

Analytics & Statistics

On Deck: SAS/STAT® 9.3

Wednesday, 2:30-3:20
Curran

Maura Stokes, SAS Institute

SAS/STAT® 9.3, coming soon to a site near you, delivers numerous enhancements to the statistical software. The PHREG procedure supports frailty models for incorporating random effects in Cox regression, and the MCMC procedure provides a RANDOM statement to facilitate fitting Bayesian models with random effects. The NLIN procedure has been updated, and the MI procedure offers additional flexibility by providing a fully conditional specification method. The new SURVEYPHREG and HPMIXED procedures are also outfitted with additional capabilities.

This talk reviews the highlights of SAS/STAT 9.22 and then describes important SAS/STAT 9.3 enhancements with practical illustrations, mainly from the SAS/STAT 9.3 documentation.

Boost your confidence in a 2x2 standard table

Wednesday, 3:30-3:50
Curran

Iuliana Barbalau, ClinOps

According to note 24170 issued by SAS® for version SAS 9.1.2 in September 2004, there are no statistical procedures which directly produce statistics such as sensitivity, specificity, positive and negative predictive values, false positive and negative probabilities and their associate confidence intervals for a 2 x 2 table. In the past, various methods have been used such as formulas implemented as data steps to obtain these results. Starting with SAS 9.2 version, PROC FREQ can produce new confidence limits for the binomial proportion including Agresti-Coull, Jeffreys, and Wilson (score) confidence limits. This paper will show how a macro call can simplify a SAS programmer job by providing accurate and easy to use calculations for sensitivity, specificity, positive and negative predictive values using PROC FREQ options: Wilson(score) or Exact.

Beyond Breslow-Day: Homogeneity Across R x C Tables

Wednesday, 4:00-4:20
Curran

Ginny Lai, ICON Clinical Research

In the epidemiological world, we often encounter the following analytical question: is the relationship between exposure and outcome different for different strata? The Breslow-Day test (with or without the Tarone adjustment) can be used in PROC FREQ to assess homogeneity across a series of 2 x 2 tables, but what if your tables are not 2 x 2? One alternative is to fit a log-linear model to a k x R x C table and test the fit with the three-way interaction removed. This can be done easily in PROC CATMOD or, with some additional effort, in PROC GENMOD and PROC GLIMMIX. A real-world data example will demonstrate the details of different methods to analyze k x R x C tables and tests for homogeneity.

An Introduction to the Analysis of Rare Events

Wednesday, 4:30-5:20
Curran

Nate Derby, Stakana Analytics

Analyzing rare events like disease incidents, natural disasters, or component failures requires specialized statistical techniques since common methods like linear regression (PROC REG) are inappropriate. In this paper, we'll first explain what it means to use a statistical model, then explain why the most common one (linear regression) is inappropriate for rare events. Then we'll introduce the most basic statistical model for rare events: Poisson regression (using PROC GENMOD or PROC COUNTREG).

Estimating Disease Prevalence from Clinical Data Using Capture-Recapture

Wednesday, 5:30-5:50
Curran

Carol Conell, Kaiser Permanente

Capture-recapture analysis provides a natural way to model disease prevalence using the longitudinal diagnostic and treatment data from Electronic Medical Records. This technique corrects for ascertainment bias due to not yet diagnosed incident cases as well as cases not being actively treated. This paper shows how SAS® can be used to transform routinely generated data into a series of repeated measures of disease identification and to apply capture-recapture analysis to the resulting series. A simple example combining survey and clinical data on alcohol and substance use disorders is presented.

Assessing resource adequacy uncertainty in California electricity summer demand using the MCMC procedure in SAS® Enterprise Miner™

Wednesday, 6:00-6:20
Curran

Miguel garcia-cerrutti, ca energy commission

Resource adequacy is the ability of the electric system to supply the sufficient resources to satisfy demand reliably at all times. One key challenge for optimally estimating resource adequacy is that demand is difficult to predict with certainty, particularly during summer in California where the main factor affecting it is the temperature and the demand sensitivity to it. Cooling demand responses to extreme high summer temperatures generate sudden changes and large variability in demand testing the ability of the system to meet its highest demand and so predicting it with some degree of certainty is the ultimate goal in resource adequacy planning. This paper discusses how to implement the Markov chain Monte Carlo (MCMC) procedure with the SAS® Enterprise Miner™ platform to effectively fit Bayesian predictive modeling to the extreme cooling demand distribution to accurately assess risk in electricity summer demand. Specifically, it discusses how extreme value theory (EVT) analysis of extreme demand can be used to estimate its probability distribution and calculate extreme demand quantiles for low probabilities. In addition, the paper discusses how to extend SAS® Enterprise Miner™ capabilities by deploying an extension node under the model menu to perform EVT analysis.

“Before and After” Models in Observational Research Using Random Slopes and Intercepts

Thursday, 2:30-3:20
Curran

David Pasta, ICON Clinical Research

In observational data analyses, it is often useful to use patients as their own controls by comparing their outcomes “before” and “after” some signal event, such as the initiation of a new therapy. It may be useful to have a control group that does not have the event but instead is evaluated before and after some arbitrary point in time, such as their birthday. In this context, the change over time is a continuous outcome that can be modeled as a (possibly discontinuous) line, with the same or different slope before and after the event. Mixed models can be used to estimate random slopes and intercepts and compare patients between groups. A specific example published in a peer-reviewed journal is presented.

Fitting latency models in epidemiological studies

Thursday, 3:30-3:50
Curran

Bryan Langholz, University of Southern California

Latency models address important questions about the timing of exposure and subsequent disease risk in epidemiologic research. However, the non-standard form of such models and the complexity of time-varying exposure histories that characterize many epidemiologic studies make such models difficult to fit using standard software packages. SAS offers software tools that can be used to fit “non-standard” latency models. In particular, SAS procedures NLP and NLMIXED provide the flexibility to specify quite general latency models and maximize the likelihood for latency parameters. While the methods require more sophistication than using a package or procedure that is specialized to the usual log-linear form, the additional programming complexity is not great. The methods are illustrated and compared by fitting latency models for radon exposure and lung cancer mortality rates from a cohort study of Colorado Plateau uranium miners. It was found that the conditional logistic likelihood is more computationally efficient than the unconditional and that PROC NLP is much faster than PROC NLMIXED.

Principal Component Regression as a Countermeasure against Collinearity

Thursday, 4:00-4:20
Curran

Chong Ho Yu, Arizona State University

There are different approaches to counteract the threat of multicollinearity in regression modeling, such as centered-score regression, orthogonalization, partial least square, and ridge regression. Principal component regression (PCR) is an under-use option because it takes multiple steps to accomplish the goal. This paper will illustrate different steps of performing PCR using the data set compiled by Programme for International Student Assessment (PISA) and a few other databases. Typically, PCR consists of four steps: 1. Principal component analysis (PCA), 2. Principal component regression (PCR) under partial least squares (PLS), 3. Factor analysis, and 4. OLS regression. Initially, PCA is run to verify whether collinear predictors could be combined to form a composite score. Further, the component structure is verified by principal component regression under PLS. While PCA suggests the proper number of principal components by indicating the loadings, PCR makes the same type of suggestion based on the PRESS statistics and variance explained in the model effects. Next, in order to obtain a set of better weights to form a composite index, factor analysis with the varimax rotation is employed. Last, the composite index is used to run the OLS regression model.

Evaluation of the Psychometric Properties of the Inventory of the Dimensions of Emerging Adulthood

**Thursday, 4:30-4:50
Curran**

Nadra Lisha, University of Southern California

It is now presumed that youth do not move directly from adolescence to adulthood, but rather pass through a transitional period, "emerging adulthood." The Inventory of the Dimensions of Emerging Adulthood (IDEA) is a self-report instrument developed to examine the attributes of this period. In the present study, a 21-item version of the IDEA was administered to a sample of 1676 "at risk" continuation (alternative) high school students in Southern California. Principal component factor analysis with orthogonal rotation revealed three factors. Overall, the measure demonstrated high internal consistency. Construct validity analyses indicated that the measure was correlated with demographics, risk behaviors, and psychological measures. We conclude that the IDEA is a useful instrument for measuring emerging adulthood in at-risk populations.

Why Dummy Variable Makes You SMART, and How to Do it SEXY

**Thursday, 5:00-5:20
Curran**

Brian (Guanghui) Sun, Rick Hansen Institute

When considering a categorical variable as an independent variable in a regression analysis, should you directly include the categorical variable into your model by using CLASS statement or instead include multiple dummy variables converted from levels of the categorical variable? Theoretically those two approaches are equivalent in terms of estimation results. The CLASS statement offers various possibilities of variable parameterization. However, the availability and power of CLASS statement differs in different SAS regression procedures. So the dummy variable approach is still very appealing for the consistency, flexibility and transparency of variable parameterization. In practice the dummy variable approach also has advantages in terms of ODS statistical graphics output, variable Boolean calculation, and model identification for better fit. However, coding dummies can be very tedious and painful. This paper introduces two approaches for automatic dummy variable generation, which makes the application of dummy variables easy and efficient.

Consistency of Treatment Effect cross Regions

**Thursday, 5:30-5:50
Curran**

Yunfeng Li, Cerexa, Inc.

Multi-Regional Clinical Trials are becoming increasingly common. The evaluation of the consistency of treatment effect cross regions is not only important to support the robustness of the conclusions, but is also frequently requested by local regulatory agencies in support of registration. There are numerous recent publications on this topic. Most of these focus on superiority trials with a continuous endpoint. In this presentation, we will explore the consistency of treatment effect across regions in the context of a non-inferiority trial with a binary endpoint using simulations. It is anticipated that this approach will help plan for the potential impact of observed regional differences and thus help to improve study design and understand the impact on sample size calculations.

What's happening if there is "B" in the general linear modeling output?

Thursday, 6:00-6:20

Jason Shilong Kuang, Kelley Blue Book, Inc.

Curran

The SAS® GLM procedure is an efficient tool in statistical data analysis, especially when we have categorical variables (also called class variables) as predictors. In SAS® practice, you may see the letter "B" showing up sometimes in the parameter estimate output, and you may wonder what's happening in our model. What is the cause of this? Can we still trust our model? Can we verify the modeling output by some alternative procedures? Where should we put more attention for the similar situations in future?

In this paper, we answer those questions by demonstrating some intuitively understandable examples, with the corresponding theoretical statistical background attached. With those in mind, you can easily turn the general linear modeling into a more powerful tool! SAS® ROCKS!

Applications Development

Reading and Writing RTF Documents as Data: Automatic Completion of CONSORT Flow Diagrams

Wednesday, 2:30-2:50
Fillmore A&B

Art Carpenter, CA Occidental Consultants

Whenever the results of a randomized clinical trial are reported in scientific journals, the published paper must adhere to the CONSORT (CONsolidated Standards Of Reporting Trials) statement. The statement includes a flow diagram, and the generation of these CONSORT flow diagrams is always problematic, especially when the trial is not the typical two-arm parallel design. Templates of the typical two-arm design flow diagram are generally available as RTF documents, however the completion of the individual fields within the diagram is both time consuming and prone to error. The SAS Macro language was used to read a the RTF template file for the CONSORT flow diagram of choice, fill in the fields using information available to the SAS program, and then rewrite the table as a completed RTF CONSORT flow diagram. This paper describes the process of reading and writing RTF files.

Working the System: Our Best SAS® Options

Wednesday, 3:00-3:20
Fillmore A&B

Patrick Thornton, SRI International

This paper provides an overview of SAS system options, and discusses our best options—those we have found most beneficial, interesting or both. There are many ways to work the SAS System by changing the settings of one or more of 100s of system options. These settings may have sweeping effects that influence, for example, DATA steps (e.g. REPLACE, MERGENOBY=), the log (MSGLEVEL=), data sets (e.g. VALIDVARNAME), listing output (e.g. NODATE, NONUMBER, ORIENTATION=, FORMDLIM=), and other destinations (e.g. PDFSECURITY=). Option settings are invaluable for working with format catalogs (e.g. FMTSEARCH=) and they are essential for using, developing and debugging macro programs (e.g. MPRINT, MLOGIC, MPRINT, and SYMBOLGEN, or new to SAS 9.2 MCOMPILE, MCOMPILENOTE). Designed for an intermediate user of SAS, this paper demonstrates our best options.

SAS® Macros in an Era of Apps

Wednesday, 3:30-3:50
Fillmore A&B

Sy Truong, Meta-Xceed, Inc.

