



SESUG Speaker Sharing Program

Presentation Titles, Abstracts and Author Bios

Program Goals:

- Speaker sharing is one of the many ways SESUG and SAS support local SAS user groups.
- SESUG and SAS will co-sponsor and provide SAS user groups in the South East a yearly presentation selected from our list of topics and delivered by our SESUG EC featured speakers. Speaker Travel Costs are fully paid by SESUG and SAS.

Program Specifics:

- Local and in-house SAS user groups in the SESUG region (Alabama, Florida, Georgia, Kentucky, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia, Washington DC, Puerto Rico) are eligible to participate. User groups in the former NESUG region are also eligible to apply.
- Sponsorship is limited to one speaker each calendar year. Speakers **do** facilitate requests to provide more than one presentation at a User Group Meeting.
- SESUG fulfills local SAS user group requests for speakers based upon speaker availability and reasonable travel costs.
- In support of the SESUG Speaker Sharing Program, SAS provides matching funds to assist SESUG in maximizing the effectiveness and coverage of the Speaker Sharing Program.

Requesting a Speaker:

- Please plan as far in advance as possible so that SESUG can accommodate your request.
- After reviewing the available topics, speakers, and speaker bios **BELOW**, select your “top three choices” of speakers, and email your request to [Darryl Putnam](#). Please include:
 - Desired **date / time / location** of meeting
 - User Group name and contact information
 - **Three** preferred speakers, and the presentation titles of interest.
- SESUG will review your request and, if fulfilled, SESUG will provide you with confirmation of:
 - Speaker / title(s)
 - Dates
 - Travel arrangements.
- NOTE: Requests are fulfilled based on speaker availability and reasonable travel expenses.



SESUG Speaker Sharing Program

Alphabetic List of Presentation Titles

[A Beginner's Babblefish: Basic Skills for Translation Between R and SAS®](#); Sarah Woodruff

[A Macro to Unravel Macros](#); Sarah Woodruff

[An Introduction to SAS® Arrays](#); Andrew T. Kuligowski

[Array, Hurray, Array; Consolidate or Expand Your Input Data Stream Using Arrays](#); William E. Benjamin Jr.

[Can you read this into SAS® for me? "Using INFILE and INPUT to Load Data"](#); Peter Eberhardt

[Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study](#); Abbas S Tavakoli, DrPH, MPH, ME

[Creating Zip Code Level Maps With SAS®](#); Barbara B. Okerson Ph.D.

[Data in the Doughnut Hole: Using SAS® to Report on What is NOT There](#); Sarah Woodruff

[Data Management Plans: Stages, Components, and Activities](#); Abbas S Tavakoli, DrPH, MPH, ME

[Easier than You Think: Creating Maps with SAS® Enterprise Guide®](#); Stephanie Thompson

[Easy Ways to get SAS® Enterprise Guide® to Send Tables and Graphs to Microsoft Office Suite, Without Writing Code](#); William E. Benjamin Jr.

[Evaluating Sample Code for an Interview](#); Stephanie Thompson

[Examining Mediator and indirect effects of Loneliness in Social Support on Social Well-Being Using Baron and Kenny and a Bootstrapping Method](#)
[Data Management Plans: Stages, Components, and Activities](#)
[Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study](#); Abbas S Tavakoli, DrPH, MPH, ME

[Extend the Power of SAS® to Use Callable VBS and VBA Code Files Stored in External Libraries to Control Excel Formatting Routines](#); William E. Benjamin Jr.

[Flat Pack Data: Converting and ZIPping SAS® Data for Delivery](#); Sarah Woodruff

[Four Thousand Reports Three Ways](#); Stephanie Thompson

[FREQ Out – Exploring Your Data the Old School Way](#); Stephanie Thompson

[Generating Custom Shape Files for Data Visualization - note this is a Quick Tip and can be done with another choice](#); Stephanie Thompson

[Getting the Most Out of Your SAS Conference](#); Andrew T. Kuligowski

[Getting Your Data into SAS®](#); Stephanie Thompson



SESUG Speaker Sharing Program

[Give the Power of SAS® to Excel Users Without Making Them Write SAS Code as a “Take-out”](#); William E. Benjamin Jr.

[I Object: SAS® Does Objects with DS2](#); Peter Eberhardt

[I’m a SAS® Programmer. Why Should I JMP®?](#); Barbara B. Okerson Ph.D.

[In Search of the LOST CARD](#); Andrew T. Kuligowski

[Interactions between the DATA Step and External Files – INFILE / INPUT and more](#); Andrew T. Kuligowski

[Intro to Proc Report with PDF Output](#); Barbara B. Okerson Ph.D.

[ISO 101: A SAS® Guide to International Dating](#); Peter Eberhardt

[Introduction to Data Simulation](#); Dr. Jason Brinkley

[Introduction to Hierarchical Linear Modeling \(HLM\)](#); Dr. Jason Brinkley

[Keeping Up Appearances: Turning Specifications into SAS® Format Libraries and Statements](#); Sarah Woodruff

[Leave Your Bad Code Behind: 50 Ways to Make Your SAS Code Execute More Efficiently](#); William E. Benjamin Jr.

[Looking Beneath the Surface of Sorting](#); Andrew T. Kuligowski

[Managing Your SAS® Assets](#); Stephanie Thompson

[Messy Data in SAS](#); Dr. Jason Brinkley

[More than Models: The Data Mining Team](#); Stephanie Thompson

[New to SAS® and New to Programming? What You Need to Do Before Typing Code](#); Stephanie Thompson

[Parsing Useful Data Out of Unusual Formats Using SAS®](#); Andrew T. Kuligowski

[Perusing, Choosing, and Not Mis-using: Non-parametric vs Parametric Tests in SAS \(R\)](#); Venita DePuy

[PROC SURVEY... Says!: Selecting and Analyzing Stratified Samples](#); Darryl Putnam

[Pruning the SASLOG – Digging into the Roots of NOTES, WARNINGS, and ERRORS®](#); Andrew T. Kuligowski

[Regression Models for Count Data](#); Dr. Jason Brinkley

[SAS Users Can Command Microsoft Excel to Automatically Create Graphs from SAS ExcelXP Output](#); William E. Benjamin Jr.

[Smoke and Mirrors!!! Come See How the _INFILE_ Automatic Variable and SHAREBUFFERS Infile Option Can Speed Up Your Flat File Text-Processing Throughput Speed](#); William E. Benjamin Jr.



SESUG Speaker Sharing Program

So, You've Got Data Enterprise Wide (SAS, ACCESS, EXCEL, MySQL, Oracle, and Others); Well, Let SAS Enterprise Guide Software Point-n-Click Your Way to Using; William E. Benjamin Jr.

Solving the 1,001-Piece Puzzle in 10 (or Fewer) Easy Steps: Using SAS@9 .cfg, autoexec.sas, SAS Registry, and Options to Set Up Base SAS@; Peter Eberhardt

Talking Past Each Other? How to Communicate with Medical Writers When Preparing Clinical Research Manuscripts for Journal Submission; Stephanie Thompson

The Baker Street Irregulars Investigate: Perl Regular Expressions and CDISC; Peter Eberhardt

The Building Blocks of SAS@ Datasets – S-M-U (Set, Merge, and Update); Andrew T. Kuligowski

The Little Engine That Could: Using LIBNAME Engine Options to Enhance Data Transfers Between SAS@ and Microsoft Excel Files; William E. Benjamin Jr.

