

Short Courses of ISCA 2019 Applied Statistics Symposium

SC01: A Variety of Mixed Models

8:30am-12:30pm, June 9, 2019

David A. Dickey, North Carolina State University

Abstract:

Mixed models are those with fixed and random effects. In ordinary mixed models, one estimates the fixed effects using estimated generalized least squares where the variance-covariance matrix of the data is estimated as part of maximum likelihood or REML (Restricted, or Residual, Maximum Likelihood) algorithm. After reviewing how to distinguish random from fixed effects, this course will describe the overall methodology and show several examples of its application including random coefficient models, repeated measures and hierarchical models. A review of nonlinear models is included and the additional complexities arising from the inclusion of random effects illustrated. A third type of model, the generalized linear mixed model, is discussed with examples. Such a model arises when the response is not normally distributed but rather is in the exponential family of distributions. Outstanding examples of the exponential family are the binomial and Poisson distributions. Emphasis is on concepts, examples, when to apply each type of model, and how to interpret each. Examples use SAS™ but the ideas presented are independent of software.

™ SAS is the registered trademark of SAS Institute, Cary, NC

About the instructor:

Dr. David A. Dickey is W. N. Reynolds Professor (emeritus) of Statistics at North Carolina State University. He is known for the Dickey-Fuller test for unit roots in time series. He is a Fellow of the American Statistical Association. He has spoken at the ASA's JSM, ASQ, and CSP meetings and many times at SAS Global Forum and regional SAS Users' Group meetings. Dickey has co-authored several books and dozens of papers. He was major advisor to 16 PhD students at NCSU and served on hundreds of graduate student committees across campus. Dickey is a member of NCSU's Academy of Outstanding Teachers and Academy of Outstanding Faculty Engaged in Extension. He received the D.D. Mason Faculty Award in 1986 and 2018 and the Outstanding Extension Service Award in 2007. Dickey was a founding faculty member of NCSU's Institute for Advanced Analytics, holds an associate appointment in Economics, and is a member of the Financial Math faculty. He taught at Randolph Macon College and The College of William and Mary in Virginia for 3 years before earning his PhD in 1976 under Wayne Fuller and spending the next 43 year at NC State.

SC02: Longitudinal Data Analysis and Latent Growth Curve Modelling in Public Health

1:30pm-5:30pm, June 9, 2019

Din Chen, University of North Carolina at Chapel Hill

Abstract:

Longitudinal data are very commonly collected by several time points which produced longitudinal data to evaluate the growth curve. However, the longitudinal data are temporally correlated which directly violate the fundamental assumption of independence in typically regression modelling and therefore erroneous conclusions and recommendations could be made if the data are not analyzed appropriately. Therefore, an appropriate analysis of longitudinal data is important to capture the intra-individual growth changes and inter-individual variabilities. This workshop is then designed to show how to do longitudinal data analysis using R package “*nlme/lmerTest*” and latent growth curve modelling using *Mplus*.

This workshop aims to address the longitudinal data analysis. We will review the classical longitudinal data analysis methods from multi-level and hierarchical modelling to test for within-individual longitudinal change and between-individual variability. Then a latent growth curve modelling will be introduced to analyze multi-domain longitudinal data using *Mplus*. Real data on study of 405 Hong Kong Chinese women who underwent cancer surgery will be used as a real example in the class to model the evidence of rate change in their mood and social adjustment at 1, 4, and 8 months post-surgery.

About the instructor:

Dr. Din Chen is a Fellow of ASA. He is now the Wallace H. Kuralt distinguished professor in biostatistics, University of North Carolina at Chapel Hill. He was a professor in biostatistics at the University of Rochester and the Karl E. Peace endowed eminent scholar chair in biostatistics at Georgia Southern University. Professor Chen is also a senior statistics consultant for biopharmaceuticals and government agencies with extensive expertise in clinical trials and bioinformatics. He has more than 150 referred professional publications and co-authored/co-edited 23 books on randomized clinical trials, statistical meta-analysis, public health statistical methods, causal inferences and statistical Monte-Carlo simulation and public health applications. Professor Chen has more than 30 years teaching and research experience in statistics/biostatistics. He has been invited to give short courses at JSM, CSP, ICSA, Biopharmaceutical Applied Statistics Symposium, Deming Conferences. He was honored with the "Award of Recognition" by the Deming Conference Committee for highly successful biostatistics workshop tutorials. This course on latent growth curve modelling will be of interest to even broad audiences.