Software development is a dynamic environment shifting from large enterprise systems to small utilities or "apps". The success of marketplaces such as Apple's App Stores has proven that specialized small apps used for specific purposes are essential for large groups of users with diverse requirements. This paper takes a new view of SAS macros in the framework of mobile apps. SAS Macros can thus be showcased to a community of users such as an App Store and then automate the process of downloading and installing, analogous to mobile apps. Mobile App Stores has been successful in delivering millions of apps since it has made it easy for users to find their app through user reviews, documentation and screenshots. Some even allow users to test drive the apps. The Macro App Store allows users to submit the macro with sample data as a test drive. Once an app is found, it can be instantly installed and used upon a single click. The Macro App Store applies the same methodologies to many free SAS macros that were once left in obscurity.

Benefits of JAVA Within SAS: Exploring the Differences of the new JavaObj and SAS/IML Studio

Wednesday, 4:00-4:20
University at San Luis Obispo
Fillmore A&B

Diana Shealy, California Polytechnic State

Released into production in SAS® 9.2, the JavaObj is a new interface to Java programs and classes. Yet, this is not the first and only way to access Java's abilities for SAS. Another way to work with Java using SAS is with the programming language for SAS/IML Studio called IMLPlus. The purpose of this paper is to compare and contrast these two methods for using library-defined or user-defined Java classes. While this paper does not concentrate on Java's massive GUI and graphics packages, it does focus on using Java programs to calculate complex equations and algorithms. There are two main benefits to using Java to write such algorithms: code reuse, if the code is already written in Java, and the flexibility of the Java language. This paper will walk the user through basic Java code and how to access that code through both the JavaObj and through IMLPlus, as well as educate the user on both the advantages and limitations of each method.

%GetTweet: A New SAS® Macro to Fetch and Summarize Tweets

Wednesday, 4:30-4:50
Fillmore A&B

Satish Garla, Oklahoma State University

The role of Twitter as a source of valuable information for spotting trends has been much talked about in the popular press. The open Application Program Interface (API) in Twitter makes it one of the most sought after platforms for textual data analysis. While SAS® Text Miner provides a robust method for analyzing textual data, the challenge remains to fetch customized Tweets and clean textual data before any text mining.

This paper develops and discusses a new SAS® macro that can be used easily to fetch the Tweets a researcher wants from Twitter. The macro uses the search API in Twitter and the HTTP procedure in SAS to create a data set of Tweets that are customized using parameters such as combination of keywords, exact phrases, omission of specific words, and so on. The macro also purges terms such as http tags and URLs that might create problems in text mining. This paper also shows a visual analysis of "retweets" to identify influencers via network visualization graphs that are available in SAS/GRAPH® software.

Protecting Macros and Macro Variables: It Is All About Control

Wednesday, 5:00-5:50
Fillmore A&B

Art Carpenter, CA Occidental Consultants

In a regulated environment it is crucially important that we are able to control, which version of a macro, and which version of a macro variable, is being used at any given time. As crucial as this is, in the SAS® macro language this is not something that is easily accomplished. We can write an application that calls a macro that we have validated, but the user of the application can write and force the use their user written un-validated version of that same macro. Values of macro variables that we have populated can be 'accidentally' replaced by user written assignments. How can you guarantee that the end results are accurate if you cannot guarantee that the proper programs have been executed?

Although our tools are limited, there are a few options available that can be used to help us control our macro execution environment. These, along with management programs, can give the application developer better control, and greater protection, during the execution of the application.

For a successful macro language application, it is all about CONTROL!!

Automating file pulls for recurring jobs ad-hoc reports or other SAS processes

Friday, 9:00-9:20
Fillmore A&B

Ezekiel Budda, Department of Defense

Many times when running jobs, you only want the latest information. Along with running a program, you need to configure the program to read in the proper file, set up date-dependent variables and create the formats needed to output your data in a user friendly format. In general, setting up all of these things might not be much of a chore, but what about when you need to do it every year, month or day, or if you need to read in hundreds of records for a time period and you're not sure if they're all available? These issues by themselves might not be too much either, but what happens when you want the same steps performed for hundreds of programs? Do you want to have to manually add

and update code for each program, or would you rather use one line of code that is no more complicated than setting up a file reference? That is the goal and subject of this paper.

How to easily convert clinical data to CDISC SDTM

Friday, 9:30-9:50
Fillmore A&B

Ale Gicqueau, Clinovo

Sponsors are receiving clinical information of increased complexity, from multiple sources and different formats. As a result, clinical data submission has become more time-consuming, costly and error-prone. To tackle this challenge, CDISC® (Clinical Data Interchange Standards Consortium) has been establishing new non-proprietary clinical data standards to speed up data-review and improve clinical data exchange, storage and archival. Conforming to these recognized CDISC standards improves and significantly speeds up FDA submission and FDA review. In addition, converting clinical data to a standardized format will improve SAS code re-usability for many programs used in data management and biostatistics such as Edit Checks, Patient Profile, TLGs, and custom reports.

SAS is often used as an ETL tool to manually convert SAS extracts from a clinical database to SDTM format. While this is a reasonable approach, it can quickly become tedious, error-prone, and time consuming. CDISC Express is a powerful open source SAS®-based clinical data management system that automatically and systematically converts clinical data into CDISC SDTM using an Excel framework. All CDISC Express mapping definitions and rules are defined in Excel, which are dynamically converted into a SAS program that automatically performs the SDTM transformation and validation through a series of SAS macros

CDISC Express source code is freely available, well-documented and easily understandable; it can be easily modified by any SAS programmer to fit his company SAS infrastructure.

This paper will provide SAS programmers with an introduction to CDISC Express, and show how the SAS programs and configuration files are organized. We will also show how to create macros, and convert clinical data to CDISC SDTM domains.

Managing Your Intellectual Property: Optimizing Your Investment of Resources by Building a Searchable SAS® Program Repository

Friday, 10:00-10:50
Fillmore A&B

Joe Perry, Perry & Associates Consulting

The sad fact is that after years of accumulating programs either individually or as an organization the width, breadth and innovative programming techniques can get lost. Unfortunately, when looking for key code, MS Window's search engine can miss character strings and, important for the task of managing your intellectual property (IP), lack the tools to properly search this source of IP.

This paper discusses application design techniques and a macro that will create a searchable code repository that can help answer the key questions individuals and organizations should be asking about the library of work they have accumulated over the years. Questions like... "Has someone done something like this before?" "Where have I seen that code before?" and "Who is really using the departmental macros we have spent all that money building and are called out in our SOPs?"

This paper is the result of a thirty-minute Code Doctor's session but has been expanded to both solve a bigger problem and teach application development methodologies that can be used in many other areas of your professional programming life.

Turbocharging Your Microsoft Office Environment with version 4.3 of the SAS® Add-In for Microsoft Office

Friday, 11:00-11:50
Fillmore A&B

Kevin Bickford, SAS Institute

Some of the most commonly used tools by users of all skill levels across organizations are Microsoft Office tools, such as Word, Excel, and PowerPoint. The SAS® Add-In for Microsoft Office turbocharges these tools, giving users the power to explore and analyze information using SAS software within their familiar Microsoft Office environment. The SAS® integration with Microsoft Office brings a large set of capabilities enabling users to access reports, stored processes, and dashboards; carry out multidimensional OLAP analysis in spreadsheets with auto-charting; run

interactive reports for quick what-if analysis; and securely work with data directly. It brings capabilities allowing users to reuse the work done, thereby increasing user productivity.

This paper highlights the key capabilities provided by the SAS® Add-In 4.3 for Microsoft Office and the application of these capabilities across the Microsoft Office products.

Business Intelligence

Comparison of K-means, Normal Mixtures and Probabilistic-D Clustering for B2B Segmentation using Customers' Perceptions

Wednesday, 2:30-2:50
Orpheum

Satish Garla, Oklahoma State University

Cluster Analysis is a popular technique used by businesses and analysts for market segmentation. For segmentation, clustering is used to split customers in a market into meaningful groups such that the customers within a group are similar and customers between the groups are dissimilar. Several clustering methods and numerous clustering algorithms are available in existing software packages and new ones frequently appear in the literature. These methods and algorithms vary depending on how the similarity between observations is defined or on other assumptions about shapes of clusters, distributions of variables, etc. This paper describes a comparative study of three clustering methods (K-means, Normal Mixtures and Probabilistic-D) for segmenting business-to-business (B2B) customers using their perceptions.

The "hard" clustering techniques such as K-means follow a deterministic approach in calculating cluster membership whereas the "soft" clustering techniques like Normal Mixtures calculate a degree of membership or probability for each customer to belong to a cluster. The Normal Mixtures technique, trained by the expectation-maximization algorithm, uses probability estimates via an iterative classification method. A new SAS® macro was developed for application of probabilistic-D technique. The macro calculates probability of cluster membership using the Euclidean distance of each observation from cluster centers found by k-means. These two soft clustering techniques are compared with the much widely used K-means technique. The results from each method are evaluated based on purity and cluster profiles. SAS® Enterprise Miner is used for probabilistic-D clustering and for profiling clusters generated from all the three techniques while JMP® is used for K-means and Normal Mixtures. Our results show that a better understanding of markets can be achieved using soft clustering techniques.

If you imbed it...?

Wednesday, 3:00-3:20
Denver
Orpheum

Donald Price, Metropolitan State College of

Businesses and organizations, faced with ever increasing competitive pressures, need to make better decisions. Better decision making requires superior analytics – the extensive use of data, statistical, quantitative, and qualitative analysis, exploratory and predictive models, and fact-based management. These same companies are faced with a growing talent shortage – a shortage of skilled analysts. Recent anecdotal articles and academic studies highlight the need to incorporate contemporary information technology developments in the practice of marketing. This paper provides insights into the efforts of Metropolitan State College of Denver to incorporate these developments into undergraduate marketing studies by imbedding the subject of analytics into marketing curriculum.

Using Affinity Models for Buyer Behaviors

Wednesday, 3:30-3:50
Orpheum

Matthias Kehder, Modern Analytics

Companies in the early stages of adopting analytics are testing and exploring strategies that produce ROI to justify the capital investment of implementing analytics. Other companies farther down the adoption path have made that investment and found the required ROI early on. By doing so, the early investors created a repertoire of analytical business strategies and their tactical results. It is these companies that are working on the next level of analytics – global optimization at the macro and micro level. Examples are: yearly marketing spend optimized across all channels down to the most logical entity, campaign spend of a charity to reduce funding gaps or increase donations for topics of interest, reducing a bank's risk exposure of each investment instrument on a position basis, or assisting a

global company in ensuring that local product development can be adopted world wide. This last topic is the focus of this paper and is addressed using affinity models.

Modeling Interactive Advertising on TiVo: Overcoming the Challenges of Collinearity, Efficiency, and Practicality

Wednesday, 4:00-4:20

Rachel Poulsen, TiVo

Orpheum

For interactive advertising on TiVo, benchmarking the performance of one advertisement against similar advertisements is important to many areas of the business. One of the measures of success of an interactive advertisement is the Click-through rate (CTR). Understanding the combination of ad placement and air time that will maximize the CTR for each advertisement provides the sales team with an attractive product, the advertising client with the best success, and the design team with an effective user-interface. There are many challenges associated with modeling CTR. The business demand for a practical and timely model coupled with the large amount of data associated with CTR presents a computational challenge. Many of the variables used to model CTR are measurements of human behavior, resulting in problems with collinearity. Some of the variables that appear to be unrelated will describe similar human behaviors. Using CTR data collected from TiVo interactive advertising, this paper will demonstrate how to identify and correct for collinearity using PROC PRINCOMP, create an ordinary least squares regression model using model selection techniques in SAS® (PROC REG), and deliver a practical statistical model in an accurate and timely manner.

Create Compelling Visualizations with Geographic Data and JMP® 9

Wednesday, 4:30-5:20

Laura Higgins, JMP

Orpheum

JMP 9 introduces exciting graphical support for geographic data. In this presentation, you learn how to use the built-in background maps in any plot of geographic data. You also see how to connect to a Web mapping service to display specialty maps like satellite images, radar images, or roadways. Because JMP 9 can plot geographic shapes based on place names, we demonstrate how to use the flexible architecture to create your own shape files to plot any location, like a floor plan or a campus. Lastly, we show how to convert SAS® map data sets for use in JMP.