The Quick and Dirty Excel without ODS or DDE; Andrea Zimmerman

Time Series Mapping with SAS@: Visualizing Geographic Change over Time in the Health Insurance Industry; Barbara B. Ph.D. Okerson

Tips and Tricks for Introductory Workshops in SAS@ for Health Professionals; Dr. Jason Brinkley

Tricks and Tips for Using the Bootstrap in JMP Pro; Dr. Jason Brinkley

Using Big Data to Visualize People Movement Using SAS@ Basics; Stephanie Thompson

Using GENMOD to Analyze Correlated Data on Military Health System Beneficiaries Receiving Inpatient Behavioral Health Care in South Carolina Health Care Systems; Abbas, S Tavakoli, DrPH, MPH, ME

Using GLIMMIX and GENMOD Procedures to Analyze Longitudinal Data from a Department of Veterans Affairs Multisite Randomized Controlled Trial; Abbas, S Tavakoli, DrPH, MPH, ME

Using INFILE and INPUT Statements to Introduce External Data into the SAS@ System; Andrew T. Kuligowski

Using JMP@ to apply Decision Trees and Random Forests as Screening Tools for Limiting Candidate Predictors in Regression Models; Dr. Jason Brinkley

Using SAS to Examine Mediator, Direct and Indirect Effects of Isolation and Fear on Social Support Using Baron& Kenny Combined with Bootstrapping Methods; Abbas, S Tavakoli, DrPH, MPH, ME

Where Should I Dig? What to do Before Mining Your Data; Stephanie Thompson

While You Were Sleeping, SAS@ Was Hard At Work; Andrea Zimmerman

Why the Bell Tolls 108 times? Stepping Through Time with SAS; Peter Eberhardt

Windows and Unix Computers Now Have Multiple CPU's; Why Not Control Two or Three or More Parallel Executing SAS Batch Jobs from One Master Job!; William E. Benjamin Jr.

Working with Character Data; Andrew T. Kuligowski



SESUG Speaker Sharing Program

Author Bios

William E. Benjamin Jr.

William E. Benjamin, Jr. his expertise includes Base SAS® Software, and SAS Macros. William has a BS degree in computer science from Arizona State University and an MBA from Western International University. He has been a SAS software user since 1983 and a computer programmer since 1973. His programming experience spans from vacuum tube mainframes, to current PC computers. William currently owns a consulting company called OWL Computer Consultancy, LLC in Phoenix AZ. His new SAS Press book titled "Exchanging Data between SAS and Microsoft Excel: Tips and Techniques to Transfer and Manage Data More Efficiently" is now available.

Dr. Jason Brinkley

Dr. Jason Brinkley is a Senior Researcher at the American Institutes for Research health division out of Chapel Hill, NC. He maintains an Adjunct Professor of Public Health and is an affiliate of the NC Agromedicine Institute at East Carolina University where he advises on data collected on agriculture worker health (including migrant populations). Brinkley spent almost 8 years on the faculty of the department of Biostatistics at East Carolina and has worked with a variety of individuals in the medicine, public health, business, education, and the natural and social sciences. He has expertise in a wide range of statistical and biostatistical methods along with basic psychometric methods and customized data visualizations. He currently serves as NC Chapter Representative for the American Statistical Association and on the executive board for the Southeast SAS Users Group. In his role at AIR, Dr. Brinkley leads and participates in research on a variety of health services research projects including evaluation, quality improvement, access, quality of care, performance management, and health care disparities.

Peter Eberhardt

Peter is a long time SAS consultant and his company, Fernwood Consulting Group, is a SAS Alliance Partner. When Peter is not participating in SAS Conferences and User Groups across Canada, the United States, the Caribbean, and China he enjoys riding his motorcycle, wilderness camping, and hiking. There are also occasions when he has been known to work.

Venita DePuy

Venita is an experienced statistician and SAS programmer, and owner of Bowden Analytics. She specializes in providing statistical and programming services to pharmaceutical companies, CROs, and academic organizations, with an emphasis on clinical trials and CDISC. Her therapeutic experience includes cardiology, oncology, depression, veterinary medicine. She earned her Bachelor of Science in Statistics from the University of Alaska Fairbanks, then continued on to earn her Master of Statistics at North Carolina State University and PhD in Biomathematics with a Statistics minor at NCSU as well.

Andrew T. Kuligowski

Andrew T. Kuligowski has been a SAS user for well over 25 years, and is currently a Manager for HSN in St Petersburg, Florida. He has augmented his professional coding experiences in the retail, media, insurance, and petrochemical fields by speaking at various SAS conferences and user group meetings. Andrew was conference chair for SAS Global Forum 2012 in Orlando, Florida, and co-chair for SESUG'97 in Jacksonville, Florida and Tennessee SAS Users Day in Knoxville, Tennessee. In his spare time, Andrew can also be found volunteering at the Florida Aquarium in Tampa.

Barbara B. Okerson

Barbara B. Okerson, Ph.D. is a Health Information Manager at Anthem where she supports the Client Reporting and Analytics team, especially ad hoc requests with SAS software. She has worked in the healthcare industry for over 20 years and has been a SAS user for over 25 years. She serves in leadership roles at SESUG and the



SESUG Speaker Sharing Program

Anthem In-house SAS User Group. She is a SAS Certified Professional, a Certified Professional in Healthcare Quality, and a Fellow, Academy for Healthcare Management.

Darryl Putnam

Darryl Putnam has 15 years of experience designing and implementing statistical and data analyses systems to solve business problems in a wide variety of industries. Mr. Putnam has specialized in designing and implementing the SAS software solution suite to build information decision support systems for his clients. These solutions range from statistical quality control systems, geo-demographic mapping systems, behavioral modeling systems, and enterprise wide production systems. He has a B.S. and M.A. in Economics from Virginia Commonwealth University. He is always looking for the next problem to solve.

Abbas Tavakoli

Dr. Tavakoli received a Bachelor of Science in Animal Husbandry from Tehran University, Tehran, Iran 1985, a Master of Public Health in Biostatistics at the University of South Carolina in 1989, a Doctor of Public Health in Biostatistics at the University of South Carolina in 1998, and a Master of Engineering in Computer Engineering at the University of South Carolina in 2003. Dr. Tavakoli teaches statistics courses for USC College of Nursing DNP and PhD students. Dr. Tavakoli has been involved with many research projects and manuscripts during his tenure at the College. His research interests are health related topics and he has considerable experience in data management, linear modeling, logistic regression, and experimental design.

Stephanie Thompson

Over twenty years of experience in applying statistical and modeling techniques to solve business problems in various commercial and academic environments. Strong understanding of data structures, a variety of analytical tools, and operating environments. Views problems (opportunities?) in a broad context by examining the interrelations between issues and the local and broader operational framework. Demonstrated skill at effectively communicating and working across multiple functional areas and at all organizational levels. Stephanie has made dozens of presentations at local, regional, and international meetings and conferences to technical and non-technical audiences.

Andrea Wainwright-Zimmerman

Andrea Wainwright-Zimmerman has been writing computer programs since the 2nd grade and has been programming in SAS for over 15 years. She graduated from Sam Houston State University with a BS in Mathematics and an MS in Statistics. She has been working for Capital One for over 12 years now. In her spare time she sings lead in a female barbershop chorus and is an animal lover and trainer working with 4 cats, 1 dog, 3 horses, and one husband.

Sarah Woodruff

Sarah Woodruff has been programming in SAS professionally since 2006. She works on reporting and analysis for Westat's Health, Education and Transportation divisions. She served as the Operations Chair for SESUG 2012 and has served on the SESUG executive council as well as section chaired for several years. Her undergraduate work includes a BS in mathematics and statistics from Georgia State University and a BS in microbiology from University of Maryland. Sarah also holds an MS in bioinformatics.