SC03: Applied Nonlinear Modelling

8:30am-12:30pm, June 9, 2019

Tim O'Brien, Loyola University

Abstract:

Researchers often recognize that nonlinear regression models are more applicable for modelling their physical and medical processes than are linear ones for several important reasons. Nonlinear models usually fit their data well and often in a more parsimonious manner (typically with far fewer model parameters). Also, nonlinear models and the corresponding model parameters are usually more scientifically meaningful. As such, this short course focuses on these methods using a series of key illustrations accompanied by appropriate software analysis. Specifically, we cover linear regression and generalized linear models (e.g. logistic regression), Gaussian nonlinear models, and generalized nonlinear models, focusing on applications. Illustrations are given from bioassay, relative potency and drug or similar compound synergy useful in biomedical (including pharmacokinetics) and environmental sciences (including toxicology). Caveats are discussed regarding convergence, diagnostics, and the inadequacy of standard (Wald) confidence intervals – the intervals provided by several software packages. Extensions to bivariate situations (e.g., for both efficacy and safety of drugs) and censored (survival) analysis are also provided, as are implications for experimental design. Implementation using the SAS and R statistical software packages is discussed and illustrated.

About the instructor:

Dr. Tim O'Brien is a tenured professor and graduate director at Loyola University Chicago. He is a four-time Fulbright awardee, Loyola Master Teacher awardee, Best Paper awardee, and 2011-12 Jefferson Science Fellow (US State Dept., USAID), and a member of ASA, ENAR, RSS (elected) and ISI (elected). His PhD dissertation (NCSU, 1993) and research focuses on robust optimal design, generalized and normal nonlinear modelling especially in assessing potency and synergies, and predictive analytics. Dr. O'Brien has published over 50 research articles and book chapters and engages in extensive statistical consulting in industry (Amgen, BMS, BASF, Chiron, Glaxo, J&J, Janssen, Novartis), research institutes (IDI, INRA, INSERM, NIH, USAID), national and international universities, and various I/NGO's (Battelle, PHPT). Chosen as an American Statistical Association Travelling Short Course in 2005, this course has been given 24 times in 14 countries to over 730 participants. Dr. O'Brien teaches statistics courses (at Loyola and as visiting professor worldwide) spanning the spectrum from introductory concept-based courses to PhD-level theoretical courses in experimental design and nonlinear regression. His teaching experience started with his two-year Peace Corps service as a mathematics instructor in French West Africa.

SC04: Quantile Regression in Practice

1:30pm-5:30pm, June 9, 2019

Yonggang Yao, SAS Institute Inc.

Abstract:

If you ever worry about the validity of the common variance or other parametric distribution assumptions for your data analysis, quantile regression might be a relief for you. Quantile regression is a modern statistical methodology for modeling quantiles of a response variable conditional on explanatory covariates. Whereas linear regression models the conditional mean, quantile regression enables you to more fully explore your data by modeling the conditional quantiles, such as the median and the 5th and 95th percentiles. By fitting quantile regression models in a quantile-level range (called quantile process regression), you can furthermore estimate the conditional distribution of your response. Quantile process regression can be used to estimate the counterfactual distribution of a treatment group by using the quantile process model of a control group. Therefore, quantile regression is useful for health science, economics, engineering, and many other fields.

From the basic concepts and comparison to linear regression to more advanced applications and research topics, this tutorial demonstrates the benefits and potentials of using quantile regression methods and introduces computing tools for quantile model fitting, quantile predictions, conditional distribution estimation, conditional percentage estimation, and other inferences and hypothesis testing. The attendees are assumed to be familiar with basic probability distributions, linear algebra, and linear regression.

About the instructor:

Yonggang Yao is a Principal Research Statistician Developer at SAS Institute Inc. He joined SAS in 2008 after obtaining his PhD in statistics from The Ohio State University. Dr. Yao has developed several SAS quantile-regression procedures including PROC QUANTSELECT, PROC HPQUANTSELECT, and PROC QTRSELECT for standard and distributed computing environments. He is also the key supporting developer for two other SAS procedures: PROC QUANTREG for quantile regression and PROC ROBUSTREG for robust regression. Dr. Yao has taught tutorials on quantile regression at SAS Global Forums, SAS User Group Meetings, the 2015 Joint Statistical Meetings, and for the 2017 ASA traveling courses.