Practically Perfect Presentations

Wednesday, 5:30-6:20

Cynthia Zender, SAS Institute

Orpheum

PROC REPORT is a powerful reporting procedure, whose output can be "practically perfect" when you add ODS STYLE= overrides to your PROC REPORT code. This hands-on workshop will feature several PROC REPORT programs that produce default output for ODS HTML, RTF and PDF destinations. Workshop attendees will learn how to modify the defaults to change elements of PROC REPORT output, such as HEADER cells, DATA cells, SUMMARY cells and LINE output using ODS STYLE= overrides. In addition, attendees will learn how to apply conditional formatting at the column or cell level and at the row level using PROC FORMAT techniques and CALL DEFINE techniques. Other topics include: table attributes that control interior table lines and table borders, use of logos in output and producing "Page x of y" page numbering. The rest of this document is divided into two parts: the slides used in the Hands-On Workshop presentation and the attendee worksheet.

Coders' Corner

Assigning a User-defined Macro to a Function Key

Wednesday, 2:30-2:40

Mary Rosenbloom, Edwards Lifesciences, LLC

Warfield

Are you entering one or more of the same SAS Display Manager System (DMS) commands repeatedly during a session? The DMS offers a convenient way of capturing and saving frequently entered commands in a user-defined macro, and then saving the macro as a function key of your choosing. Are you typing SAS code for data exploration during program development or validation, only to delete it soon afterwards? If you are, then this code can be placed

in a macro, too, and assigned to a function key. This paper illustrates the purpose and steps one would use to assign a user-defined macro to a function key.

You Could Be a SAS® Nerd If . . .

Wednesday, 2:45-2:55
Corporation
Warfield

Kirk Paul Lafler, Software Intelligence

Are you a SAS® nerd? Wiktionary, a wiki-based Open Content dictionary, defines "nerd" as a person who has good technical or scientific skills, but is generally introspective or introverted. Another definition is a person who is intelligent but socially and physically awkward. Obviously there are other definitions for "nerd", many of which are associated with derogatory terms or stereotypes. This presentation focuses not on the negative descriptions associated with being a nerd, but on the many positive traits SAS users possess. So let's see how nerdy you actually are using the mostly unscientific, but fun, "Nerd" detector.

Macro Tabulating Missing Values, Leveraging SAS® PROC CONTENTS

Wednesday, 3:00-3:10
Warfield

Adam Chow, Dept. of Veterans Affairs

Metadata, in a broad context, is data about data. Executing a SAS PROC CONTENTS is a form of metadata. When you run a PROC CONTENTS on a SAS data set, the results tell you various attributes of the data set. Moreover, it details each variable in the data set, such as the name, the variable type (numeric or character), the variable label that makes the name's mnemonic a little bit more meaningful, and the variable format that instructs the SAS System on how the values are to be displayed externally. As insightful as PROC CONTENTS is, however, it does not tell you anything about missing values. Accordingly, by leveraging the results from PROC CONTENTS, the %CHK_MISSING macro described in this paper, tabulates missing values for each variable in a data set and reports the count and the corresponding percentage in a separate data set. This paper will: 1) summarize the macro specification; 2) examine the design methodology and syntax; and 3) demonstrate how it is used with an illustration. The %CHK_MISSING macro was developed and tested under Base SAS 9.1.3 Service Pack 4 in UNIX, SUN operating system 5.10 platform.

Hunting for Columbus' Eggs in the SAS® Programming World: A Guidance to Creative Thinking for SAS® Programmers

Wednesday, 3:15-3:25
Warfield

Alice Cheng, Independent

A Columbus" Egg, according to Wikipedia, refers to a brilliant idea or discovery that seems simple or easy after the fact. In this paper, the author presents several Columbus" Eggs that may be helpful to fellow SAS® programmers. SAS examples are given, followed by learning tips to remind readers what they actually have learned. Examples include catchy phrases to remember basic SQL structure, an innovative use of a SAS function, methods to white out/cover up unwanted information, horizontal vs. vertical approach to resolve a problem, as well as, divide and conquer technique. It is the author"s intention to not only provide readers with useful techniques, but more importantly, to stimulate them to think outside the box and hunt for their own Columbus" Eggs!

Output SAS® DATA Step Views: an Experimental Feature

Wednesday, 3:30-3:40
Warfield

Thomas Billings, Union Bank

Output DATA step views are a specialized experimental feature in the Base SAS system. They are compiled `_null_` DATA steps that are executed as output from a PROC or DATA step. Simple examples of output DATA step views are demonstrated and explained. Differences between output views and other types of views supported in SAS are described. The limitations/constraints of output views are specified, e.g., views are compiled hence the PDV (program data vector) must be fully defined and cannot change, and `_null_` DATA steps (by themselves) don't create an output SAS data set. Possible approaches to mitigate these limitations are discussed, including macros to create a set of

ATTRIB statements to define the PDV, and the possibility of using a hash object to produce an output SAS data set from a view. Potential applications of output views are described.

Exporting Variable Labels as Column Headers in Excel using SAS®

Wednesday, 3:45-3:55

Chaitanya Chowdagam, MaxisIT

Warfield

Excel output is the desired format for most of the ad-hoc reports requested for an easy interpretation and exploration of data. However, to make these reports more informative variable labels could be used instead of variable names as column header. This is harder to achieve by the conventional SAS® Export procedure, PROC EXPORT. This paper will provide easy and quick methods of exporting SAS® data to multiple Excel sheets with variable labels as column headers using EXCEL Libname and ODS Markup.

SAS Data Step Debugger: Your Liberator from Logic Errors

Wednesday, 4:00-4:10

Brandi Rhoads, Franchise Tax Board

Warfield

Have you ever written an intricate data step only to find that when you run the program the results are not as you expect? The SAS log is free of errors meaning you have rid your program of syntax errors; however, you know something is wrong. It must be a frustrating logic error! Using a simple example, this paper will illustrate basic debugger features you can use to find logic errors even in the most complex data steps. These techniques are useful to any programmer of SAS data steps, particularly those without prior knowledge of the data step debugger.

Producing Cumulative Distribution Frequency Figures with Minimal Important Difference Reference Lines to Assess Quality of Life Treatment Effects

Wednesday, 4:15-4:25

Xiao Liu, ICON Late Phase & Outcomes Research

Warfield

Quality of Life (QoL) data are generally collected in clinical trials at baseline and follow-up. One method of assessing treatment effects on QoL involves generating cumulative distribution frequency (CDF) figures. These figures illustrate the probability that a QoL score will have changes by at least a specific amount. By including vertical lines representing the Minimal Important Difference (MID) – i.e., the smallest change in QoL score perceived to be important – the cumulative percentage of patients who achieve at least the MID can be obtained for each treatment group. If a data point does not exist at the exact value of the MID, interpolation using adjacent data points is necessary. The treatment group with larger percentages at the MID is considered superior. The purpose of this paper is to describe the methodology that can be used to produce two CDF figures – each representing a QoL score with opposite directionality – and to identify the cumulative percentages for each MID (one based on 1 standard error of the mean and another based on 1/2 standard deviation). The CDF figures are generated using SAS/GRAPH and include inset tables, produced using the annotate facility, showing the percentages for each MID by treatment group.

Creating SAS® Datasets from Varied Sources

Wednesday, 4:30-4:40

Sofia Shamas, Maxisit Inc

Warfield

Often SAS® programmers find themselves dealing with data coming from multiple sources and usually in different formats. Steps have to be taken to logically relate the process and convert the variety of data into SAS® data sets before it can be analyzed. Since these sources do not follow a similar pattern, this paper is to serve as a collection of examples illustrating the conversion of data coming from various sources, such as extensible markup language (XML), comma separated values (CSV), Microsoft excel (XLS), or tab delimited (TXT) files to SAS® data sets.

Has Anybody Opened My File? Find Out Before Trying to Update it

Wednesday, 4:45-4:55

Olena Galligan, Teikoku Pharma, USA

Warfield

We've all been in the situation where we run a long code just to find out several minutes later that somebody opened

a file which your program is trying to update. Thus, no file update for you at this time. The macro demonstrated in this paper saves your time by letting you find out right away whether the file (.RTF, .DOC, DOCX, .XLS, XLSX, SAS7BDAT, .PDF) is opened by somebody and thus, whether it could be updated by your program. SCL (SAS Component Language) does most of the work, and I will describe not only my macro, but also some of the SCL elements you'll find useful in your everyday programming activities.

Adding PROC SQL to your SAS® toolbox for merging multiple tables without common variable names

Wednesday, 5:00-5:10
Diego
Warfield

Susan Wancewicz, University of California, San Diego

The SQL procedure offers an efficient method of creating a new dataset by merging tables in SAS®. Advantages of PROC SQL over the SAS Merge statement include: tables do not need to be sorted before joining them; and tables without a common variable can be joined simultaneously. This paper will lead the SAS user through the following steps in PROC SQL: The basic join, join of tables without common variables, a demonstration of grouping variables in order to obtain a new variable containing the average of an existing variable. At the end of this discussion the SAS user will add flexibility to their SAS toolbox for creating a new merged table.

Best Practices: Clean House to Avoid Hangovers

Wednesday, 5:15-5:25
Warfield

Mary Rosenbloom, Edwards Lifesciences, LLC

In a production environment, where dozens of programs are run in sequence, often monthly or quarterly, and where logs can span thousands of lines, it's easy to overlook the small stuff. Maybe a data statement fails to execute, but one already exists in the temp library from a previous program. Maybe a global macro assignment is missed or fails to execute, but a global macro of the same name already exists from a previous program. This can also happen with macros. The list goes on. This paper offers some suggestions for housekeeping steps that can be taken at the end of each SAS program to minimize the chance of a hangover.

VBA Microsoft Access 2007 Macros to Import Formats and Labels to SAS®

Wednesday, 5:30-5:40
Medical Sciences
Warfield

Maria Melguizo Castro, University of Arkansas for

It is a common practice for statisticians and data analysts to import data from Microsoft Access into SAS for analysis. Although it is relatively straightforward to transfer data between the two programs, currently there is not an option to fully transfer descriptions or data formats into SAS. This paper proposes a solution using both Visual Basic for Applications (VBA) and SAS macros to address this issue and create a more complete SAS dataset with descriptions as variable labels and with appropriate data formats.

SAS Date refresher: cleaning up legacy code using ANYDTDTEw., ANYDTCMw. and ANYDTCMEw. informats

Wednesday, 5:45-5:55
Warfield

Laura Kelly, Bank of America

Busy SAS programmers often spend too much time dealing with legacy SAS code surrounding the handling and formatting of incoming date data. SAS version 9 introduced some useful data informats -- ANYDTDTEw., ANYDTCMw. and ANYDTCMEw. This is a quick review of how to use these functions to clean up some old-fashioned legacy code that is cumbersome and often a headache to maintain.

Macro Quoting to the Rescue: Passing Special Characters

Wednesday, 6:00-6:10
Warfield

Mary Rosenbloom, Edwards Lifesciences, LLC

We know that we should always try to avoid storing special characters in macro variables. We know that there are just too many ways that special characters can cause problems when the macro variable is resolved. Sometimes, however, we just do not have a choice. Sometimes the characters must be stored in the macro variable whether we like it or not. And when they appear we need to know how to deal with them. We need to know which macro quoting functions will solve the problem, and even more importantly why we need to use them.

This paper takes a quick look at the problems associated with the resolution and use of macro variables that contain special characters such as commas, quotes, ampersands, and parentheses.

Are You Taking it Day by Day or Having an Episode? Identifying and Describing Patterns in Longitudinal Data

Thursday, 2:30-2:40
Warfield

Sarah Short, ICON Clinical Research

Data involving consecutive measurements over time arise in a variety of fields. Detecting specific patterns can be important in analysis, and some patterns that are reasonably easy to explain in common language are surprisingly tricky to identify with code that processes records sequentially one at a time. An example of this is identifying medication dosing patterns in a dataset that contains a record for each day. Clinicians may be particularly interested in looking at clinically meaningful medication dosing patterns, rather than simply calculating minimum, maximum, or mean dose over the entire treatment period. We present a method for outputting medication dose episodes from a patient-day-level dataset using a single data step with a specific goal of identifying the longest sustained dose for each patient.