SESUG Speaker Sharing Program

Presentation Abstracts (in alphabetic order by presentation title)

A Beginner's Bafflefish: Basic Skills for Translation Between R and SAS®

Sarah Woodruff

SAS professionals invest time and energy in improving their fluency with the broad range of capabilities SAS software has to offer. However, the computer programming field is not limited to SAS alone and it behooves the professional to be well rounded in his or her skill sets. One of the most interesting contenders in the field of analytics is the open source R software. Due to its range of applications and the fact that it is free, more organizations are considering how to incorporate it into their operations and many people are already seeing its use incorporated into project requirements. As such, it is now common to need to move code between R and SAS, a process which is not inherently seamless.

This paper serves as a basic tutorial on some of the most critical functions in R and shows their parallel in SAS to aid in the translation process between the two software packages. A brief history of R is covered followed by information on the basic structure and syntax of the language. This is followed by the foundational skill involved in importing data and establishing R data sets. Next, some common reporting and graphing strategies are explored with additional coverage on creating data sets that can be saved, as well as how to export files in various formats. By having the R and SAS code together in the same place, this tutorial serves as a reference that a beginner can follow to gain confidence and familiarity when moving between the two.

A Macro to Unravel Macro

Sarah Woodruff

Do you find it tedious to pinpoint problems in your macro code? Are you annoyed when the log points to the line number of the macro call rather than the code in the macro that generated the error? Would you like a little help? You want to produce macros that are utilitarian black boxes, but you need to understand what the code is actually doing as part of the development process, and that can be frustrating.

This paper presents a simple wrapper macro that is used to enclose existing code, which can call any number of other macros. When you use this technique along with the mPrint and mFile options, the output file produced by mPrint/mFile contains a completely "de-macroed" version of the program. You can then examine this output as a working program, thus making it easier to both understand and edit.

An Introduction to SAS® Arrays

Andrew T. Kuligowski

So, you've heard about SAS® arrays, but are not sure when – or why – you would use them. This presentation will provide the attendee / reader with a background in SAS arrays, from an explanation as to what occurs during compile time through to their programmatic use, and will include a discussion regarding how DO-loops and macro variables can enhance array usability. Specific examples, including Fahrenheit to Celsius temperature conversion, salary adjustments, and data transposition / counting will assist the user with effective use of SAS arrays in their own work, and provide a few caveats as to their usage, as well.



SESUG Speaker Sharing Program

Array, Hurray, Array; Consolidate or Expand Your Input Data Stream Using Arrays

William E. Benjamin Jr.

You have an input file with one record per month, but need an output file with one record per year. But you cannot use PROC TRANSPOSE because other fields need to be retained or the input file is sparsely populated. The techniques shown in this paper will enable you to be able to either consolidate or expand your output stream of data by using arrays. Sorted files of data records can be processed as a unit using "BY Variable" groups and building an array of records to process. This technique allows access to all of the data records for a "BY Variable" group and gives the programmer access to the first, last and all records in between at the same time. This will allow the selection of any data value for the final output record.

Can you read this into SAS® for me? “Using INFILE and INPUT to Load Data

Peter Eberhardt

With all the talk of “Big Data” and “Visual Analytics” we sometimes forget how important, and often hard, it is to get external data into SAS. In this paper we will review some common data sources such as delimited sources (e.g. CSV) as well as structured flat files and the programming steps needed to successfully load these files into SAS. In addition to examining the INFILE and INPUT statements, we will look at some methods for dealing with bad data. This paper assumes only basic SAS skills, although the topic can be of interest to anyone who needs to read external files.

Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study

Abbas S Tavakoli, DrPH, MPH, ME

The importance of choosing the type of statistical program that we use to analyze data in a longitudinal study are growing as a specialty due to the fact that computerized data analyses have become the standard for scientific research. There are many procedures in SAS ® that can be used to analyze longitudinal data. The purpose of this paper is to compare MIXED and GLIMMIX procedure in SAS to analyze data from a longitudinal study. A randomized trial design was used in which 185 participants were assigned to a therapeutic group by teleconference (n=92) in which participants interacted in real time with each other or the control group (n=93) who received usual psychosocial care (any support used by the patient in the course of cancer treatment). The randomization was stratified by treatment type. Data were collected at baseline, the end of the intervention, and 16 weeks from baseline. A mixed-effects repeated measures model was used to assess outcome variable of social well being (social connection) by group over time. The effect of group, time and interaction effect of group by time were examined after controlling for several confounding factors. SAS is the most powerful statistical program for data analyses in a longitudinal study.

Creating Zip Code Level Maps With SAS®

Barbara B. Okerson, Ph.D.

SAS®, SAS/GRAPH®, and ODS graphics provide SAS programmers with the tools to create professional and colorful maps. Provided with SAS/GRAPH are boundary files for U.S. states and territories, as well as internal boundaries at the county level. While much data and results can be displayed at this level, often a higher degree of granularity is needed. The U.S. Census Bureau provides ZIP code boundary files in ESRI shapefile format (.shp) by state for free download and import into SAS using SAS PROC DATAIMPORT. This paper illustrates the use of these ZIP code tabulation area (ZCTA) files with SAS to map data at a ZIP code level. Example maps include choropleth, distance, and heat maps.



SESUG Speaker Sharing Program

Examples included in this paper were created with version 9.2 of SAS on a Windows 64-bit server platform and use Base SAS, SAS/STAT and SAS/GRAPH. The techniques represented in this paper require a SAS/GRAPH license but are not platform-specific and can be adapted by beginning through advanced SAS users.

Data in the Doughnut Hole: Using SAS® to Report on What is NOT There

Sarah Woodruff

Typically the request for a table, graph or other report concerns data that currently exists and needs to be explored. However, there is often a need to examine data that is expected to be present but currently is not. Yet how can we report on something that is not there? This paper describes and explores ways to create simulations of what is expected and then match those with data that is actually present. Making comparisons between what has been anticipated and what actually exists then opens up the ability to report on what is not there. This paper also explores how to test the differences between the groups based on a variety of conditions and how to determine what is legitimately absent. Reporting on the absent data is then described, including how to check that everything which is expected is actually present and be able to report that as well. DATA step techniques are combined with PROC SQL and PROC REPORT to create a step-wise process which can be easily modified to fit most any set of specifications.

Data Management Plans: Stages, Components, and Activities

Abbas S Tavakoli, DrPH, MPH, ME

Data management strategies have become increasingly important as new computer technologies allow for larger and more complex data sets to be analyzed easily. As a consequence, data management has become a specialty requiring specific skills and knowledge. Many new investigators have no formal training in management of data sets. This paper describes common basic strategies critical to the management of data as applied to a data set from a longitudinal study. The stages of data management are identified. Moreover, key components and strategies, at each stage are described.

Easier than You Think: Creating Maps with SAS® Enterprise Guide®

Stephanie Thompson

Have you ever wanted to display data on a map to add more punch to your analysis? Sometimes seeing things geographically can put things into a completely new perspective. Letting your data tell a story on a map is easier than you might think. A two-step approach with SAS® Enterprise Guide® is all you need. From using supplied shape files right through customizing your graph, this paper walks you through the process. Some best practices are also included to make sure your customers are impressed. They will never know how easy it really was.

Easy Ways to get SAS® Enterprise Guide® to Send Tables and Graphs to Microsoft Office Suite, Without Writing Code

William E. Benjamin Jr.