SC05: Re-Sampling: Theory, Uses, Novel Uses, and Mis-uses

8:30am-12:30pm, June 9, 2019

Richard Simon, formerly National Cancer Institute and currently R Simon Consulting

Abstract:

The fields of biology and medicine have seen the introduction of modern technologies for measuring biological processes within cells. Instead of having to develop tailored assays for each gene or protein of interest, the new assays provide genome-wide information about gene expression, epigenetic silencing as well as point mutations and copy number genomic variations. These development in technology have led to new biomedical problems and the development of new statistical methods for the analysis of $p > n$ data where the number of variables is greater than the number of cases. Re-sampling methods have seen increased use as replacements for model-based methods which may give misleading results in such circumstances. It has also resulted in new emphases on classification and prediction and re-sampling methods have found new uses for the evaluation of prognostic and predictive classifiers. In this half day course, we will explore a variety of uses of re-sampling including some novel uses for model evaluation. We will start with a review of some theory for re-sampling methods and addressing some views on the validity of such methods.

This half day course is intended for biostatisticians as most examples and novel uses will be from that area. No specific pre-requisite is required except some experience with R. The teaching plan will include lectures, and the illustration of worked examples using software distributed to the participants.

About the instructor:

Dr. Richard Simon retired recently from his fourth-years' service at the National Cancer Institute, serving most recently as Director of the Biometric Research Program and Chief of the Computational and Systems Biology Branch. He now has his own consulting practice: <http://rsimon.us>. He is the author of the book: "*Genomic Clinical Trials and Predictive Medicine*." He also coauthored two other books: "*Design and Analysis of DNA Microarray Investigations*" and "*Design and Analysis of Clinical Trials for Predictive Medicine*". He is a former member of the FDA Oncologic Drug Advisory Committee and a Fellow of the American Statistical Association. He is the recipient of the 2013 Karl Peace award "for contributions that have played a pivotal role in bridging the gap among statistics, clinical research, and translational medicine to improve human health" and the recipient of the 2017 Marvin Zelen award for leadership in Statistical Science from the Department of Biostatistics at Harvard University.

SC06: Causal Analysis of Observational Data Using SAS/STAT® Software

1:30pm-5:30pm, June 9, 2019

Michael Lamm and Clay Thompson, SAS Institute Inc.

Abstract:

The analysis of data from observational studies is an increasingly common task faced by data scientists and applied statisticians. Like analyzing data from randomized experiments, drawing causal inferences from observational studies requires careful consideration of the underlying data generating process and research design. Indeed, because of the presence of confounding variables, these requirements are even more critical for observational studies. Hence, specialized analytic tools and estimation methods have been developed for estimating causal effects from observational data. This presentation demonstrates the uses of these tools and methods.

This tutorial begins by reviewing the theory of causal effect estimation and discussing the difficulties that arise in the absence of randomization. It then introduces propensity score matching, inverse probability weighting, and doubly robust methods for estimating a treatment effect from observational data and explores the use of directed causal graphs to represent the data generating process. Causal graph analysis identifies sources of association and bias and evaluates whether a study design supports the estimation of a causal effect with a valid causal interpretation. Several examples, worked out in detail, illustrate approaches for constructing and evaluating the underlying models, comparing the estimation methods, and examining the assumptions required for the identification and estimation of treatment effects. Each example emphasizes elements of a rigorous and comprehensive workflow for causal analysis from observational or imperfectly randomized studies. The examples in this tutorial use SAS/STAT® software. No prior experience with these estimation and graphical methods is assumed.

About the instructors:

Dr. Michael Lamm is a Senior Research Statistician Developer in the Advanced Regression Research Department at SAS Institute. Among his responsibilities are the development of software for longitudinal data analysis and causal inference. Before coming to SAS, he received his PhD in statistics and operations research from the University of North Carolina at Chapel Hill.

Dr. Clay Thompson is a Senior Research Statistician Developer in the Multivariate Models Research Department at SAS Institute, where he develops algorithms and software for the analysis of causal effects using graphical models. Before joining SAS, he worked as a quantitative systems pharmacologist in the pharmaceutical industry. He received a PhD in applied mathematics from North Carolina State University.

SC07: Advanced Topics in Propensity Score Methods

8:30am-12:30pm, June 9, 2019

Wei Pan, Duke University

Abstract:

Propensity score methods have become a widespread practice in observational studies to reduce selection bias. However, researchers still often encounter in-depth methodological challenges in practice. For example, how to check assumptions? How to select covariates? How to deal with multiple treatments? How to deal with complex data such as clustered, longitudinal, and survey data? How to conduct outcome analysis using propensity score methods? How to conduct sensitivity analysis after applying propensity score methods? This course will build upon the fundamental principle and concept of propensity score methods to tackle these in-depth methodological challenges along with relevant R packages, such as *twang*, *Zelig*, and *rbounds*. Through lectures on the advanced topics in propensity score methods and hands-on activities for using the R packages, this course will help researchers better understand propensity score methods and, therefore, improve the validity of their observational studies. Instructions for downloading and installing the R packages with examples of real-world data will be provided to participants in advance through a course website. A working knowledge of propensity score methods is required. Participants are encouraged to bring their own computers for hands-on activities on the data provided in the course, and they are also welcomed to work on their own real-world data.