Sam and Max's Adventure in SAS® Macro Land. (Quick and Easy tips for Macro programming)

Thursday, 2:45-2:55
Service
Warfield

Sam/Seunghee Chung, Educational testing

Sam was walking through SAS world one day when she started seeing strange characters all around her. She was in Macro Land! She had been there a few times before but, %SYSFUNC, &SQLOBS, and CALL SYMPUT were all strange and kind of scary. Then she spotted a tall figure leaning against a PROC. He said "Hi, I'm Macro Max. Do you need some help?" When Sam said "Yes," and Macro Max said "Don't be scared. %SYSFUNC is really a nice guy who just wants to help. And &SQLOBS can save you a few lines. Why don't I introduce you to them and some other macro buddies who can make macro coding easier and increase your efficiency?" That day, Sam made many macro friends that make her life much easier.

Quick ODS OUTPUT Statements

Thursday, 3:00-3:10
Warfield

Patrick Thornton, SRI International

The two short SAS® macro programs in this paper facilitate almost instantaneous access to the data sets of results that are available from a given procedure through ODS OUTPUT. Typically programmers manually create ODS OUTPUT statements by submitting a procedure bracketed by ODS TRACE ON and ODS TRACE OFF, copy the ODS object names from the log, and resubmit the procedure with an ODS OUTPUT statement that incorporates the object names. Using the macro programs in this paper, you can often achieve the same result easier by bracketing the first run of a procedure with calls to the macro programs, followed by a second run of the procedure and an ODS OUTPUT CLOSE. The first macro program, called just prior to a procedure, uses PROC PRINTTO and FILENAME to direct the procedure log and output to temporary locations. ODS TRACE ON is used to list the ODS object names in the log. The second macro program, called just after the procedure, reads the log and creates an ODS OUTPUT statement that incorporates the ODS object names, restores the default log and output destinations, and turns trace off. A second call to the procedure followed by ODS OUTPUT CLOSE then creates all the data sets available from the procedure via ODS and creates a single visible log and listing output. This paper is suitable for a beginning programmer and these macros have been tested on a Windows OS.

Conditional Processing Using the Case Expression in PROC SQL®

Thursday, 3:15-3:25
Corporation
Warfield

Kirk Paul Lafler, Software Intelligence

The SQL® procedure supports conditionally selecting result values from rows in a table (or view) in the form of a case expression. Similar to an IF-THEN construct in the DATA step, a case expression uses one or more WHEN-THEN clause(s) to conditionally process some but not all the rows in a table. An optional ELSE expression can be specified to handle an alternate action should none of the expression(s) identified in the WHEN condition(s) be satisfied. This presentation will illustrate the basic syntax associated with the two forms of case expression and examples of each.

Tag It, Bag It, Put It Out into Excel

Thursday, 3:30-3:40
Warfield

Ethan Miller, SRI International

So you have to put data out from SAS_i to Excel, again. You don't want the clients to ignore your output like they did last time. To make your Excel output command attention, PROC REPORT and TAGSETS.MSOFFICE2K are your weapons of choice. This paper will illustrate how to create hyperlinks, highlight rows, and conditionally shade cells (traffic lighting). This technique is an easy way to put out data in Excel for people who do not use SAS. This paper was written for those with beginning skills; the code was written using SAS version 9.2 on Windows OS.

Using PROC SQL to Calculate FIRSSTOBS

Thursday, 3:45-3:55
Warfield

David Tabano, Kaiser Permanente

The power of SAS® programming can at times be greatly improved using PROC SQL statements for formatting and manipulating data, and there is considerable literature available on comparing the strengths of using PROC SQL over the DATA step. This paper expands on such programming comparisons by reviewing the use of the DATA step and PROC SQL to select first observation(s) of a particular incident with a specified date range across repeated measures in datasets.

Efficient Techniques and Tips in Handling Large Datasets

Thursday, 4:00-4:10
Warfield

Jason Shilong Kuang, Kelley Blue Book, Inc.

When we work on millions of records, with hundreds of variables, it is crucial how we are processing our data. To make SAS® really ROCK, we need to pay more attention to SAS® program efficiency, since a single data step or some SQL query may take a few hours in dealing with such large datasets. In this paper, we present a few practical efficient techniques and hands-on tips in handling large datasets, including the application of INDEX, separating one single step into multi-step to improve efficiency, the classic Where vs. If statement, some tips in joining large datasets in PROC SQL etc.

To see the efficiency of those techniques, we also provide for each case with experimental example output, how much for the time-resource consuming, "apple-to-apple" comparison between the processes with and without those techniques. With those tips in our large data practice, we can save a lot of space and time, SAS® ROCKS!

Customized Graphics Made Simpler

Thursday, 4:15-4:25
Warfield

Sofia Shamas, Maxisit Inc

With the production release of SAS® Graph Template Language (GTL) and the introduction of Statistical Graphics (SG) procedures, it has become much easier to customize graphical output. These new SAS® features combined with the Output Delivery System (ODS) allows user to customize graphs with less efforts than traditional SAS® graph methods. This paper will illustrate using specific SAS® codes, how to utilize GTL and SG procedures to create eye-catching graphs. Scatter Plot and Kaplan Meier will be used as examples in this presentation.

HTMLBlue That's Really, Really Where We're Going To!

Thursday, 4:30-4:50
Warfield

Warren Kuhfeld, SAS Institute

This talk describes the new style of the SAS windowing environment (display manager) including new defaults and the new HTMLBlue style. In SAS 9.3, ODS Graphics is enabled by default, the default destination is HTML, and the default style is HTMLBlue. He says he only needs 20 minutes for this presentation.

Using PROC CONTENTS and a Macro to Convert Internal Data Values to their Associated Format Values

Thursday, 5:00-5:10
Warfield

Mary Rosenbloom, Edwards Lifesciences, LLC

Working with formatted data can be cumbersome and error-prone. Presented here is a method for using PROC CONTENTS and a macro to generate data-driven code that can be used to effectively convert the numeric internal data values to the character values dictated by the associated format.

Sending Emails in SAS® to Facilitate Clinical Trial

Thursday, 5:15-5:25
Warfield

Frank Fan, Clinovo

Email has drastically changed our ways of working and communicating. In clinical trial data management, delivering reports to users is part of our daily activities. Most of our colleagues prefer receiving reports by email rather than on the web or in a shared drive. Email is indeed one of the most convenient ways of communication.

In this paper, we will present how to use SAS V9 to send reports via email. Before an email can be sent via SAS, the SAS configuration file needs to be modified. After addressing the configuration file, SAS syntax to send an email will be discussed. We will use several examples to illustrate how we can implement such practices.

Data Capture, Validation, Manipulation, & Integration

SAS® Formats and the Format Procedure

Thursday, 8:30-9:20
Warfield

Christine Riddiough, SAS Institute

You will learn some of the basics of SAS formats and the FORMAT procedure. This session discusses how formats can facilitate SAS programming, from displaying data in a variety of ways to grouping and looking up data. You will learn how to create and maintain permanent formats and how to use formats for a variety of tasks. For example, you can use formats to group data and to conditionally highlight procedure output.

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

Thursday, 9:30-9:50
LLC
Warfield

William Benjamin Jr, Owl Computer Consultancy

You have an input file with one record per month, and 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 input or 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.

The Many Uses of SQL Subqueries

Thursday, 10:00-10:20

Tasha Chapman, State of Oregon - DCBS

Warfield

Subqueries are very useful tools in SQL. A subquery is an embedded query nested within another query. They're most often used when a particular value that is needed for the outer query is unknown, but can be searched for using an SQL query. This paper acquaints the reader with the concept of subqueries, and provides some examples of how they might be used in the SELECT, FROM, and WHERE clause.

Helping Students Become Effective Industry Statisticians: Supplementing Science with Data Savvy

Thursday, 10:30-10:50
Nebraska-Lincoln
Warfield

Aleksandra Stein, Celerion/University of

In the strictly regulated environment of the pharmaceutical industry, a Contract Research Organization (CRO) provides a wide range of services including, but not limited to, "design of a protocol, selection or monitoring of investigations, evaluation of reports, and preparation of materials to be submitted to the Food and Drug Administration." [FDA 21 CFR 312.3(b)] With regulatory agencies' recent endorsement of CDISC data submission standards, statistical programmers and statistical scientists now need a solid, functional understanding of data to be effective. Academia, however, has not yet embraced a data-focused training of statisticians. This paper examines the emerging discrepancy between the scientific knowledge students are equipped with and the focus on data required to work in the industry. We also share our methods for enhancing communication and easing the transition between academia and industry.

Probabilistic Record Linkage in SAS®

Thursday, 11:00-11:50
Warfield

Glenn Wright, UCSF

Record linkage is the process of deciding which records from one or more databases refer to the same entity, even when those records do not match exactly on any combination of identifiers. For example, we may wish to determine that a record for "Glenn Thomas Wright" living at "1234 West Fifth Street" refers to the same person as a record for "Glen T Wright" living at "1234 Fifth Street," but not the same person as a record for "Denis Hulett" living at "678 Ninth Street." Probabilistic record linkage is a family of record linkage techniques that assigns similarity scores to pairs of records and treats all pairs that score above a certain threshold as matches. This paper describes a method for implementing probabilistic record linkage in SAS, using PROC SQL and other tools.

Data Presentation

The Wonderful Wizard Of ODS - Using The EXCELXP Tagset Coupled With Proc Template, And Proc Report, To Twist Your World From Gray To All The Colors Of The Rainbow

Thursday, 8:30-8:50
Orpheum

Steven Black, W. L. Gore & Associates

In this paper I provide an easy to follow example using the power of ODS to get out of the gray and white into a world of color using the powerful EXCELXP tagset, TEMPLATE and REPORT procedures. I will show several methods of traffic lighting while providing a wonderful example of an Excel document full of colors, borders, and printability, melting away the wicked witches of confusion, bewilderment, and post processing. Enabling each SAS user the ability to lift off on their own without getting caught up by the flying monkeys.

SAS® Data Query and Edit Checks Web 2.0

Thursday, 9:00-9:50
Orpheum

Sy Truong, Meta-Xceed, Inc.

Querying data used to require sophisticated algorithms and programming with SQL expertise or SAS data step techniques. Now, there are many user-friendly graphical tools to build expressions without having to program code to access relational databases. However, there are few tools for creating queries from a web interface upon SAS data.

This paper describes a real world project where data managers who extract data from a relational database and transfer it to SAS data. Data managers need a user-friendly query tool to verify the cleanliness of the resulting SAS data. Since the data managers are not SAS programmers, the web interface eases the learning curve allowing them to perform tasks such as: creating compound query expressions, deriving temporary imputed variables, saving and combining saved queries. These are some of the methods described in this paper as it explains how to use SAS data step, HTML and XML to accomplish an effective user experience. The initial target audience for this project was to empower data managers to perform tasks only SAS programmers were able to do. However, the user-friendly interface extends it to a larger audience of non-techie power users.

Now You Can Annotate Your Statistical Graphics Procedure Graphs

Thursday, 10:00-10:50
Orpheum

Dan Heath, SAS Institute

Have you ever had a situation where you wanted to add custom labeling or other graphical features to your statistical graphics (SG) procedure output but could not find a way to do it? Even with all of the plot types and options of the SG procedures, you might have run into this situation. With SAS® 9.3, you can add these features directly by using the new annotation support in the SG procedures.

The annotation data set definition used by these procedures has been redesigned to take advantage of ODS Graphics functionality. This paper will discuss this functionality in detail, highlighting new features in this system that include the following:

rich text support (including superscripts, subscripts, and Unicode)

transparency support for all primitives

more drawing spaces for easier placement of annotations

Let's Give 'Em Something to TOC about: Transforming the Table of Contents of Your PDF File

Thursday, 11:00-11:50
Orpheum

Chevell Parker, SAS Institute

In PDF files, the table of contents provides a map that helps your audience to navigate the document easily. However, the default table of contents that is generated by the SAS® Output Delivery System (ODS) destination, while informative, is fairly utilitarian. Your procedures and DATA steps generate tables and graphs that have meaning to you and your audience. Likewise, the table of contents should also be as meaningful as possible by clarifying the contents of your PDF. This paper explains and demonstrates step by step how to use the following statement, options, and procedures to customize your table of contents:

- the ODS PROCLABEL statement
- the CONTENTS= and the DESCRIPTION= options
- the DOCUMENT destination and procedure
- the TEMPLATE procedure

These tools provide you with the flexibility and the power to customize your table of contents so that you really leave your audience with something to TOC about!