This paper will demonstrate several ways to output tables, graphs, databases, and files that are in formats that Microsoft products are able to process. All of these methods require no programming, and just use little known or hidden techniques that are available directly from the SAS® Enterprise Guide user interface. The focus will be



SESUG Speaker Sharing Program

directly on how a new user to SAS® Enterprise Guide can prepare and deliver data to the Microsoft Office (MS Office) suite of tools.

Evaluating Sample Code for an Interview

Stephanie Thompson

Requesting sample SAS® code from job candidates before an interview is a good way to gauge their level of experience, ability, and style. Once you have reviewed the code, prepare your questions for the candidate. You want to make sure that they 1) wrote the code themselves, 2) understand what they wrote, and 3) are familiar enough with it to explain why things were done the way they were. This paper will provide you with some guidelines on how to get answers to these questions in an interview. Code samples that I have received over the last few years and the questions that arose will be used as examples. This is also a good time to ask if the candidate can think of another way to accomplish the same task using a different approach. Evaluating work samples can be a way to gain invaluable insight into your candidate's skills and this paper will help you get the most out of it that you can.

Examining Mediator and indirect effects of Loneliness in Social Support on Social Well-Being Using Baron and Kenny and a Bootstrapping Method Data Management Plans: Stages, Components, and Activities Compare MIXED and GLMMIX to Analyze Breast Cancer Longitudinal Study

Abbas S Tavakoli, DrPH, MPH, ME

This study presentation examines mediator effect and indirect effect of loneliness in social support on social well being by using two methods: Baron & Kenny, and Bootstrapping. A cross-sectional data were used here from the longitudinal study randomized trial design in which 185 participants were assigned to the therapeutic group (n=93) who received by teleconference with participants interacting in real time with each other and control group (n=92) who received usual psychosocial care (any support used by the patient in the course of cancer treatment. Baron and Kenny (1986) steps and Hayes (2004) we reused to examine mediation effect. Results of Baron indicated that the relationship between social support and social well being was significant ($c = .634$ (total effect) ($p = .0001$)) and that there was significant relationship between mediator and predictor variable ($\alpha = -2.11$ ($p = .0001$)). Also, previously found significant relationship between social support and social well being ($c' = .595$ (direct effect) ($p = .0001$)) when both social support and loneliness were in the model. The indirect effect was $.039$ and Sobel test was significant ($P = .028$). Therefore, there was no mediator effect for loneliness in the relationship between social support and social well being. The results of bootstrapping methods indicated the direct effect was $.591$ (95% CI: $.589-.593$ for normal theory and $.481-.690$ for percentile) and indirect effect was $.040$ (95% CI: $.039-.040$ for normal theory and $.006-.087$ for percentile). The result showed both methods had significant indirect effect.

Extend the Power of SAS® to Use Callable VBS and VBA Code Files Stored in External Libraries to Control Excel Formatting Routines

William E. Benjamin Jr.

Did you ever wish you could use the power of SAS® to take control of EXCEL and make EXCEL do what you wanted WHEN YOU WANTED? Well one letter is the key to doing just that, the letter X as in the SAS X Command that opens the door to all operating system commands from SAS. The Windows operating system comes with a facility to write a series of commands called scripts. These scripts have the ability to open and reach into the internals of EXCEL. Scripts can load, execute and remove VBA macro code and control EXCEL. This



SESUG Speaker Sharing Program

level of control allows you to make EXCEL do what you want, without leaving any traces of a macro behind. This is Power.

Flat Pack Data: Converting and ZIPping SAS® Data for Delivery

Sarah Woodruff

Clients or collaborators often need SAS data converted to a different format. Delivery or even storage of individual data sets can become cumbersome, especially as the number of observations or variables grows. The process of converting SAS data sets into other forms of data and saving files into compressed ZIP storage has become not only more efficient, but easier to integrate into new or existing programs. This paper describes and explores various methods to convert SAS data sets to other formats as well as effective strategies to ZIP data sets along with any other files that might need to accompany them.

PROC IMPORT and PROC EXPORT have been long standing components of the SAS toolbox, so much so that they them in code being run in batch or to include them in programs that may be run interactively but “hands free”. The syntax of each is described with a particular focus on moving between SAS, STATA and SPSS, though some attention is also given to Excel. Once data sets and their attendant files are ready for delivery or need to be put into storage, compressing them into ZIP files becomes helpful. The process of using ODS PACKAGE to create such ZIP files is laid out and can be connected programmatically to the creation of the data sets or documents in the first place.

Four Thousand Reports Three Ways

Stephanie Thompson

How do you go about generating over four thousand PDF reports in up to three different versions? When a large, southern research university decided to add up to five optional questions per class section and up to five more questions at the prefix level to their core set of fifteen questions on the student evaluation of faculty survey, it seemed like a project that would never be completed. If the additional questions weren't enough, the reports were being revamped at the same time to improve their appearance for delivery on the web. Each report had a tabular section and two customized box and whisker plots. Thanks to ODBC / SAS Access, PROC SQL, macro, DATA Step programming, PROC GPLOT, goptions, and ODS it all came together. This paper summarizes how each SAS® component was used and contributed to the completion of the project.

FREQ Out – Exploring Your Data the Old School Way

Stephanie Thompson

The tried-and-true FREQ procedure just doesn't get the attention it deserves. But, as they say, it is an oldie but a goodie. Sometimes you just need a quick look at your data and a few simple statistics. PROC FREQ is a great way to get an overview of your data with a limited amount of code. This hands-on workshop explores everything from the basic framework of the procedure to how to customize the output. It also presents an overview of some of the options that are available.



SESUG Speaker Sharing Program

Generating Custom Shape Files for Data Visualization - note this is a Quick Tip and can be done with another choice

Stephanie Thompson

There are times you may want to represent your data graphically but the existing options do not work. Using a custom shape file or even a file with coordinates you create yourself can get you what you want. This quick presentation will show you how.

Getting the Most Out of Your SAS Conference

Andrew T. Kuligowski

Your company has approved your attendance at the conference. They've agreed to an outlay of money to cover your registration, transportation, food, and lodging. Equally important, they've made a commitment of your time, allowing you to be at the event instead of sitting at your computer terminal. They fully anticipate that you will return their investment through the knowledge you bring back. How do you ensure that you make the best use of your time – to get the most out of your conference experience?

I've been a speaker at many SAS User Group conferences and events over the years. In that time, I think I've learned some tricks to best utilize my time at user group events. This presentation will be a short, informal look into how I make the best use of my time; hopefully, it will assist you in getting the most out of your conference.

Getting Your Data into SAS®

Stephanie Thompson

Where is your data stored?

- Oracle tables
- SQL Server tables
- Microsoft Access
- Microsoft Excel
- Text file
- All over the place

Over the years there has been a proliferation of ERP systems and other ways to collect and store data. Many times you need data from different systems to complete a single analysis. Sometimes getting data out can seem like quite a challenge. Luckily, SAS has the capability to access many different types of data but also different ways to do so. This workshop will cover how to use SAS to access data from a variety of sources through both presentation and live demonstration. Some practical tips on which methods work best or are fastest will also be covered.

The following methods to access data will be covered in the workshop:

- SAS libname engine
- PROC SQL
- ODBC
- PROC IMPORT (including the Import Wizard)



SESUG Speaker Sharing Program

Give the Power of SAS® to Excel Users Without Making Them Write SAS Code as a Take-out™

William E. Benjamin Jr.