About the instructor:

Dr. Wei Pan is an Associate Professor of Research Design and Statistics at the Duke University School of Nursing. His research work focuses on causal inference (propensity score methods, resampling, power and effect size), advanced statistical modeling (multilevel, structural, and longitudinal), meta-analysis, psychometrics, and their applications in the social, behavioral, and health sciences. He has published numerous referred journal articles on both methodological and applied research studies, including an edited book on propensity score methods entitled, *"Propensity Score Analysis: Fundamentals and Developments."* He has been invited or selected to offer more than ten professional development courses and workshops on propensity score methods at annual conferences of various professional organizations, such as the American Statistical Association (JSM and CSP), the American Public Health Association, the International Chinese Statistical Association (ICSA), the American Educational Research Association, the American Evaluation Association, and the South African Statistical Association.

SC08: Power Prior: Incorporating Historical Data for Bayesian Inference and Designs of Clinical Trial

1:30pm-5:30pm, June 9, 2019

Ming-Hui Chen, University of Connecticut

Abstract:

The power prior has been widely used in many applications covering a variety of disciplines. The power prior is intended to be an informative prior constructed from historical data. This short course provides a comprehensive overview of the power prior and its applications to date.

The course starts with a brief introduction of the basic formulation of power prior and then presents various variations of power prior, including the full power prior, the normalized power prior, the partial discounting power prior, and the partial borrowing power prior. The theory of the power prior and its properties, including the theoretical justification, connections to hierarchical models, and frequentist properties of the posterior estimates using power priors in linear models as well as generalized linear models will be reviewed. The determination of a guide value of the power parameter will be discussed. A revised information gain measure will be introduced and illustrated on how to quantify the information gain or loss while incorporating historical data. A detailed analysis of benchmark dose toxicology data will be presented.

The course also covers an important application in the design of clinical trials from the Bayesian perspective. A general Bayesian methodology for the design of non-inferiority clinical trials with a focus on controlling type I error and power will be presented. Bayesian methods are then applied to the design of a non-inferiority medical device clinical trial with historical data from previous trials to demonstrate superiority of the Bayesian methods in sample size reduction.

About the instructor:

Dr. Ming-Hui Chen is currently Professor and Head of the Department of Statistics at the University of Connecticut (UConn). He was elected to Fellow of the International Society for Bayesian Analysis in 2016, Fellow of the Institute of Mathematical Statistics in 2007, and Fellow of the American Statistical Association in 2005. He has published over 375 statistics and biostatistics methodological and medical research papers in mainstream statistics, biostatistics, and medical journals. He has also published five books, including two advanced graduate-level books on Bayesian survival analysis and Monte Carlo methods in Bayesian computation. He served as President of the International Chinese Statistical Association (ICSA) in 2013, Program Chair and Publication Officer of SBSS of the American Statistical Association (ASA) and the ASA Committee on Nomination for 2016-2017 to nominate candidates for ASA President/Vice President. Currently, he serves as Co Editor-in-Chief of Statistics and Its Interface and an Associate Editor of JASA, JCGS, and LIDA.

SC10: Statistical Analysis of Network Data

One day (8:30am-12:30pm and 1:30pm-4:30pm), June 9, 2019

Eric Kolaczyk, Boston University

Abstract:

Networks have permeated everyday life through everyday realities like the Internet, social networks, and viral marketing. Their use has become especially prevalent in the biological and life sciences, particularly in computational biology and neuroscience. Accordingly, network analysis is an important growth area in the quantitative sciences, with roots in social network analysis going back to the 1930s and graph theory going back centuries. Measurement and analysis are integral components of network research, and statistical methods therefore play a critical role in network analysis. This course will provide a broad treatment of foundational topics relevant to statistical analysis of network data across the disciplines. Material will be organized according to a statistical taxonomy, with presentation entailing a conscious balance of conceptual and technical aspects. Additionally, practical application of network analysis will be demonstrated in the context of the R software environment. The focus in the morning will be on manipulation, visualization, and descriptive analysis of network data. In the afternoon, we will draw on topics from network sampling and inference, the modeling of networks and network-indexed processes, and networked experiments. Specific examples of network analysis will be taken from a variety of domain areas, with emphasis on computational biology and neuroscience and on social networks.