Using SAS® GTL to Visualize Your Data When There is Too Much of It to Visualize

Thursday, 2:30-3:20
Orpheum

Nate Derby, Stakana Analytics

In many of the SAS Institute publications about the new ODS statistical graphics, there is an introductory statement that defines an “effective” graph as one that reveals “patterns, differences and uncertainty that are not readily apparent in tabular output” (Kuhfeld, 2010; Rodriguez, 2008; Rodriguez and Cartier, 2009). Good graphs are also said to “provoke questions that stimulate deeper investigation, and ... add visual clarity and rich content to reports and presentations.” Developing a good graph becomes a challenge, however, when input data map to crowded displays with overlapping points or lines. Such is the case with the Framingham Heart Study of 5209 subjects captured in the sashelp.heart data set and a series of 100 cumulative booking curves for the airline industry. In addition, interleaving

time series plots can be difficult to interpret, and patterns can be missed when lattice plot panels are charted out-of-order. In the paper, transparency, layering, data point rounding, median calculation, and color coding are among the techniques that are evaluated for their effectiveness to add visual clarity to graphics output. The following Graph Template Language (GTL) statements are referenced in the paper: ENTRY, HISTOGRAM, SCATTERPLOT, LINEPARM, REFERENCELINE, BANDPLOT, and SERIESPLOT plus layouts OVERLAY, DATAPANEL, LATTICE, and GRIDDED. GTL is chosen over SG PROCEDURES because of its greater graphics capability.

Is the Legend in your SAS/Graph® Output Still Telling the Right Story?

Thursday, 3:30-3:50
Orpheum

Alice Cheng, Independent

In clinical studies, researchers are often interested in the effect of treatment over time for multiple treatments or dosing levels. Usually, in a graphical report, the measurement of treatment effect is in the vertical axis and a second factor, such as time and visit number, in the horizontal axis. Multiple lines are displayed in the same figure; each line represents a third factor, such as treatment or dosing groups. It is critical that the line appearances (color, symbol and style) are consistent throughout the entire clinical report, as well as, across clinical reports from related studies. Flavin and Carpenter (2004) showed that the GPLOT procedure, by default, did not guarantee consistency in line appearances. They provided macro and non-macro solutions to this problem. With the introduction of Statistical Graphics in SAS® version 9.2, there are multiple approaches to resolve this problem. In this paper, the author will cover the following topics:

The Nature of SGPLOT (How are Line Attributes Assigned?)

Re-visit the Line Inconsistency Problem by means of SGPLOT Procedure

5 Solutions to Resolve Line Inconsistency Issues

The SGPANEL Procedure (as a Special Case of SGPLOT Procedure)

A Guide for Using SAS® and OpenXML to Create Charts in MS Excel®

Thursday, 4:00-4:20
Orpheum

James Van Campen, SRI International

A system of SAS macros developed by the author for generating charts in Excel is presented. First, there is an overview of OpenXML as it applies to Excel. Second, there is a discussion of how to create Excel workbooks with SAS. Finally, an example of an Excel workbook including a chart and the SAS code used to create it are presented. Intermediate SAS programming skills are assumed, including familiarity with the SAS macro facility. All examples use SAS 9.2 and MS Office 2007.

Tips and Techniques for Automating the SAS® Add-In for Microsoft Office with Visual Basic for Applications

Thursday, 4:30-5:20
Orpheum

Tim Beese, SAS Institute

Do you want to run SAS® Stored Processes in Microsoft Excel and set the prompt values from your worksheet? Do you want to filter your SAS data sets based on the values of specific cells? Do you want to customize your Microsoft Office content with buttons and other controls that allow you to open data, run SAS code and refresh results? This is now possible!

The SAS® Add-In 4.3 for Microsoft Office adds many new features that allow users to interact with their content through Visual Basic for Applications (VBA). Users can insert and refresh data, stored processes, and reports using VBA. It is also possible to provide prompt values to stored processes, filter and sort strings for data, and control where results are displayed. Combining these features with the existing functionality provided by VBA in Microsoft Office, integration of SAS within Microsoft Office will become easier and more powerful!

Graphing a Progression of Time Series Plots with ODS Graphics

Friday, 8:30-8:50

Nate Derby, Stakana Analytics

Orpheum

Graphing is an essential step for exploratory data analysis and statistical modeling. However, when graphing an ordered progression of time series plots, it can be difficult to effectively display the progression without looking disorganized and chaotic. This paper shows a couple of approaches to this problem using the GPLOT procedure from SAS®/GRAPH® software and the LAYOUT OVERLAY, LAYOUT DATAPANEL and SERIESPLOT statements from the Graphic Template Language (GTL) in ODS statistical graphics.

Making the Most Out of Multilabel Formats

Friday, 9:00-9:20

Tasha Chapman, State of Oregon - DCBS

Orpheum

This paper will discuss how to use multilabel formats, how to control the sort order of your formats, how to print rows with missing data, and how to maintain indentations and leading spaces in ODS output.

Multilabel formats allow a SAS® user to display over-lapping ranges. Among other things, multilabel formats allow a user to create hierarchical formats, showing multiple levels of groups and sub-groups in the same table.

Multilabel formats have been available since SAS Version 8, and are currently compatible with Proc Tabulate, Proc Means and Proc Summary. Multilabel formats cannot currently be used with Proc Report (as of version 9.2). The following examples were created in SAS version 9.1.3, using multilabel formats in Proc Tabulate and output using ODS PDF (SASweb style).

Data Visualization of User Interface Navigation Using Interactive Treeview Diagrams

Friday, 9:30-9:50

Aruna Buddana, TiVo Inc

Orpheum

Building an intuitive user interface depends on the ability to look at the usage of different pathways that users frequently use and constantly improvise them. TiVo's user interface famously known for its simplicity has been developed to add more features and functionalities and the research team often stumbles on questions including – what are the different paths that the users take to utilize a feature? Can the navigational paths be displayed, given the huge amounts of data?

In an attempt to answer these, we are researching effective ways to visually show the navigational paths and found that the DS2TREE macro integrated with Treeview applet from SAS9 is simpler and visually effective. The macro recognizes the input dataset and specifies how the variables in the dataset can be utilized to generate an interactive and customized node/link diagrams that perfectly shows the hierarchical relationships.

This paper discusses the application of DS2TREE macro in the user interface data and demonstrates the interactive graph capabilities offered by the Treeview applet. SAS Base and SAS Graph 9.2 are the SAS products used in the paper.

The Perfect Marriage: The SAS® Output Delivery System (ODS) and Microsoft Office

Friday, 10:00-10:50

Chevell Parker, SAS Institute

Orpheum

A compatible relationship is one in which the associated entities complement each other's strengths and compensate for each other's weaknesses. This relationship is certainly the case when you combine the power of ODS and the flexibility of Microsoft Office. This paper describes how you can create outstanding spreadsheets via the harmonious union of ODS with Microsoft Office applications. The paper explores the relationships among ODS destinations (for example, ExcelXP and MSOffice2K) as well as custom tagsets (for example, the TableEditor tagset). In addition, the paper explains how these components enable you to generate highly presentational spreadsheets.

This paper covers the following topics:

- specific layout techniques such as paneling tables and graphics and positioning output in a worksheet
- common issues (related to styles, cell formatting, and so on) that occur when you export data customization of your output in Microsoft Word using the RTF destination and the tagset

- generated output that can be read into Microsoft Access and other Office applications

SAS® Style Templates: Always in Fashion

Friday, 11:00-11:50
Orpheum

Cynthia Zender, SAS Institute

The syntax for style templates changed in SAS® 9.2 and became much easier to use. With the new CLASS statement, there is no need to debate the use of the STYLE statement versus the now defunct REPLACE statement. Yes, the REPLACE statement has gone away, and now we have the CLASS statement. In addition, you can import a CSS file into a SAS style template. All of these exciting new changes add up to stylish and fashionable output using ODS style templates.

This paper provides an introduction into the use of style templates in SAS 9.2. Methods of determining the correct style elements will be shown along with several concrete examples of making style template changes. The use of the IMPORT statement will also be demonstrated. In addition to these topics, a job aid will be provided that outlines the most commonly used style elements and their attributes.

Hands-on-Workshop

Ready To Become Really Productive Using PROC SQL?

Wednesday, 2:30-3:50
Conference Theatre

Sunil Gupta, Gupta Programming

Using PROC SQL, can you identify at least four ways to: select and create variables, create macro variables, create or modify table structure, and change table content? Learn how to apply multiple PROC SQL programming options through task-based examples. This hands-on workshop reviews topics in table access, retrieval, structure and content, as well as creating macro variables. References are provided for key PROC SQL books, relevant webinars, podcasts as well as key SAS® technical papers.

Quick Results with PROC SQL®

Wednesday, 4:00-5:20
Corporation
Conference Theatre

Kirk Paul Lafler, Software Intelligence

Structured Query Language (SQL) is a universal language that allows you to access data stored in relational databases or tables. This hands-on workshop presents core concepts and features on using PROC SQL to access data stored in relational database tables. Attendees learn how to define, access, and manipulate data from one or more tables using PROC SQL quickly and easily. Numerous examples are presented on how to construct simple queries, subset data, produce simple and effective output, summarize data with summary functions, and join two tables.

Statistical Graphics for Clinical Research Using ODS Graphics Designer

Thursday, 8:30-9:50
Conference Theatre

Wei Cheng, Isis Pharmaceuticals, Inc.

Statistical graphics play an important role across various stages in clinical research. They help investigators to explore and understand raw data in the early stage of statistical analysis, as well as present final analysis result in the formal publications. The graphs need to be specifically designed and carefully drawn to best represent data and analysis. While this can be done by SAS® programming using traditional SAS DATA steps and SAS/GRAPH procedures, the process is time consuming and time is spent to find the right options or annotation syntax.

With SAS 9.2M3, ODS Graphics Designer (Designer) becomes production software. This is an interactive “point and click” application that can be used to design and create custom graphs. This new application greatly enhances the ability to effectively generate statistical graphs for clinical research.

In this hands-on workshop, we will show you the application interface and walk you through creating some commonly used statistical graphs for clinical research. The intended audience doesn't need to know SAS/GRAPH syntax, but wants to create high-quality statistical graphs for clinical trials. Examples will use scrambled data from real world in CDISC format.

Creating Stylish Multi-Sheet Microsoft Excel Workbooks the Easy Way with SAS®

Thursday, 10:00-11:20
Conference Theatre

Vince DelGobbo, SAS Institute

This paper explains how to use Base SAS® 9 software to create multi-sheet Excel workbooks (for Excel versions 2002 and later). You learn step-by-step techniques for quickly and easily creating attractive multi-sheet Excel workbooks that contain your SAS output using the ExcelXP ODS tagset and ODS styles. The techniques that are presented in this paper can be used regardless of the platform on which SAS software is installed. You can even use them on a mainframe! Creating and delivering your workbooks on-demand and in real time using SAS server technology is discussed. Although the title is similar to previous papers by this author, this paper contains new and revised material not previously presented.

PROC TABULATE: Doing More

Thursday, 2:30-3:50
Conference Theatre

Art Carpenter, CA Occidental Consultants

Although PROC TABULATE has been a part of Base SAS® for a very long time, this powerful analytical and reporting procedure is very under utilized. TABULATE is different; it's step statement structure is unlike any other procedure. Because the programmer who wishes to learn the procedure must essentially learn a new programming language, one with radically different statement structure than elsewhere within SAS, many do not make the effort.

The basic statements will be introduced, and more importantly the introduction will provide a strategy for learning the statement structure. The statement structure relies on building blocks that can be identified and learned individually and in concert with others. Learn how these building blocks form the structure of the statement, how they fit together, and how they are used to design and create the final report will be presented.

Once the foundation is laid a number of intermediate level techniques, options, and statements that provide the TABULATE programmer with a wide range of power and flexibility are presented. These techniques include those that are often underutilized even by the more experienced TABULATE programmer.

Getting Up to Speed with PROC REPORT

Thursday, 4:00-5:20
Conference Theatre

Kimberly LeBouton, KJL Computing

Learning the basics of PROC REPORT can help the new SAS® user avoid hours of headaches. PROC REPORT can often be used in lieu of PROC TABULATE or DATA_NULL_ reporting—two areas that have driven the new SAS user crazy!!! With the added capabilities of ODS, PROC REPORT can look as sharp as an EXCEL report. This paper will show how to use PROC REPORT in both a windowing and non-windowing environment using SAS Version 9.