Merging the ability to use SAS® and Microsoft Excel can be challenging. However, with the advent of SAS® Enterprise Guide®, SAS® Integration Technologies, SAS® BI Server software, JMP® software, and SAS® Add-In for Microsoft Office; this process is less cumbersome. Excel has the advantages of being cheap, available, easy to learn, and flexible. On the surface, SAS and Excel seem widely separated without these additional SAS products. But wait, BOTH SAS AND EXCEL CAN INTERFACE WITH THE OPERATING SYSTEM. SAS can run Excel using the command and Excel can run SAS as an "APPLICATION." This is NOT DDE; each system works independently of the other. This paper gives an example of Excel controlling a SAS process and returning data to Excel.

I Object: SAS® Does Objects with DS2

Peter Eberhardt

The DATA step has served SAS® programmers well over the years, and although it is powerful, it has not fundamentally changed. With DS2, SAS has introduced a significant alternative to the DATA step by introducing an object-oriented programming environment. In this paper, we share our experiences with getting started with DS2 and learning to use it to access, manage, and share data in a scalable, threaded, and standards-based way.

I'm a SAS® Programmer. Why Should I JMP®?

Barbara B. Okerson, Ph.D.

JMP® software, when used as a stand-alone package, provides a variety of ways of understanding, visualizing and communicating what your data is telling you. Adding JMP software and functionality to a SAS® programming environment can result in the best of both worlds. Topics covered include previewing SAS data, running SAS procedures from JMP, using JMP for further exploration of SAS results, and using SAS geographic data with JMP. Through these features and others in JMP's point-and-click environment, JMP extends the power and functionality of SAS.

Examples used in this paper were created with JMP 8 and SAS 9.2 that coexist on a Windows 2008 R2 Enterprise Server. The techniques represented are not platform-specific and can be adapted by beginning through advanced SAS and JMP users.

In Search of the LOST CARD

Andrew T. Kuligowski

"Everyone who's not here, raise your hand." It's an old joke, but it points out the difficulty of identifying persons or things that are not present. The SAS® System has its own version of this chestnut, the SASLOG message indicating that there are one or more gaps in one's input data:

NOTE: LOST CARD.

This presentation will focus on the creation and use of ad hocs to explore input data, in order to locate the positions where the input data might be incomplete. The goal will be to identify where the missing data should be, so that you can code around the limitations of your data.



SESUG Speaker Sharing Program

Interactions between the DATA Step and External Files – INFILE / INPUT and more

Andrew T. Kuligowski.

Chances are, your raw data was not created within the SAS® System. There is a good likelihood that your data may also need to be packaged and passed along to another non-SAS package.

This presentation will provide basic answers to two questions common to new SAS users:

- How do I get my data into SAS for analysis?
- How do I get my data out of SAS?

The focus for this presentation will be on two pairs of DATA step statements: INFILE / INPUT and FILE / PUT. We will discuss syntax and usage, citing various types of files as examples.

Introduction to Data Simulation

Dr. Jason Brinkley

Creating synthetic data via simulation can often be a powerful tool for a wide variety of analyses. The purpose of this workshop is to provide a basic overview of simulating data for a variety of purposes. Examples will include power calculations, sensitivity analysis, and exploring nonstandard analyses. The workshop is designed for the mid-level analyst who has basic knowledge of data management, visualizations and basic statistical analyses such as correlations and t-tests.

Introduction to Hierarchical Linear Modeling (HLM)

Dr. Jason Brinkley

HLM, also known as multi-level analysis, is a more advanced form of simple linear regression and multiple linear regression. Multilevel analysis allows variance in outcome variables to be analysed at multiple hierarchical levels, whereas in simple linear and multiple linear regression all effects are modeled to occur at a single level. Thus, HLM is appropriate for use with nested data.

Intro to Proc Report with PDF Output

Barbara B. Okerson, Ph.D.

Proc Report is a SAS tool that combines and enhances the features of the print, means, and tabulate procedures to summarize and display data. The flexibility of the report procedure allows the design of reports as per requested specifications with every aspect fully customizable. Examples create client-ready reports in PDF format without post-processing. Customizations illustrated include: splitting column labels, creating spanning headers, grouping and ordering results, currency formatting, and cell-width modifications.

ISO 101: A SAS® Guide to International Dating

Peter Eberhardt

For most new SAS programmers SAS dates can be confusing. Once some of this confusion is cleared the programmer may then come across the ISO date formats in SAS, and another level of confusion sets in. This paper will review SAS date, SAS datetime, and SAS time variables and some of the ways they can be managed. It will then turn to the SAS ISO date formats and show how to make your dates international.



SESUG Speaker Sharing Program

Keeping Up Appearances: Turning Specifications into SAS® Format Libraries and Statements

Sarah Woodruff

Specifications documents concerning the desired appearance of SAS data are often provided in an Excel spreadsheet format. While such an arrangement provides ease of use to the person creating it, typically the client to whom the final data delivery will eventually be returned, having format information set up this way is not directly conducive to its use in SAS. This paper describes a process by which formats presented in this way can be easily converted into both SAS format libraries and format statements. The process centers around the use of standard reports from Oracle Clinical to provide the basic information, but the methods can be applied to any specifications document in Excel. Along with the final format products, this encompasses evaluating the appropriateness of format names and making broadly applicable changes as needed, ensuring unique variable names, particularly if data is coming from multiple sources, and building in appropriate quality control steps along the way. This conversion requires both DATA step work and several common procedures (IMPORT, CONTENTS, FREQ), but is accomplished entirely in Base SAS and does not need any special products. Flexibility remains in the process to compile or subdivide both the format catalog and statements as needed by the user or as applicable based on the requirements of a particular deliverable.

Leave Your Bad Code Behind: 50 Ways to Make Your SAS Code Execute More Efficiently

William E. Benjamin Jr.

This laundry list of tips shows 50 ways to help SAS programmers make their code run faster and more efficiently. Topics include maximizing each DATA step, making simple tasks take less code, using macro variables to simplify maintenance, using built-in features, optimizing code to save disk space, using sorts for more than just sorting, and ways to make the program code just read better.

Looking Beneath the Surface of Sorting

Andrew T. Kuligowski

Many things that appear to be simple turn out to be a mask for various complexities. For example, as we all learned early in school, a simple drop of pond water reveals a complete and complex ecosystem when viewed under a microscope. A single snowflake contains a delicate crystalline pattern. Similarly, the decision to use data in a sorted order can conceal an unexpectedly involved series of processing and decisions. This presentation will examine multiple facets of the process of sorting data, starting with the most basic use of PROC SORT and progressing into options that can be used to extend its flexibility. It will progress to look at some potential uses of sorted data, and contrast them with alternatives that do not require sorted data. For example, we will compare the use of the BY statement vs. the CLASS statement in certain PROCs, as well as investigate alternatives to the MERGE statement to combine multiple datasets together.

Managing Your SAS® Assets

Stephanie Thompson

Hiring talented SAS® professionals can be an expensive proposition. While it is great to get another position for your department, replacing someone can be a real downer. Why did they leave? Was it the work? Did we make



SESUG Speaker Sharing Program

a poor hiring decision? Was it just a better opportunity? Could they have been happy here? Maybe they just did not like me.

All of these questions are valid, and you can keep guessing and never really know. As they say, an ounce of prevention is worth a pound of cure. What can you do as a manager to keep your best talent happy and challenged? Keeping those in your department engaged is a big piece of the puzzle.

Not a manager? Well, you're not off the hook either. There are things you can do to help your manager manage you better. I've been on both sides of this fence, and this paper will talk about some of the things that will help keep everyone happy. Think of your company's SAS talent as SAS assets!

Messy Data in SAS

Dr. Jason Brinkley

Transitions between character and numeric data, working with dates and time, summarizing large data into smaller datasets by counts or groupings.

