

**Boston Area SAS® Users Group Announcement
SAS Blowout Event!
Co-hosted by BASUG and SAS Institute Inc.**

Friday, September 29, 2017

Look - It's a Quarterly Meeting!

It's a Mini-Conference!


It's BASUG's Annual SAS Blowout!

Choose from 7 Talks by SAS/STAT Developers !

Networking Lunch !

Ice Cream Social !

So many presentations, so little time. That's why we're offering concurrent sessions in the afternoon. Come for all or part of the day, and be sure to stay for our make-your-own sundae social event, where you can schmooze with your colleagues and this incredible team of presenters from SAS.

When	Friday, September 29, 2017 8:15AM – 4:00 PM
Where	Microsoft New England Research and Development Center (NERD) ¹ One Memorial Drive Conference Center, First Floor Cambridge, MA 02142 857-453-6000 
Directions	Please visit the meeting site directions page
How to register	Individual, on-line registration required. Please visit the event registration page NO WALK-INS!
Payment	\$20 – if paid on-line by Tuesday, September 19 th \$25 – if paid on-line by NOON on Thursday September 28 th \$35 – at-the-door (checks only)
Contact	If you have questions about the meeting, contact the meeting organizers: Jean Roth, Rita Volya, Quentin McMullen, and Paul Grant at: the meeting organizers:

Agenda*

8:15 AM	Sign in and Refreshments	
8:45 AM	Announcements	
9:00 AM	<u>An Introduction to Bayesian Analysis with SAS/STAT® Software</u> <u>Maura Stokes, SAS</u>	
9:50 AM	Break	
10:00 AM	<u>Step Up Your Statistical Practice with Today's SAS/STAT Software</u> <u>Robert Rodriguez, SAS</u>	
10:55 AM	Break	
11:05 AM	<u>Best Practices for Machine Learning Applications</u> <u>Brett Wujek, SAS</u>	
12:00 PM	Networking Lunch	
	Breakout Room #1	Breakout Room #2
1:00 PM	<u>An Overview of SAS Visual Data Mining and Machine Learning on SAS Viya</u> <u>Brett Wujek, SAS</u>	<u>Fitting Your Favorite Mixed Models with PROC MCMC</u> <u>Fang Chen, SAS</u>
1:50 PM	Break	Break
2:05 PM	<u>Propensity Score Methods for Causal Interference with the PSMATCH Procedure</u> <u>Maura Stokes, SAS</u>	<u>Advanced Hierarchical Modeling with the MCMC Procedure</u> <u>Fang Chen, SAS</u>
2:55 PM	Closing Remarks	
3:00 PM	Ice Cream Social !!!	

**Note: Times (and sequence) are approximate and subject to change. Please re-visit the BASUG website (www.basug.org) for updated information.*

Speaker biographies and abstracts

An Introduction to Bayesian Analysis with SAS/STAT® Software

Maura Stokes

The use of Bayesian methods has become increasingly popular in modern statistical analysis, with applications in numerous scientific fields. In recent releases, SAS® has provided a wealth of tools for Bayesian analysis, with convenient access through several popular procedures as well as the MCMC procedure, which is designed for general Bayesian modeling. This paper introduces the principles of Bayesian inference and reviews the steps in a Bayesian analysis. It then describes the built-in Bayesian capabilities provided in SAS/STAT®, which became available for all platforms with SAS/STAT 9.3, with examples from the GENMOD and PHREG procedures. How to specify prior distributions, evaluate convergence diagnostics, and interpret the posterior summary statistics is discussed.

Propensity Score Methods for Causal Inference with the PSMATCH Procedure

Maura Stokes

In a randomized study, subjects are randomly assigned to either a treated group or a control group. Random assignment ensures that the distribution of the covariates is the same in both groups and that the treatment effect can be estimated by directly comparing the outcomes for the subjects in the two groups. In contrast, subjects in an observational study are not randomly assigned. In order to establish causal interpretations of the treatment effects in observational studies, special statistical approaches that adjust for the covariate confounding are required to obtain unbiased estimation of causal treatment effects. One strategy for correctly estimating the treatment effect is based on the propensity score, which is the conditional probability of the treatment assignment given the observed covariates. Prior to the analysis, you use propensity scores to adjust the data by weighting observations, stratifying subjects that have similar propensity scores, or matching treated subjects to control subjects. This paper reviews propensity score methods for causal inference and introduces the PSMATCH procedure, which is new in SAS/STAT® 14.2. The procedure provides methods of weighting, stratification, and matching. Matching methods include greedy matching, matching with replacement, and optimal matching. The procedure assesses covariate balance by comparing distributions between the adjusted treated and control group.



Maura Stokes is Senior R & D Director of the Statistical Applications Department in the Advanced Analytics Division at SAS Institute. She received her DrPH from the Department of Biostatistics at the University of North Carolina in 1986. A SAS user since 1977, she joined SAS in 1985. Lead author of *Categorical Data Analysis with the SAS System*, which she wrote with Gary Koch and Charles Davis, Maura has taught workshops and tutorials on applied statistical topics for many years. Maura is a Fellow of the American Statistical Association.

[Return to Agenda](#)

[Return to Top](#)

Step Up Your Statistical Practice with Today's SAS/STAT® Software

Robert Rodriguez

Has the rapid pace of SAS/STAT® releases left you unaware of powerful enhancements that could make a difference in your work? Are you still using PROC REG rather than PROC GLMSELECT to build regression models? Do you understand how the GENMOD procedure compares with the newer GEE and HPGENSELECT procedures? When should you turn to PROC ICPHREG rather than PROC PHREG for survival modeling? This paper will increase your awareness of modern tools in SAS/STAT by providing high-level comparisons with well-established tools and explaining the benefits of enhancements and new procedures. The paper focuses on new tools in the areas of regression model building, generalized linear

models, survival analysis, and mixed models. When you see the advantages of these tools, you will want to put them into practice.