Graphing Made Easy with SG Procedures

Friday, 8:30-9:50
Conference Theatre

Susan Slaughter, Avocet Solutions

New with SAS® 9.2, ODS Graphics introduces a whole new way of generating graphs using SAS. With just a few lines of code, you can create a wide variety of high-quality graphs. This paper covers the three basic Statistical Graphics (SG) procedures—SGPLOT, SG PANEL and SGSCATTER. SGPLOT produces single-celled graphs. SG PANEL produces multi-celled graphs that share common axes. SGSCATTER produces multi-celled graphs that may use different axes. This paper shows how to use each of these procedures to produce different types of graphs, how to send your graphs to different ODS destinations, how to access individual graphs, and how to specify

properties of graphs, such as format, name, height and width.

PROC TABULATE: Getting Started

Friday, 10:00-11:20
Conference Theatre

Art Carpenter, CA Occidental Consultants

Although PROC TABULATE has been a part of Base SAS® for a very long time, this powerful analytical and reporting procedure is very under utilized. TABULATE is different; it's step statement structure is unlike any other procedure. Because the programmer who wishes to learn the procedure must essentially learn a new programming language, one with radically different statement structure than elsewhere within SAS, many do not make the effort.

The basic statements will be introduced, and more importantly the introduction will provide a strategy for learning the statement structure. The statement structure relies on building blocks that can be identified and learned individually and in concert with others. Learn how these building blocks form the structure of the statement, how they fit together, and how they are used to design and create the final report will be presented.

Once the foundation is laid a number of intermediate level techniques, options, and statements that provide the TABULATE programmer with a wide range of power and flexibility are presented. These techniques include those that are often underutilized even by the more experienced TABULATE programmer.

Health Outcomes & Healthcare Research Methodologies

Are You in Need of Validation? Psychometric Evaluation of Questionnaires Using SAS®

Thursday, 8:30-8:50
Warfield

Eric Elkin, ICON

Questionnaires should undergo evaluation of their psychometric properties before they are relied on for making decisions. Psychometric validation is used in various fields including education, psychology, and health research to define the properties of questionnaires (sometimes referred to as scales, surveys, or instruments). Several presentations at prior SAS user group meetings have focused on factor analysis and related topics in order to develop scale structure. Instead, this presentation will assume that the multi-item scale has already been developed. It may need validation as a brand new scale to support the hypothesized scale structure or it may be an existing scale which is being used in a new population. The examples are taken from health-related quality-of-life research and will follow the "Guidance for Industry" published by the FDA (Dec. 2009). The focus will be on the classical test theory approach to psychometric validation including internal consistency, test-retest, and inter-rater reliability; construct and known-groups validity; and responsiveness. The presentation will include tips and tricks for SAS code as well as interpreting and presenting results. It is intended for those who are new to this field or as a review for those who already work in it.

Use CDISC SDTM as a data middle-tier to streamline your SAS® infrastructure

Thursday, 9:00-9:20
Warfield

Kalyani Chilukuri, clinovo

In many life science companies, SAS is used to streamline biometric processes. Some of these activities include but are not restricted to Edit Checks, study metrics, patient profile, analysis data sets, TLGs for submission or even CDISC conversion. Unfortunately, we have a tendency to reinvent the wheels for each new study and adapt the existing code to match the new study's data structure. This process is often tedious and can lead to errors, as we do not fully understand the assumptions of the modified legacy SAS programs.

The convergence of open innovations such as the CDISC SDTM standard and the free SAS-based CDISC Express has brought a better alternative by providing an easy way for sponsor companies and CROs to develop an internal SDTM middle-tier. CDISC SDTM has provided a common data structure for raw clinical data while CDISC Express makes it drastically faster and cheaper to convert any clinical database to the SDTM format using free software. Using this paradigm, companies are able to fully embrace the concept of global standards for all their SAS processes, and leverage their entire standard reporting SAS assets with minimum programming. As CDISC Express is SAS-based, the most complex SDTM transformation can be done in one macro call and seamlessly integrated with internal SAS programs.

There are many benefits to this approach:

- Code re-usability ! No need for re-validation of standard SAS programs
- Better documentation, reliability of SAS standard study programs ! Reduced programming errors ! Faster study data-cleaning and analysis
- Easier FDA submission

This session will demonstrate this concept by showing in practice how to:

- Easily format your study to SDTM by using CDISC Express
- Re-run SAS Edit Checks, Enrollment graph, Patient Profile, ADaM and TLG for a new study with no SAS programming
- Quantify the savings generated with this new approach

Linkage of hospitalization records with time series data to define chronic disease events in a prospective epidemiologic study

Thursday, 9:30-9:50
Warfield

Nadia Chung, City of Hope

Hospital administrative databases provide a rich source of detailed information on a wide range of diagnoses, procedures, and clinical endpoints. Prospective epidemiologic studies are useful for assessing exposure-disease relationships, but identifying and verifying the multiple disease endpoints that occur among study participants during follow-up is costly and time-consuming. Direct linkage of hospital administrative databases with prospective epidemiologic data enables investigators to study temporal trends in disease incidence and mortality and assess potential changes in both risk factors and outcomes over time. However, resolving multiple records of hospitalizations among individual patients, assigning rules and priorities to different hospitalizations, and resolving conflicting information from successive patient visits in order to maintain consistent and high-quality endpoint definitions among these linked data can be challenging. We linked data from the prospective California Teachers Study (CTS) with data on in-patient hospitalizations from the California Office of Statewide Health Planning and Development (OSHPD), in addition to national and state mortality files to explore incidence, mortality, and risk- factors associated with stroke, which is one of the leading causes of death and disability in the United States. Using ICD-9-CM, Medicare Severity Diagnosis Related Group (MS-DRG) and Diagnosis Related Group (DRG) codes from OSHPD, and ICD-9 and ICD-10 codes from mortality databases, we identified over 3,000 stroke events among over 100,000 CTS participants from 1997 to 2008. This paper will describe our approach for handling the analytic challenges of classifying and analyzing these stroke endpoints after linking hospital administrative data with large prospective epidemiologic cohort studies.

A SAS® Macro Program to Calculate Medication Adherence Rate for Single and Multiple Medication Use

Thursday, 10:00-10:20
California
Warfield

Li-Hao Chu, Kaiser Permanent Southern

Different medication adherence measurements have been developed as a result of the complexity of medication consumption and different study designs. After an extensive review of methods applied in pharmacy related administrative databases, we found that Medication Possession Ratio (MPR) and Proportion Days Covered (PDC) are the two most common measurements used in pharmacoepidemiological and pharmaco-economic studies and are also recommended by the International Society for Pharmaco-economics and Outcomes Research (ISPOR).

This paper summarizes and presents a SAS macro program that covers different metrics for calculating medication adherence rate for single medication and multiple medication use. The results will allow clinicians and researchers to compare and select a metric that fits best to their research interest.

Estimating Sample Size through Simulations

Thursday, 10:30-10:50
Warfield

Wuchen Zhao, University of Southern California

Determining sample size is one critical and important procedure for designing an experiment. The sample size for most statistical models can be easily calculated by using the POWER procedure. However, the PROC POWER

cannot be used for a complicated statistical model. This paper reviews a more generalized method to estimate the sample size through a simulation approach by using SAS® software. The simulation approach not only applies to the simple but also to a more complex statistical design.

Using SAS Procedure Proc Mixed to Assess the Effects of School Health Policy Requirements on Change in Student Health Outcomes

Thursday, 11:00-11:20
Warfield

Lori Miller, University of Washington

The SAS® procedure Proc Mixed is a powerful yet flexible program for fitting complex hierarchical linear models. As such, a school effects and a school health policy effects model is generally designed for and statistically analyzed using hierarchically structured data sets in which individuals are nested within naturally occurring hierarchies (e.g., students within schools). However, using the SAS procedure Proc Mixed does not preclude the substantial data processing needed to prepare the data for analysis of a large secondary database that includes a racial/ethnically diverse student population of various socioeconomic levels. It is worth noting that the current thought among highly experienced SAS users is that the creation and integration of dummy variables for categorical variables used within a hierarchical linear model was not necessary and could not be executed using the SAS procedure Proc Mixed. Thus, the objective of this paper is to demonstrate both the process and difficulties of aptly structuring a custom data set for modeling and statistical analysis, how to create dummy coded variables for each represented student race/ethnicity and socioeconomic status with more than two categories, and how to integrate the dummy coded variables with the use of a chosen dummy coded reference group into the final mixed model procedure to enable smooth convergence of a two-level school health policy effects model. Incorporating these key data processing steps prior to the statistical analysis can expand the use of SAS procedure Proc Mixed for public policy, public health nursing, demographic, healthcare and, racial and ethnic health disparities research.

Causal Methods for Observational Data

Thursday, 11:30-11:50
Warfield

Amanda Stevenson, University of Texas at Austin

Comparative effectiveness research often uses non-experimental observational data (like hospital discharge records or nationally representative surveys) to draw causal inference about the effectiveness of interventions for health. These ex post inferences require the careful use of specialized statistical methods in order to account for issues like selection bias and unmeasured heterogeneity. This paper briefly discusses the strengths and weakness of some of the most common causal methods for comparative effectiveness evaluation and provides instructions for using SAS® to implement propensity score matching, double difference, instrumental variables methods and regression discontinuity.

Effects of Timing of Surgery in Children with Ureteropelvic Junction Obstruction: A Pediatric Health Information System (PHIS) Database Study using SAS®

Friday, 8:30-8:50
Campus
Warfield

Andrea Masias, University of Colorado Anschutz Medical

Ureteropelvic junction (UPJ) is an obstruction in the ureter that can lead to hydronephrosis which is a swelling of the kidney due to a lack of urine passage. It is the most common cause of congenital hydronephrosis and affects roughly 60,000 children annually¹. Pyeloplasty is a surgical intervention used to correct the condition. The increased number of antenatal ultrasounds being performed has resulted in an earlier diagnosis of the condition and in some cases earlier pyeloplasty. Based on previous research, we hypothesize that those who receive early pyeloplasty will require more readmissions to the hospital within one year following the initial pyeloplasty, will have longer lengths of hospital stay and will have higher overall adjusted hospital charges compared to those who receive delayed pyeloplasty. We will test these hypotheses using SAS® software version 9.2.

Clinical Data Integration for Observational Studies Using Electronic Health Records: Examples with Defining Pregnancy and Gestational Diabetes

Friday, 9:00-9:20
Institute
Warfield

Beinan Zhao, Palo Alto Medical Foundation Research

Electronic Health Records (EHRs) have increasing potential for secondary use in observational research studies, including studies of drug safety and comparative effectiveness of interventions. However, clinical data from different sources must be successfully integrated and thoughtfully prepared to create robust research definitions. As one step in using EHR data, it is often necessary to determine episodes of a health condition from multiple data sources. In this paper, we demonstrate one approach in defining such episodes with pregnancy and gestational diabetes as examples of how we integrate clinical data for observational studies using EHRs at a multi-clinic, outpatient healthcare system in Northern California.

Combining PROC SQL Summary Functions with Logical Expressions

Friday, 9:30-9:50
Warfield

Glenn Wright, UCSF

In SAS, a true expression such as "2+2=4" or "3>=0" resolves to the numeric value 1, and a false expression such as "2+2=5" resolves to the numeric value 0. Such expressions can be used as input for PROC SQL summary functions in order to create customized summary variables that draw upon multiple records. For example, one could use this technique on a database of healthcare claim lines in order to create client summary records showing which clients received a certain combination of services on at least one day during a certain time period. One can also combine these techniques with PROC SQL JOINS to create summary variables that combine data from multiple sources.

Analyze This? Supporting Clinical Decisions Graphically When Not Enough Data is Available A Study Case: Challenges in NCI CTCAE Version 4 Grading

Friday, 10:00-10:20
Roche Group
Warfield

Sheila Dayog, Genentech, A Member of the

Working with laboratory clinical data is challenging enough, but another challenge has arisen when applying the updated National Cancer Institute Common Terminology Criteria, Version 4.0 for Adverse Events (NCI CTCAE). Version 4.0 incorporates stronger clinical observation into the some of the laboratory grading descriptions, thus removing straight forward range-based numeric algorithms. Additionally, these observations can present themselves a number of ways and the task of programmatic laboratory grade assignment whilst considering these assessments offers challenges. Potential for inconsistent grading exists, possibly leading to incorrect clinical data interpretation. This paper will discuss these challenges, offering the use of SAS® graphics data visualization to assist in identifying any discordant relationships between laboratory grade assignments and the adverse event grades that occur throughout the conduct of an oncology clinical trial.

Clinical Data Review on iPad® with DEFINE.XML

Friday, 10:30-11:20
Warfield

Sy Truong, Meta-Xceed, Inc.