More than Models: The Data Mining Team

Stephanie Thompson

Gathering data from various sources, preparing it for modeling, imputing, partitioning, testing various models, choosing the best, presenting it to your boss, then failure? Data mining is about more than just variables and models. Developing an understanding of each variable may take more than just running some summary statistics and deciding if it is ordinal or nominal. The need for comprehension is even more critical when the data you use are from different areas of the organization. Putting together a group of subject matter experts in the early phases of a data mining project can make a big difference in the outcome of your project. They can help you eliminate extraneous or duplicative variables and put others in context to help you better understand and interpret results. This paper will discuss how subject matter experts can aid in data mining using examples from several actual projects. Learn how to leverage the knowledge to derive a better conclusion and avoid costly errors.

New to SAS® and New to Programming? What You Need to Do Before Typing Code

Stephanie Thompson

People from all kinds of backgrounds are using SAS® software to meet their business needs. Some come with previous programming experience and some have never typed a line of code in their career. Have you been told what each proc does, that each DATA step must end in a run command, and then been told to write a SAS program? Have you been given code to launch and told to "just change the date and run it every week"? Beware. There is a lot more to SAS programming than knowing some syntax and launching a job. Being a successful SAS programmer means bringing yourself up to speed with some basic programming skills. Before typing even the first line of code into the Enhanced Editor, some preparation needs to be done. Doing this will help you more than you can imagine. This paper provides the basics of programming that can help you write good code and ensure you are getting the answer you intended to get.



SESUG Speaker Sharing Program

Parsing Useful Data Out of Unusual Formats Using SAS®

Andrew T. Kuligowski

Most “Introduction to Programming” courses will include a section on reading external data; the first assumption they make will be that the data are stored in some sort of documented and consistent format. Fortunately, in the “real world”, a lot of the data we deal with has the same basic assumption of occurring in a documented, consistent format – a lot of it, but not all of it. This presentation will address some techniques that can be used when we are not dealing with cleanly formatted data, when the data we want is in a less-than-ideal format, perhaps intermingled or seemingly buried with unnecessary clutter. It will discuss the principles of using SAS® to parse a file to extract useful data from a normally unusable source. This will be accomplished by citing examples of unusual data sources and the SAS Code used to parse it.

Perusing, Choosing, and Not Mis-using: Non-parametric vs Parametric Tests in SAS (R)

Venita DePuy

Most commonly used statistical procedures, such as the ttest, are based on the assumption of normality. The field of non-parametric statistics provides equivalent procedures that do not require normality, but often require assumptions such as equal variances. Parametric tests (which assume normality) are often used on non-normal data; even non-parametric tests are used when their assumptions are violated. This paper will provide an overview of parametric tests and their non-parametric equivalents; what assumptions are required for each test; how to perform the tests in SAS®; and guidelines for when both sets of assumptions are violated.

Procedures covered will include PROCs ANOVA, CORR, NPAR1WAY, TTEST and UNIVARIATE. Discussion will include assumptions and assumption violations, robustness, and exact versus approximate tests.

PROC SURVEY... Says!: Selecting and Analyzing Stratified Samples

Darryl Putnam

Statisticians and analysts need to design stratified survey plans and analyze the results of those surveys. Gone are the days when the analyst can ignore survey design tools when drawing inferences from the surveys. By forgoing the SAS® survey analysis procedures, estimates of the mean and standard error will be incorrect. By combining DATA STEP processing with the SAS® survey analysis procedures of PROC SURVEYSELECT and PROC SURVEYMEANS, we can determine the sample size, allocate the sample size across strata, and then draw correct inferences. This paper will demonstrate how to use these survey design and analysis tools with a stratified sample of an inventory audit.

Pruning the SASLOG – Digging into the Roots of NOTEs, WARNINGs, and ERRORs®

Andrew T. Kuligowski

You've sat through constant design meetings. You've endured countless requests for "just one more little change". You even managed to find a creative solution to that nagging technical problem. But, you persevered, and despite all of the obstacles, you've managed to eliminate the final syntax error in your newest SAS® routine. Time to sit back and relax -- uh, not quite ...



SESUG Speaker Sharing Program

The primary focus of this presentation will be on techniques to ensure comprehension of your input data. We will look at several messages that are often found in the SASLOG, such as:

NOTE: MERGE statement has more than one data set with repeats of BY values.

that imply that there may be gaps in your knowledge of your data! Special emphasis will be placed on the use of ad-hoc queries to assist in finding data anomalies that can cause problems with your SAS code. It is assumed that the reader has a basic understanding of the SASLOG, including its composition, format, and the SAS system options which control its content.

Regression Models for Count Data

Dr. Jason Brinkley

Outcomes in the form of counts are becoming an increasingly popular metric in a wide variety of fields. For example, studying the number of hospital, emergency room, or in-patient doctor's office visits has been a major focal point for many recent health studies. Many investigators want to know the impact of many different variables on these counts and help describe ways in which interventions or therapies might bring those numbers down. Traditional least squares regression was the primary mechanism for studying this type of data for decades. However, alternative methods were developed some time ago that are far superior for dealing with this type of data. The focus of this workshop is to illustrate how count regression models can outperform traditional methods while utilizing the data in a more appropriate manner. Most of the first half of the workshop will focus on Poisson Regression techniques but some time will be devoted to using Negative Binomial Regression when the data are overdispersed and using Zero-Inflated techniques for data with many more zeroes than is expected under traditional count regression models. Time will be spent applying such models to example data so that the participant can see when and how to utilize these methods. A strong grasp of traditional linear regression methods is suggested.

SAS Users Can Command Microsoft Excel to Automatically Create Graphs from SAS ExcelXP Output

William E. Benjamin Jr.

The SAS ODS Tagset ExcelXP creates *.xml output, and *.xml output cannot contain graphs. So how can SAS programmers get graphs into your Excel workbooks? One way is to build them in Excel yourself. This paper shows you how to create data using SAS, and then command Microsoft Excel to read the data, create a graph or fully reformat a worksheet, without putting an Excel macro into the output Excel Workbook. And the program will do it all while you watch, including for multiple sheets in a workbook. The SAS code, and Excel code, shown is a fully integrated system to create and format macro-free Excel workbooks, using SAS® 9 and Excel 97 and above.

Smoke and Mirrors!!! Come See How the _INFILE_ Automatic Variable and SHAREBUFFERS Infile Option Can Speed Up Your Flat File Text-Processing Throughput Speed

William E. Benjamin Jr.

Processing speed for small files is never really a problem; one eye blink looks and feels much the same as two or three blinks of the eye. With today's computer speeds any machine will do a lot of work in three blinks of the eye.



SESUG Speaker Sharing Program

However, when processing hundreds of thousands, or millions of records blinks can become pauses, yawns, naps, or even an over-night sleepover. When a Base SAS® data step reads in text files and writes out nearly the same file the work proceeds as follows; (a) read input file data into input buffer, (b) move data values to data step variables (c) process data variables, (d) move data to output buffer, (e) write output file from output buffer. SAS reads your source code every time you run the code and builds the “current” processing steps (unless the code is compiled separately first) then executes them. By packing more work into each instruction less time is wasted between instructions, this means more work in less real time. The SHAREBUFFERS option eliminates moving data from the input to the output buffer (a BIG time saver). And the `_infile_` automatic variable can be used in function calls on either side of the “=” sign. This eliminates some or all of the data step variable processing like this; (a) read input file data into a common buffer, (b) process data values, (c) write data output text file from common buffer. The resulting time saved can result in faster execution of the program.