Robert N. Rodriguez is a senior director in SAS Research & Development with responsibility for the development of statistical software, including SAS/STAT and SAS/QC. He received his PhD in statistics from the University of North Carolina at Chapel Hill and worked as a research statistician at General Motors Research Laboratories before joining SAS in 1983. Bob is a Fellow of the American Statistical Association and served as ASA president in 2012.

[Return to Agenda](#)
[Return to Top](#)

Best Practices for Machine Learning Applications

Brett Wujek

Building representative machine learning models that generalize well on future data requires careful consideration both of the data at hand and of assumptions about the various available training algorithms. Data are rarely in an ideal form that enables algorithms to train effectively. Some algorithms are designed to account for important considerations such as variable selection and handling of missing values, whereas other algorithms require additional preprocessing of the data or appropriate tweaking of the algorithm options. Ultimate evaluation of a model's quality requires appropriate selection and interpretation of an assessment criterion that is meaningful for the given problem. This paper discusses many of the most common issues faced by machine learning practitioners and provides guidance for using these powerful algorithms to build effective models.

An Overview of SAS® Visual Data Mining and Machine Learning on SAS® Viya

Brett Wujek

Machine learning is in high demand. Whether you are a citizen data scientist who wants to work interactively or you are a hands-on data scientist who wants to code, you have access to the latest analytic techniques with SAS® Visual Data Mining and Machine Learning on SAS® Viya. This offering surfaces in-memory machine-learning techniques such as gradient boosting, factorization machines, neural networks, and much more through its interactive visual interface, SAS® Studio tasks, procedures, and a Python client. Learn about this multi-faceted new product and see it in action.



Dr. Brett Wujek is a Senior Data Scientist with the Enterprise Miner R&D team in the SAS Advanced Analytics division. He helps evangelize and guide the direction of advanced analytics development at SAS, particularly in the areas of machine learning and data mining. His formal background is in design optimization methodologies, receiving his PhD from the University of Notre Dame for his work developing efficient algorithms for multidisciplinary design optimization.

Check out this interview with Brett, discussing [SAS Machine Learning](#).

[Return to Agenda](#)
[Return to Top](#)

Fitting Your Favorite Mixed Models With PROC MCMC

Fang Chen

The popular MIXED, GLIMMIX, and NLMIXED procedures in SAS/STAT® fit linear, generalized linear, and nonlinear mixed models, respectively. These procedures take the classical approach of maximizing the likelihood function to estimate model parameters, using methods such as maximum likelihood and restricted maximum likelihood. The flexible MCMC procedure in SAS/STAT® can fit these same models by taking a Bayesian approach. Instead of maximizing the likelihood function, PROC MCMC draws samples (using a variety of sampling algorithms) to approximate the posterior distributions of model parameters, which is the key to Bayesian analysis. Similar to the mixed modeling procedures, PROC MCMC provides estimation, inference and prediction. This paper describes how to use the MCMC procedure to fit Bayesian mixed models and compares the Bayesian approach to how the classical models would be fit with the familiar mixed modeling procedures. The paper also discusses unique aspects of the Bayesian approach that are not related to the classical approach. Several examples illustrate the approach in practice.

Advanced Hierarchical Modeling with the MCMC Procedure

Fang Chen

Hierarchical models, also known as random-effects models, are widely used for data that consist of collections of units and are hierarchically structured. Bayesian methods offer flexibility in modeling assumptions that enable you to develop models that capture the complex nature of real-world data. These flexible modeling techniques include choice of likelihood functions or prior distributions, regression structure, multiple levels of observational units, and so on. This paper shows how you can fit these complex, multilevel hierarchical models by using the MCMC procedure in SAS/STAT® software. PROC MCMC easily handles models that go beyond the single-level random-effects model, which typically assumes the normal distribution for the random effects and estimates regression coefficients. This paper shows how you can use PROC MCMC to fit hierarchical models that have varying degrees of complexity, from frequently encountered conditional independent models to more involved cases of modeling intricate interdependence. Examples include multilevel models for single and multiple outcomes, nested and non-nested models, autoregressive models, and Cox regression models with frailty. Also discussed are repeated measurement models, latent class models, spatial models, and models with nonnormal random-effects prior distributions.



Fang Chen is a Senior Manager of Bayesian Statistical Modeling in Advanced Analytics Division at SAS. Among his responsibilities are development of Bayesian analysis software and the MCMC procedure. Prior to joining SAS, he received a PhD in statistics from Carnegie Mellon University in 2004

[Return to Agenda](#)
[Return to Top](#)

BASUG Contacts

Mailing Address:

BASUG
PO Box 170253
Boston, MA 02117

Email the [BASUG Webmaster](#)

- (1) The Microsoft New England Research & Development Center (NERD) is a research and software innovation campus located in the heart of Cambridge, Massachusetts. The NERD vertical campus spans two buildings with its primary presence and conference center located at One Memorial Drive and a recently renovated and expanded space located at One Cambridge Center. NERD is home to some of Microsoft's most strategic teams including Microsoft Research New England, Microsoft Application Virtualization (App-V), SharePoint Workspace, Microsoft Technical Computing, Microsoft Advertising, Microsoft Lync, Microsoft Office 365 and more. NERD has become a hub of activity for the local tech community and has hosted more than 500 events and welcomed more than 40,000 visitors during the past two years.