepts and techniques are applicable across SAS and operating system versions, and can be fruitfully extended into use with non-Base products such as SAS/GRAPH, SAS/IntrNet, *et al.*

Recommended Background: Familiarity with SQL and macro language

Intended Audience: Intermediate to advanced SAS programmers who are interested in developing general-purpose utilities.

Materials Provided Attendees: Hard copy of slides; hard copy of listings, logs, etc that expand on material in the slides; CD with sample code used in the workshop

Merging, Combining and Subsetting Files (Tricks, Traps, and Techniques)

Instructor: Malachy (Mal) J. Foley

When: Sunday, Oct 31, 1:00pm-5:00 pm

Abstract: This workshop discusses over 30 common errors in merging, combining, or otherwise manipulating data. All of these errors result in erroneous data. Most of them occur with no SAS message or warning! In addition to examining errors, this covers a wide range of topics such as interleaving, subsetting, concatenations, the IN= data set option, BY groups, FIRST.variable, program data vectors (PDV), Cartesian products, one-to-one merges, match merges, and fuzzy merges.

The workshop starts with the basics and continues to build up to complex and tricky examples of data set manipulation. The only prerequisite for the course is a working knowledge of the SAS DATA Step. Yet, this workshop will give intermediate and advanced programmers a great review and some surprises. Come see what mysteries lurk in manipulating SAS files!

Recommended Background: Working knowledge of the SAS DATA Step

Intended Audience: Anyone who manipulates SAS data sets

Materials Provided Attendees: Hard-copy of slides PLUS more in-depth material

Workshop Instructor Profiles

Here are the biographies for instructors providing pre-conference workshops at SESUG '04. We have chosen them to present workshops in conjunction with our annual conference because of their reputation within the SAS user community for being subject-matter experts who have excellent presentation skills and for their enthusiasm to share their knowledge with others.

David L. Cassell

David L. Cassell is co-director of Design Pathways, a consulting company based in Corvallis, Oregon providing survey sampling design and analysis, experimental design, statistical analysis, SAS programming, and SAS systems design to public and private sector clients. He is also a contractor for the U.S. Environmental Protection Agency, where his primary responsibility is design and analysis of environmental survey designs for the U.S. EPA's EMAP program.

A 25-year SAS Software user, David has applied SAS System capabilities to generate and analyze survey sampling designs for environmental monitoring that have been implemented across the United States, as well as in a number of foreign countries. David's expertise with the SAS System has been recognized by his peers within the user community in several ways including being named to the SAS-L Hall of Fame. David has given award-winning presentations at numerous SAS user group meetings on topics ranging from survey sampling to the integration of SAS and PERL.

Patricia B. Cerrito

Patricia B. Cerrito is a Professor of Mathematics and Biostatistician at the University of Louisville. She also collaborates with medical researchers through the Jewish Hospital Center for Advanced Medicine in Louisville, Kentucky. She has worked with SAS Enterprise Miner since its introduction, and was awarded a pilot project with the SAS Institute to develop medical applications of SAS Text Miner.

Dr. Cerrito has received funding from the National Institutes of Health to use text and data mining techniques to investigate large, clinical databases derived from electronic patient records, and frequently integrates methods in Enterprise Miner with methods in SAS/Stat. She has been a SAS user for over 25 years. Recently she has given a WebCast in conjunction with the SAS Institute on the use of Text Miner to investigate risk-adjusted severity models used to rank hospitals as to quality. She is under contract with Books by Users to complete a textbook on Enterprise Miner.

Frank Dilorio

Frank Dilorio is President of CodeCrafters, Inc. and the author of "SAS Applications Programming: A Gentle Introduction" and (with Ken Hardy) "Quick Start to Data Analysis with SAS." Both titles are part of SAS Institute's Books by Users series and have

sold over 25,000 copies. Frank has been active in the SouthEast SAS Users Group (SESUG) since its inception, co-chairing the 1994 and 1996 conferences.

He has, much to his astonishment after doing the math, nearly 30 years' experience with SAS software. When not writing *about* SAS, Frank writes *in* SAS, primarily data management and reporting applications in the pharmaceutical industry.

Steven J. First

Steven J. First, President and Owner of Systems Seminar Consultants (SSC), is one of the country's foremost experts in SAS applications and support. He is SAS Certified Professional V8.

Since starting the company by himself in 1981, Steve has developed SSC into a full-service provider of SAS training that has been attended by over 12,000 students. One of the early pioneers to recognize the power of SAS software, Steve has a commitment to quality and customer service. Steve has established a reputation for making his clients' work easier, not more complicated. His philosophy has always been to look for the simple things that make life easier.

His experience with SAS since 1976 includes manufacturing, government, university, and financial applications. Steve comes from a traditional data processing background and has worked as an applications programmer, systems programmer, systems trainer and SAS consultant.

Steve is past president and founder of WISAS (Wisconsin Illinois SAS User's Group) and has been invited to speak at SUGI (SAS Users Group International) meetings on many occasions.

Malachy J. Foley

Malachy J. Foley is a Senior SAS Programmer/Analyst in the Department of Biostatistics at the University of North Carolina at Chapel Hill (UNC). Mal has programmed business and research applications for more than 35 years. His career in computing includes being an international consultant, a department manager, a university professor, and a CEO. He teaches computing courses at the undergraduate, graduate, and professional levels. He frequently presents papers and gives seminars at local, regional, national, and international SAS users' groups.