Clinical data review on a tablet computer is a natural step in the evolution of automated information delivery. Clinical Research Associates and Medical Monitors require a unique method of accessing information during the conduct of a clinical trial and also at the pivotal conclusion, or submission of a trial. The challenge remains to overcome the technical complexity of metadata models such as DEFINE.XML. This paper describes these challenges as it elaborates upon methods of using SAS to deliver clinical data in a format that can be easily navigated on an iPad.

Some of the methodologies described include:

- Capturing data with standardized metadata of DEFINE.XML
- Navigating to specific clinical domains with standardized data structures with multi-touch
- Searching for subject data such as Adverse Events on iPad
- Viewing clinical data while mobile and offline
- Zooming in on images such as CRF or graphs with pinch gestures

These technologies are new and will pave ways for tools in this "post-PC World". However, data review of clinical information remain as old as the drug development process.

Posters

SAS® DM Statements Unveiled

Wednesday, 2:00-2:50
Cypress A&B

Megha Agarwal, Clinovo

SAS programmers usually lose a lot of time (and temper) when setting up a correct environment before submitting a program. For example, every time you run a program, you need to close all the datasets that are created in the program, clear the log and output windows, save the program etc. Then once the program is run, you need to open a specific dataset to check the results. In a nutshell, you always follow some fixed, time-consuming processes, before and after executing the program.

Before knowing about the DM (Display Manager), I used to say, "if only all this could be done automatically". Then DM statements came to my rescue. A DM statement stands for Display Manager Statements. It submits SAS Program Editor, Log, Procedure Output or text editor commands as SAS statements. This paper will explain the purpose, functionalities and usage of DM statements using several examples.

The intended audience is all levels of SAS users.

Is Your SAS Library a Disk Hog? Here's How to Put it on a Diet

Wednesday, 2:00-2:50
Cypress A&B

Ross Bettinger, Modern Analytics

We have developed a set of SAS macros to deal with big SAS libraries that hog large amounts of disk space. We may put them on a diet by squeezing the unnecessary bytes out of them without losing any accuracy of numeric variables or the contents of character variables. Using these macros may result in significant reductions in disk space. In one case, a SAS library containing 177 datasets which required 44.9 GB was reduced to 21.3 GB, a reduction in disk space of 52.6%.

ChiD, A χ^2 -Based Discretization Algorithm

Wednesday, 2:00-2:50
Cypress A&B

Ross Bettinger, Modern Analytics

We have developed a discretization algorithm, based on Kerber's ChiMerge and Liu and Setiono's Chi2, that automatically chooses the best set of cutpoints for dividing a continuous variable into a set of contiguous discrete intervals. The algorithm, ChiD, uses class information to perform supervised discretization based on maximizing the logworth of the significance of a χ^2 statistic computed from adjacent intervals of the continuous variable being discretized. The ChiD algorithm generates cutpoints that match the quality of those computed by the Enterprise Miner Decision Tree algorithm as measured by the accuracy of classification models built using ChiD cutpoints versus original, undiscretized data.

Solve Eight Queens Puzzle with SAS® Macro

Wednesday, 2:00-2:50
Cypress A&B

Jian Dai, Clinovo

We demonstrate how to solve eight queens puzzle by using recursive SAS macro to implement the depth-first search algorithm together with backtracking.

Using Vovici API in SAS® 9.2

Wednesday, 2:00-2:50
Cypress A&B

Richard Koopmann, Capella University

With the release of SAS 9.2, programmers were able to communicate directly with third-party services via two new procedures (depending on the type of service being accessed): the HTTP procedure for Representational State

Transfer (REST) Protocol-based services, the SOAP procedure for Simple Object Access (SOA) Protocol-based services. Given these tools, SAS can interact with data on remote servers directly.

The current paper demonstrates how the SOAP procedure is being employed for more automated ETL processes from Vovici, a third-party survey hosting service, using the SOA Protocol. The returned xml-based survey results, when paired with the xml-based survey definition, are used as input for regular reporting processes.

A Taylor Series Based Linearization for Forecasting Time Series Data?

Wednesday, 2:00-2:50
Cypress A&B

Anpalaki Ragavan, UNiversity of Nevada, Reno

Forecasting time series data with time trend models assume permanent deterministic pattern across time hence are best suited for data that are not dominated by random fluctuations and are normally distributed. Many observed time series do not have constant, linear, or quadratic time trends as assumed in time trend models. However, one can compensate for the inadequacies by using Taylor series based linearization methods. Data, for which the joint distribution is difficult or impossible to obtain such as data with correlated errors, with a large number of and/or crossed random effects, and multiple types of subjects are excellent candidates for Taylor series linearization based fits, which, uses expansions to approximate a model, based on pseudo data with fewer nonlinear components. The process of computing the linear approximation is repeated several times before a suitable model is obtained, which is an iterative process. On convergence, the new parameter estimates are used to update the linearization, which results in an improved linear model. In this research the traditional time trend modeling approach is compared to a Taylor series based linearization approach to forecast time series data with large crossed random effects, non-linearity and correlated errors. Multiple time series data collected over a ten year period modeled using time trend models a compared to the Taylor series based linearization. The relationships among variables were identified through a series of steps during exploratory analysis. Trend models were fitted and data forecasted using FORECAST, ARIMA and UCM procedures and compared to a Taylor series linearization based model fitted using PROC GLIMMIX.

SAS® since 1976

Wednesday, 2:00-2:50
Cypress A&B

Zubair Shaik, Oklahoma State University

In the past four decades SAS® software has been extensively used by many businesses for effective decision making. With advancements in research and technology many researchers have proposed different ways of using SAS® from decade to decade. It is interesting to explore the trends in research topics for various industries in SUGI/SAS Global Forum conferences since 1976. We apply Text Mining using SAS® Enterprise Miner 6.2 to discover trends in the usage of SAS® tools in various industries via analyzing all abstracts published in SUGI/SAS Global Forum since 1976.

JMP® Scripting: Advanced Techniques

Wednesday, 2:00-2:50
Cypress A&B

Erin Vang, Global Pragmatica

An introduction to the JMP Scripting Language (JSL) by Erin Vang, author of the first JMP Scripting Guide (JMP version 4) and custom scripting consultant with Global Pragmatica (<http://globalpragmatica.com>), this talk will present an example of how to build a custom analysis with a nice point-and-click interface.

SAS Program as a Backup Tool for an Adaptive Randomization System

Wednesday, 2:00-2:50
Cypress A&B

Yajie Wang, VA Palo Alto HCS

The randomization of patients in a clinical trial depends on the continuous availability and normal functioning of the Internet and IVR (Interactive Voice Response) system. Interruptions to the availability of the Internet or the ability of the IVR system to pick up calls from trial sites require prompt action. This is especially troublesome for adaptive randomization because the new treatment assignment is not pre-coded and is based on information from all the previously randomized patients. A backup tool is presented here to allow the continuation of adaptive randomization

during interruptions of the Internet or the IVR system.

Resources

WUSS Conference Android App

Friday, 8:30-8:50
Curran

Sy Truong, Meta-Xceed, Inc.

As information transitions from hidden databases and obscure networks into mainstream social networks like Twitter, Facebook and Youtube, users leverage the social aspects of communication to make sense out of the deluge of data. Social media can provide extra context and add more meaning when it is recommended from a trusted friend as compared to spam. Attending WUSS SAS Conference in the past used to involve carrying a thick proceedings and a program book while deciphering what presentation is most relevant and useful for each attendee. As SAS conferences grow and the number of sections and papers become more diverse, this process becomes overwhelming. Since WUSS by its own nature is a social event, it lends itself to social networks and mobile apps as it helps individuals make sense and bring context to the presentations. An example is how Twitter has become more pervasive since it gives a new voice for expressing opinions and provide instant information to collectively capture the ethos of the conference as a community and records it for perpetuity.

The Android Market has recently surpassed the number of free Apps compared to Apple AppStore. Android smart phones and tablet computers empower users to obtain information when they are mobile. This information can be more meaningful when used in conjunction with mobile social media tools. This paper presents a quintessential use of these technologies for SAS users as they navigate SAS conference with the WUSS 2011 Android App.

Improving Your SAS® Skills and Expertise - Career Path Considerations

Friday, 9:00-9:20
Curran

Kirk Paul Lafler, Software Intelligence Corporation

Many SAS® users believe they have mastered the necessary skills to be successful. But as technology evolves, it becomes increasingly more important to continue learning, and consider specializing. Even if you already consider yourself a good SAS programmer or user, with interests and abilities in several areas of the SAS software, specialization is becoming more of a necessity. This presentation will explore the many options available to users for improving their skills and expertise.

This is how we do it: teaching SAS® in the Cal Poly Statistics Department

Friday, 9:30-9:50
Curran

Jimmy Doi, City of Hope / Cal Poly SLO

Providing students with a strong foundation in SAS programming is extremely valuable for those working towards a degree that involves data management or statistical analysis. This training is useful not only for students who will enter the job force right after graduation, but also for those who intend to pursue a graduate degree. Statisticians and data analysts with SAS programming skills are in high demand, and by preparing students with solid SAS programming, data analysis, and research skills they will be ready to fill real-world positions right after graduation. Our program at Cal Poly San Luis Obispo is one of many university-level programs that help provide such a foundation. Our success lies in our methods of teaching SAS to our undergraduates, and also in the resources we utilize that are available to students and faculty using SAS. This presentation will describe some of our experiences with SAS teaching methods and resources that we have employed to try to educate the next generations of SAS programmers and data analysts.

What is a SAS Mentor and why do I need one?

Friday, 10:00-10:20
Institute
Curran

Stanley Fogleman, Harvard Clinical Research

A SAS® mentor can be a valuable resource to junior programmers just starting out, or programmers who may have used other high level languages and are just starting to use SAS. The paper will focus on preparing a plan for a

hypothetical junior programmer, with the goal of becoming a more proficient programmer over a one or two year period. SAS, unlike some other languages does not lend itself to being self-taught and some things that might look like an obvious choice can actually lead to performance penalties. The most important thing that mentors can provide is guidance – for things worth studying more and things that might be interesting, but you might only use infrequently.

Top Ten SAS® Sites for Programmers: A Review

Friday, 10:30-10:50
Corporation
Curran

Kirk Paul Lafler, Software Intelligence

What are the top ten SAS® sites for users, beginning with sas.com and jmp.com. We then expand to sasCommunity.org, support.sas.com, and six other popular sites that assist you in training and programming. If you use Google or another search engine to search for content, you will get over a million hits. In this presentation we narrow the list down to ten.

At Your Service: Your Roadmap to Support from SAS®

Friday, 11:00-11:50
Curran

Kathy Council, SAS Institute

At your service. How do you make the most of the products and services from SAS® Publishing, SAS Education, SAS Technical Support, and support.sas.com? You have world-class support from SAS at your fingertips. But where do you start? How do you navigate the sea of information available to you from SAS? How do you find the resources you need to do your job? This paper will provide you with practical tips, tricks, and techniques to find exactly what you need to use SAS. You'll save time, learn about upcoming content and, best of all, become a more proficient and expert user of SAS.

This paper will benefit the new user and the seasoned SAS user alike. The intended audience is anyone interested in learning more about how to effectively use the services available from SAS.

SAS Essentials

SAS Essentials I: SAS® Functions for a Better Functioning Community

Thursday, 8:30-9:35
Fillmore A&B

AnnMaria De Mars, The Julia Group

Employers want work done by experienced programmers. New programmers want to gain experience. Communities want, but often can't afford, more people to analyze data and to explain those analyses in plain language. One way to break out of that vicious circle of "can't get a job without experience, can't get experience without a job" is to take advantage of the many opportunities to use freely available data to help inform your community. The beauty of it is that using these data sets can provide you the experiences that will advance you from being a novice programmer.

This paper uses open data to demonstrate how new SAS programmers can download, analyze and present results from large and complex data sets. In the process, examples are presented of the experiences programmers can gain in making the appropriate design choices, applying a broad range of functions and procedures.

Character functions of PUT, VVALUE, %LENGTH and INDEX can all be used for text processing. Analyses are presented demonstrating when to use which function and why.

PROC SQL, PROC TRANSPOSE, PROC MEANS, PROC FREQ and macro programming are used in new and old ways to get a new look at data in the Trends in Mathematics and Science Study to provide information teachers can use in their classrooms.