So, You've Got Data Enterprise Wide (SAS, ACCESS, EXCEL, MySQL, Oracle, and Others); Well, Let SAS Enterprise Guide Software Point-n-Click Your Way to Using William E. Benjamin Jr.

This paper will walk a SAS user through the process of opening and using files from different (non-SAS) generated sources and processing the data with SAS Enterprise Guide (EG). The focus of this paper will be to get access to the data, since most functions available to EG are drop down menus and wizards which help teach a new user how to process data. For completeness a simple chart will be shown with instructions about how to each type of data shown. This paper will not show how to get the data into non-SAS files; it will be assumed that the user's enterprise will guide them through that process.

Solving the 1,001-Piece Puzzle in 10 (or Fewer) Easy Steps: Using SAS®9 .cfg, autoexec.sas, SAS Registry, and Options to Set Up Base SAS® Peter Eberhardt

Are you frustrated with manually setting options to control your SAS® Display Manager sessions but become daunted every time you look at all the places you can set options and window layouts? In this paper, we look at various files SAS accesses when starting, what can (and cannot) go into them, and what takes precedence after all are executed. We also look at the SAS Registry and how to programmatically change settings. By the end of the paper, you will be comfortable in knowing where to make the changes that best fit your needs.

Talking Past Each Other? How to Communicate with Medical Writers When Preparing Clinical Research Manuscripts for Journal Submission

Stephanie Thompson

Clinical research manuscripts are often a blend of qualitative prose and quantitative data. And, the people who help the author(s) prepare each section — introduction, methods and materials, results, and discussion — are typically from very different parts of the research organization and have different backgrounds. Speaking a common language makes the process easier. This paper will present some tips on how to improve communication between medical writers and SAS® programmers, such as biostatisticians. Presented from both points of view, what does and doesn't work will be discussed in a point, counter-point fashion.



SESUG Speaker Sharing Program

The Baker Street Irregulars Investigate: Perl Regular Expressions and CDISC

Peter Eberhardt

A true detective needs the help of a small army of assistants to track down and apprehend the bad guys. Likewise, a good SAS® programmer will use a small army of functions to find and fix bad data. In this paper we will show how the small army of regular expressions in SAS can help you.

The paper will first explain how regular expressions work, then show how they can be used with CDSIC.

The Building Blocks of SAS® Datasets – S-M-U (Set, Merge, and Update)

Andrew T. Kuligowski

S-M-U. Some people will see these three letters and immediately think of the abbreviation for a private university and associated football team in Texas. Others might treat them as a three-letter word, and recall a whimsical cartoon character created by Al Capp many years ago. However, in the world of the SAS® user, these three letters represent the building blocks for processing SAS datasets through the SAS DATA step. S, M, and U are first letters in the words SET, MERGE, and UPDATE – the 3 commands used to introduce SAS data into a DATA step. This presentation will discuss the syntax for the SET, MERGE, and UPDATE commands. It will compare and contrast these 3 commands. Finally, it will provide appropriate uses for each command, along with basic examples that will illustrate the main points of the presentation.

The Quick and Dirty Excel Workbooks Without DDE or ODS

Andrea Wainwright-Zimmerman

There is a simple trick using the X command in SAS® that allows you to write out your SAS data to an already formatted Excel workbook with graphs and/or pivot tables already built. This presentation will describe how to accomplish this, as well as the limitations of this method.

The Little Engine That Could: Using LIBNAME Engine Options to Enhance Data Transfers Between SAS® and Microsoft Excel Files

William E. Benjamin Jr.

Many people are not aware that the SAS® Access for PC Files product will allow SAS Programmers to access an Excel spreadsheet in much the same way as any other SAS file. There are of course some restrictions, but there are also a lot of options that help remove some of the bumps in the road. The LIBNAME statement allows the user to define an Excel file in SAS terms and gives the programmer access to LIBNAME and data set options to control how the Excel file is defined, accessed, and yes even how the data will be formatted. This paper will describe some of those options.



SESUG Speaker Sharing Program

Time Series Mapping with SAS®: Visualizing Geographic Change over Time in the Health Insurance Industry

Barbara B. Okerson, Ph.D.

Changes in health insurance and other industries often have a spatial component. Maps can be used to convey this type of information to the user more quickly than tabular reports and other non-graphical formats. SAS® provides programmers and analysts with the tools to not only create professional and colorful maps, but also the ability to display spatial data on these maps in a meaningful manner that aids in the understanding of the changes that have transpired. This paper illustrates the creation of a number of different maps for displaying change over time with examples from the health insurance arena.

Tips and Tricks for Introductory Workshops in SAS® for Health Professionals

Dr. Jason Brinkley

It can sometimes be the case that general health professionals need some basic SAS training in order to effectively create simple reports and manipulate incoming data. The presenter will share his experiences in leading SAS Workshops Series in a university setting across the course of several years. Heading a team of university faculty members, the presenter has designed, implemented, and refined short term SAS overview training for general health professionals. While multiple topics have been discussed in these workshops, some have fared better with a general health professional audience than others. Topics will include tips on introducing code based work to individuals with no previous experience, workshop format, good practices on instruction and delivery, and introducing SAS macros in an example based manner.

Tricks and Tips for Using the Bootstrap in JMP Pro

Dr. Jason Brinkley

The bootstrap has become a very popular technique for assessing the variability of many different or unusual estimators. Starting in JMP Pro 10 the bootstrap feature was added to a wide variety of output options; however, there has not been much development as to the possible uses of this somewhat hidden feature. This paper will discuss a handful of uses that can be added to routine analyses. Examples include confidence interval estimates of the 5% trimmed mean, validation of covariates in regression analysis, comparing the differences in Spearman correlation estimates across two groups, and eigenvalues in principal components analysis. The examples will show the extra depth that can be easily added to routine analyses.

Using Big Data to Visualize People Movement Using SAS® Basics

Stephanie Thompson

Visualizing the movement of people over time in an animation can provide insights that tables and static graphs cannot. There are many options but what if you want to base the visualization on large amounts of data from several sources? SAS is a great tool for this type of project. This paper will summarize how visualizing movement was accomplished using several datasets, large and small, and the various SAS PROCs to pull it together. The use of a custom shape file will also be highlighted. The end result is a gif which can be shared that provides insights not available with other methods.



SESUG Speaker Sharing Program

Using INFILE and INPUT Statements to Introduce External Data into the SAS® System

Andrew T. Kuligowski

The SAS® System has numerous capabilities to store, analyze, report, and present data. However, those features are useless unless that data is stored in, or can be accessed by, the SAS System. This presentation is designed to review the INFILE and INPUT statements. It has been set up as a series of examples, each building on the other, rather than a mere recitation of the options as documented in the manual. These examples will include various data sources, including DATALINES, sequential files, and CSV files.