SAS Essentials II: Better-Looking SAS for a Better Community

Thursday, 9:45-10:50
Fillmore A&B

AnnMaria De Mars, The Julia Group

Experienced programmers don't just write code that runs, they also look professional. That doesn't refer to their designer wardrobe (a quick glance around at your co-workers probably told you that) but to their code, log and output. In this paper, better- looking and better-designed programs are demonstrated using PROC FREQ and macros. SAS system level options are examined for their effectiveness in producing better-looking logs and output. PROC FORMAT, PROC TABULATE, PROC PRINT and ODS are used to create better-looking reports. SAS/Graph and Graph-N-Go are used to make good-looking graphs.

Now here's the catch - where do you get the time and opportunity to try out new techniques? The examples used here are from projects done for various community service organizations, from sports organizations to public schools. In the end, your program looks good, your output looks good and you've improved both your programming skills and your community. Use your SAS skills to produce reports for your child's sports league and you'll never have to chaperone in the freezing cold again! And who says nice guys finish last?

SAS Essentials III: Statistics for Hamsters

Thursday, 11:00-11:50
Fillmore A&B

AnnMaria De Mars, The Julia Group

"For next year, I would like to see a workshop offered on statistics so easy a hamster can understand it. Bring your own hamster." - Workshop Attendee Here it is! No actual hamsters were involved, but the statistics in this session were all previously presented to many classes of middle school students (who sometimes have the attention span of hamsters). The examples use the national dataset and a 1% sample of California residents, data from the American Community Survey, downloaded from the U.S. Census website. Teachers suggested questions of interest to students, including employment, income and education by race and ethnicity. Statistics were produced using SAS. Graphics were created using JMP 8 and SAS 9.2. These results were incorporated as part of a lesson that used SAS output and graphics to illustrate the concepts of frequency distribution, histogram, mean, median, mode, pie charts, correlation, sample selection and group differences. This session is recommended not only for those interested in a refresher in basic statistics but also for anyone who would like to apply their SAS skills to supporting their local schools through volunteering as a guest speaker. Integrating SAS with the curriculum can show students applications of programming and statistics to the social studies, science and mathematics they are learning in schools and to issues important in their own lives. Try it! You may be overwhelmingly surprised by the welcome you get from teachers and students alike.

Tips and Techniques for the SAS® Programmer

Thursday, 2:30-3:20
Fillmore A&B

Helen Carey, Carey Consulting

If you are new to programming or even are an experienced programmer, you will benefit by learning tips and techniques that will help you be more productive. This presentation offers tips in programming, efficiency, work habits, and where to find answers to your SAS questions. These tips come from our own experience and from learning from others through their presentations and papers.

SAS\ Macro Programming for Beginners

Thursday, 3:30-4:20
Fillmore A&B

Susan Slaughter, Avocet Solutions

Macro programming is generally considered an advanced topic. But, while macros certainly can be challenging, it is also true that the basic concepts are not difficult to learn. This paper is designed for people who know the basics of SAS programming, but know nothing about SAS macro programming. We explain how the macro processor works, and how to use macros and macro variables. Using these techniques you can create flexible, reusable code that can save you time and effort.

Introduction to SAS® Programming

Thursday, 4:30-5:20
Fillmore A&B

Lily Xu, Ohlone College

Data is everywhere today and SAS skills are in high demand. This workshop is designed for those who have little or no previous programming experience in SAS.

From the workshop, you will learn how to:

- write simple SAS programs
- read raw data files and existing data sets
- create temporary and permanent SAS data sets
- generate simple statistical reports

You will also have the opportunity to gain some hands-on programming experience on basic SAS procedures.

Tutorials

An Introduction to Creating Multi-Sheet Microsoft Excel Workbooks the Easy Way with SAS®

Wednesday, 2:30-3:20
Sequoia

Vince DelGobbo, SAS Institute

Transferring SAS data and analytical results between SAS and Microsoft Excel can be difficult, especially when SAS is not installed on a Windows platform. This talk provides basic information on how to use Base SAS®9 software to create multi-sheet Microsoft Excel workbooks (for Excel versions 2002 and later). You will learn techniques for quickly and easily creating attractive, multi-sheet Excel workbooks that contain your SAS output using the ExcelXP ODS tagset. The techniques that are presented in this talk can be used regardless of the platform on which SAS software is installed. You can even use them on a mainframe! More in-depth information on this topic will also be presented, if time permits.

Tips and Tricks for SG Procedures and GTL for Clinical Graphs

Wednesday, 3:30-4:20
Sequoia

Sanjay Matange, SAS Institute

Did you know that you can create an adverse event graph using a vector plot? Or, that you can label dosage levels for a medications plot using a scatter plot? How do you place a reference line between two values on a category axis? Statistical graphics (SG) procedures and the graph template language (GTL) provide you myriad ways to mix and match statements to create graphs. What you can achieve is based on creative usage of the statements.

This presentation includes tips and tricks you can use in SG procedures and GTL programs to build your graphs. We use examples from clinical trials and health and life sciences domains to illustrate the techniques using real- world graphs like LFT panels, patient profiles, adverse event plots, and more.

Most examples use the second maintenance release of SAS® 9.2, but this presentation also includes a sneak preview of some powerful new features to be released with SAS® 9.3.

Creating Statistical Graphics with ODS in SAS® Software

Wednesday, 4:30-6:10
Sequoia

Warren Kuhfeld, SAS Institute

Effective graphics are indispensable in modern statistical analysis. SAS 9.2 provides ODS Graphics, new functionality used by statistical procedures to create statistical graphics as automatically as they create tables. ODS Graphics is also used by new SAS/GRAPH® procedures that are designed for graphical exploration of data. This tutorial is intended for statistical users and covers the use of ODS Graphics from start to finish in statistical analysis. You will learn how to:

- Request graphs created by statistical procedures.
- Use the new SGPLOT, SGPANEL, SGSCATTER, and SGRENDER procedures in SAS/GRAPH to create customized graphs.
- Access and manage your graphs for inclusion in Web pages, papers, and presentations.
- Modify graph styles (colors, fonts, and general appearance).

- Make immediate changes to your graphs using a point-and-click editor.
- Make permanent changes to your graphs with template changes.
- Specify other options related to ODS Graphics

Getting Freqy with PROC REPORT

Thursday, 8:30-9:20

Ethan Miller, SRI International

Sequoia

Are you unfamiliar with PROC REPORT or do you feel a bit overwhelmed by it sometimes? Are you familiar with PROC REPORT but want to take it to the next level? This paper is for you! With just a brief investment of time, you too will be able to incorporate PROC REPORT into your SAS® repertoire. This paper will illustrate how to produce frequencies, cross tabulations, and means using PROC REPORT, and show various style options that can be used with ODS. It will also show how to (1) create totals using compute blocks; (2) format the report; (3) add titles, subscripts, superscripts, and custom footnotes that include statistics; and (4) report missing data in different ways.

Powerful and Sometimes Hard-to-find PROC SQL® Features

Thursday, 9:30-10:20

Kirk Paul Lafler, Software Intelligence

Corporation

Sequoia

The SQL Procedure contains many powerful and elegant language features for SAS® users in the know. This presentation explores topics that will help users unlock the many features, options, and other hard-to-find gems found in the SQL universe. Topics include querying and subsetting data; conditional processing and restructuring data by using case expressions; constructing and using virtual tables known as a view; access information from Dictionary tables; joining two or more tables to explore data relationships; and query performance tuning

Understanding the Anatomy of a SAS® Deployment: What's in My Server Soup?

Thursday, 10:30-11:20

Connie Robison, SAS Institute

Sequoia

Did you want to better understand the pieces of a SAS metadata-based deployment and where they go and what they do? This paper highlights the major components of a metadata-based SAS deployment of solutions and business intelligence. Using the concepts of topology and tiers as a framework, this paper illustrates the location of key components and how they can be distributed across machines.

An Introduction to SAS® Character Functions

Thursday, 2:30-3:20

Ronald Cody, self-employed

Sequoia

SAS® software is especially rich in its assortment of functions that deal with character data. This class of functions is sometimes called STRING functions. With over 30 new character functions in SAS® 9, the power of SAS to manipulate character data is even more impressive.

Some of the functions we will discuss are: LENGTH, SUBSTR, COMPBL, COMPRESS, VERIFY, INPUT, PUT, TRANWRD, SCAN, TRIM, UPCASE, LOWCASE, || (concatenation), INDEX, INDEXC, AND SPEDIS. Some of the new and exciting SAS® 9 functions that we will cover are the "ANY" and "NOT" functions, the concatenation functions (and call routines), COMPARE, INDEXW, LENGTHC, PROPCASE, STRIP, COUNT, and COUNTC.

Power Analysis: What is Available and What You Need to Know

Thursday, 3:30-4:20

Gerry Hobbs, West Virginia University

Sequoia

Planning should be the first step in any scientific experimentation. A critical part of planning is the determination of the sample size necessary to achieve the goals of the particular project. It is clearly a waste of resources to have too many experimental units (subjects, animals, ingots or whatever). It is equally a waste of resources to carry out an

investigation with so few experimental units that there is little hope of demonstrating whatever effect it is that we wish to demonstrate.

SAS® Software provides tools that can be used in order to guide us in our effort to, at least, estimate sample size requirements. It is not always understood that there is a considerable amount of information that is required for power and sample size calculations done by hand (perish the thought). No less information is needed if we want software do the calculations for us.

The purpose of this tutorial presentation is to discuss the information required, a priori, in order to calculate sample size/power and to discuss the implementation of those calculations in the SAS environment.

Becoming a Better Programmer with SAS® Enterprise Guide® 4.3

Thursday, 4:30-5:20

Andy Ravenna, SAS Institute

Sequoia

Both existing and new users of SAS® are turning to SAS Enterprise Guide to write and run their code. Long-time users are accustomed to typing all their code in the Program Editor window and simply clicking Submit. New users do not have this same set of expectations and are more willing to point and click on occasion. But the truth is becoming clear; the winning programmer will be the one who has the expertise to use the best of both worlds—either coding or clicking, depending on which is more efficient for a given task.

SAS Enterprise Guide 4.3 contains new functionality that can help anyone become a better programmer. This paper addresses the all-important question: When is it appropriate to code, and when to click? The purpose here is to expose new users as well as those familiar with SAS to tips and best practices that will enable them to return to the office as better programmers.

SAS Macros for the Working Statistician

Thursday, 5:30-6:20

William Krebs, Self-employed

Sequoia

In the conventional wisdom of the pharmaceutical industry, the SAS Macro facility is regarded as a data processing tool, not as a statistical tool in its own right. In contrast to that perception, this paper presents simple SAS Macros that implement advanced statistical techniques not provided in the SAS System. Examples include Fieller's theorem, errors-in-variables regression, and robust regression methods.

HELP, My SAS® Program isn't Working: Where to Turn When You Need Help

Friday, 8:30-9:20

Kimberly LeBouton, KJL Computing

Sequoia

Instead of a quick code fix, I often assist the SAS user by troubleshooting their issue with my knowledge of SAS and the SAS community. For over 20 years, I have provided SAS Technical Support, and this paper will present strategies I have used to work through simple to complex technical problems, and will include access to my "cheat sheets".

Up To Speed With Categorical Data Analysis

Friday, 9:30-10:20

Maura Stokes, SAS Institute

Sequoia

Categorical data analysis remains a substantial tool in the practice of statistics, and its techniques continue to evolve with time. This paper reviews some of the basic tenets of categorical data analysis today and describes the newer techniques that have become common practice. The use of exact methods has expanded, including additional assessments of statistical hypotheses, conditional logistic regression, and Poisson regression. Bayesian methods are now available for logistic regression and Poisson regression, and graphics are a regular component of many analyses.

This paper describes recent techniques and illustrates them with examples that use SAS/STAT® software.

Statistical Analyses Using SAS® Enterprise Guide®

Friday, 10:30-11:20

R. Scott Leslie, MedImpact Healthcare Systems,

Inc.

Sequoia

Conducting statistical analyses involves choosing proper methods, understanding model assumptions and displaying clear results. The latest releases of SAS® Enterprise Guide® offer conveniences, such as point-and-click wizards and integrated syntax help, to ease the burden on users. This tutorial demonstrates how to perform statistics quickly and easily using some handy features of Enterprise Guide. Examples of multiple linear regression, logistic regression and survival analysis are covered as well as some hints on how to navigate Enterprise Guide menus. This tutorial is intended for SAS users with beginning to intermediate experience with the above mentioned statistics or those with less experience with Enterprise Guide.