Using GENMOD to Analyze Correlated Data on Military Health System Beneficiaries Receiving Inpatient Behavioral Health Care in South Carolina Health Care System

Abbas S Tavakoli, DrPH, MPH, ME

Many SAS® procedures can be used to analyze large datasets with correlated data. This study was a secondary analysis of data obtained from the South Carolina Revenue and Fiscal Affairs Office (RFA), which includes medical claims from all health care systems in South Carolina (SC). We used the SAS procedure, GENMOD, to analyze a large dataset with correlated data from Military Health Care (MHS) system beneficiaries who received inpatient behavioral health care in South Carolina Health Care Systems from 2005 to 2014. Behavioral health (BH) was defined by Major Diagnostic Code (MDC) 19 (mental disorders and diseases) and 20 (alcohol and drug use). MDCs are formed by dividing all possible principal diagnoses from the International Classification Diagnostic (ICD-9) codes into 25 mutually exclusive diagnostic categories. The sample included a total of 6,783 BH visits and 4,827 unique military service members, veterans, and their adult and child dependents who have MHS insurance coverage. PROC GENMOD included a multivariate GEE model with type of BH hospitalization (mental health or substance abuse) as the dependent variable; and gender, race, age group, and discharge year as predictors. Hospital ID was used in the repeated statement with different correlation structures. Gender was significant for both independent correlation ($p = .0001$) and exchangeable structure ($p = .0003$). However, age group was significant using the independent correlation ($p = .0160$), but non-significant using the exchangeable correlation structure ($p = .0584$). SAS is a powerful statistical program for analyzing large, correlated datasets with categorical outcomes.

Using GLIMMIX and GENMOD Procedures to Analyze Longitudinal Data from a Department of Veterans Affairs Multisite Randomized Controlled Trial

Abbas S Tavakoli, DrPH, MPH, ME

Many SAS® procedures can be used to analyze longitudinal data. This study employed a multisite randomized controlled trial design to demonstrate the effectiveness of two SAS procedures, GLIMMIX and GENMOD, to analyze longitudinal data from five Department of Veteran Affairs Medical Centers (VAMCs). Older male veterans ($n = 1222$) seen in VAMC primary care clinics were randomly assigned to two behavioral health models, integrated ($n = 605$) and enhanced referral ($n = 617$). Data were collected at baseline, and 3, 6, and 12 month follow-up. A mixed-effects repeated measures model was used to examine the dependent variable, problem drinking, which was defined as count and dichotomous from baseline to 12 month follow-up. Sociodemographics and depressive symptoms were included as covariates. First, bivariate analyses included general linear model and chi-square tests to examine covariates by group and group by problem drinking outcomes. All significant covariates were included in the GLIMMIX and GENMOD models. Then, multivariate analysis included mixed models with Generalized Estimation Equations (GEEs). The effect of group, time, and the interaction effect of group by time were examined after controlling for covariates. Multivariate results were inconsistent for GLIMMIX and GENMOD using Lognormal, Gaussian, Weibull, and Gamma distributions. SAS is a powerful statistical program in data analyses for longitudinal study.



SESUG Speaker Sharing Program

Using JMP® to apply Decision Trees and Random Forests as Screening Tools for Limiting Candidate Predictors in Regression Models

Dr. Jason Brinkley

There are many techniques for evaluating candidate predictors in regression models. Some, such as stepwise regression, have been well studied and with known limitations. Others, such as shrinkage techniques (i.e. LASSO), have become increasingly popular and are showing true potential in helping to provide quality regression models for estimating effects in a multiple variable environment. Many of these techniques become difficult in the world of big data, especially if that data is long (many columns or variables) and short (fewer numbers of observations or rows). Tree based methods are a good alternative as a framework for data exploration and identification of variables to be used in building a quality regression model. Indeed tree methods provide an entirely different framework for model building that can oftentimes provide better predictions. However, when the main goal is still effect estimation, tree methods can be a very useful screening tool. This work examines the practice based on several examples using options very commonly found in both JMP® and JMP Pro software. The focus is on using EITHER single classification trees as well as so-called 'Random' or 'Bootstrap' Forests.

Using SAS to Examine Mediator, Direct and Indirect Effects of Isolation and Fear on Social Support Using Baron& Kenny Combined with Bootstrapping Methods

Abbas S Tavakoli, DrPH, MPH, ME

This study presentation examines mediator, direct and indirect effects of isolation and fear on social support by using two methods: Baron & Kenny, and Bootstrapping. This paper used a cross-sectional data from the longitudinal study randomized trial design in which 185 participants were assigned to the therapeutic group (n=93) who received by teleconference with participants interacting in real time with each other and control group (n=92) who received usual psychosocial care (any support used by the patient in the course of cancer treatment. Baron and Kenny (1986) steps and Hayes (2004) were used to examine for direct and indirect effects. Results of Baron indicated that the relationship between fear and social support was significant ($c = -1.151$ (total effect) ($p = .0001$)) and that there was significant relationship between isolation and fear ($\alpha = 1.22$ ($p = .0001$)). Also, previously significant relationship between fear and social support was not significant ($c' = -.40$ (direct effect) ($p = .1876$)) when both fear and isolation were in the model. The indirect effect was -1.11 and Sobel test was significant ($P = .0001$). The results of bootstrapping methods indicated the direct effect was $-.41$ (95% CI: $-.42, -.40$ for normal theory and $-.41$ (95% CI: $-.99, .14$ for percentile) and indirect effect was -1.06 (95% CI: $-1.09, -1.08$ for normal theory and $-1.09, -1.55$ for percentile). The result showed both methods had significant indirect effect.

Where Should I Dig? What to do Before Mining Your Data

Stephanie Thompson

Data mining involves large amounts of data from many sources. In order to successfully extract knowledge from data, you need to do a bit of work before running models. This paper covers selecting your target and data preparation. You want to make sure you find golden nuggets and not pyrite. The work done up front will make sure your panning yields results and is not just a trip down an empty shaft.



SESUG Speaker Sharing Program

While You Were Sleeping.....SAS® Was Hard At Work

Andrea Wainwright-Zimmerman

Automating and scheduling SAS code to run over night has many advantages, but there are also many pitfalls to be aware of. This presentation discusses how to make sure you save the log, get e-mail messages at critical points, as well as makes other recommendations and considerations on how to make it all work.

Why the Bell Tolls 108 times? Stepping Through Time with SAS

Peter Eberhardt

For many SAS programmers, new or even advanced, the use of SAS date and datetime variables is often very confusing. This paper addresses the problems that the most of programmers have. It starts by looking at the basic underlying difference between the data representation and the visual representation of date, datetime and time variables. From there it discusses how to change data representations into visual representations through the use of SAS formats. Since date manipulation is core to many business process, paper also discusses date arithmetic first by demonstrating the use of simple arithmetic to increment dates; then by moving on to SAS functions which create, extract and manipulate SAS date and datetime variables. Finally, paper demonstrates the use of the %sysfunc macro function and the %let statement to present date, datetime and time variables. This paper is introductory and focuses on new SAS programmers, however, some advanced topics also covered.

This paper is introductory and is intended for new SAS programmers.

Windows and Unix Computers Now Have Multiple CPU's; Why Not Control Two or Three or More Parallel Executing SAS Batch Jobs from One Master Job!

William E. Benjamin Jr.

This paper will show the processes required to start multiple, parallel batch copies of SAS on the SAME Microsoft Windows or UNIX based computer. It may be easy to click the SAS Short-cut button or type SAS in a command prompt window a few times, but those jobs are independent of each other, and use resources to update the screen images for the operator (you). The process proposed here allows one SAS job to start others as child batch processes, with no screen update requirements, with their own parameters and monitor their completion. This paper exposes some of the tricks required to get two (or more) batch copies of SAS running on a Windows or Unix computer.

Working with Character Data

Andrew T. Kuligowski

The DATA Step allows one to read, write, and manipulate many types of data. As data evolves to a more free-form state, SAS's ability to handle character data becomes increasingly important. This presentation will address character data from multiple vantage points. For example, what is the "default" length of a character string, and why does it appear to change under different circumstances? What sort of formatting is available to us for character data? How can we examine and manipulate character data?..This paper will aim at a beginning to intermediate audience, and will stress an introduction to the numerous character functions available in SAS®, including the basic LENGTH, and SUBSTR, plus many others.