About the instructor:

Dr. Eric Kolaczyk is a tenured professor and director of the program in statistics in the Department of Mathematics and Statistics at Boston University, where he is also a Data Science Faculty Fellow and an affiliated faculty member in the program in bioinformatics, the program in computational neuroscience, and the division of systems engineering. His major research interests for 15 years have revolved around the statistical analysis of network-indexed data and include both the development of basic methodology and inter-disciplinary work with collaborators in bioinformatics, computer science, geography, neuroscience, and sociology. He has authored a book: *“Statistical Analysis of Network Data: Methods and Models”* and coauthored another book: *“Statistical Analysis of Network Data with R”*. Dr. Kolaczyk has served as associate editor on several journals, including JASA and JRSS-B. He has also served as (co)organizer for workshops focused on networks and network data, including as lead organizer for a year-long program at SAMSI in 2010-11. He is an elected fellow of the American Statistical Association (ASA) and the Institute for Mathematical Statistics (IMS), as well as of the American Association for the Advancement of Science (AAAS). He is also an elected senior member of the Institute for Electrical and Electronics Engineers (IEEE), and an elected member of the International Statistical Institute (ISI).

SC11: Data Science, Big Data, and Deep Learning for Statistician

One day (8:30am-12:30pm and 1:30pm-4:30pm), June 9, 2019

Hui Lin, Netlify

Ming Li, Amazon.com Inc.

Abstract:

With recent big data, data science and deep learning revolution, enterprises ranging from FORTUNE 100 to startups across the world are hungry for data scientists and machine learning scientists to bring actionable insight from the vast amount of data collected. In the past a couple of years, deep learning has gained traction in many application areas and it becomes an essential tool in data scientist's toolbox. In this course, participant will develop a clear understanding of the big data cloud platform, technical skills in data sciences and machine learning, and especially the motivation and use cases of deep learning through hands-on exercises. We will also cover the "art" part of data science and machine learning to guide participants to learn typical agile data science project flow, general pitfalls in data science and machine learning, and soft skills to effectively communicate with business stakeholders. The big data platform, data science, and deep learning overviews are specifically designed for audience with statistics education background. This course will prepare statisticians to be successful data scientists and machine learning scientist in various industries and business sectors with deep learning as focuses.

About the instructor:

Dr. Hui Lin is leading and building data science department at Netlify since 2018. Before Netlify, she was a Data Scientist at DowDuPont. She was a leader in the company of applying advanced data science to enhance Marketing and Sales Effectiveness. She provided data science leadership for a broad range of predictive analytics and market research analysis from 2013 to 2018. She is the co-founder of Central Iowa R User Group, blogger of scientistcafe.com and 2018 Program Chair of ASA Statistics in Marketing Section. She enjoys making analytics accessible to a broad audience and teaches tutorials and workshops for practitioners on data science. She holds MS and Ph.D. in statistics from Iowa State University.

Dr. Ming Li is currently a Research Scientist at Amazon. He organized and presented 2018 JSM Introductory Overview Lecture: Leading Data Science: Talent, Strategy, and Impact. He was the Chair of Quality & Productivity Section of ASA for 2017. He was a Data Scientist at Walmart and a Statistical Leader at General Electric Global Research Center. He obtained his Ph.D. in Statistics from Iowa State University at 2010. With deep statistics background and a few years' experience in data science, he has trained and mentored numerous junior data scientist with different background such as statistician, programmer, software developer, database administrator and business analyst. He is also an Instructor of Amazon's internal Machine Learning University and was one of the key founding member of Walmart's Analytics Rotational Program which bridges the skill gaps between new hires and productive data scientists.

SC12: Advancing Clinical Development through Precision Medicine and Innovative Designs: Concepts, Rationale, and Case Studies

One day (8:30am-12:30pm and 1:30pm-4:30pm), June 9, 2019

Sandeep M Menon, Pfizer Inc. / Boston University/ Tufts University School of Medicine

Weidong Zhang, Pfizer Inc.

Abstract:

Precision medicine has paved the way for a new era of delivering tailored treatment options to patients according to their biological profiles. Advancement of the biotechnologies such as next generation sequencing technology (NGS) and other omics technologies have enabled us to interrogate a patient's many molecular biomarkers and associate them with disease and drug responses. In addition, incorporation of biomarker information in the innovative clinical trial design has presented drug developers unprecedented opportunities to bring a successful drug to patients in needs. The first part of this course will focus on the concept of precision medicine, biomarker discovery and its application in clinical development. Comprehensive review of omics data and major technologies will be presented. Statistical considerations and challenges such as data normalization, dimension reduction and biomarker threshold development will be discussed. Strategies in Bayesian framework leveraging historical biomarker data for quantitative decision making in early clinical trials will also be presented in detail. The second part of this course will focus on the strategy of the study design that is important to critically determine biomarker performance, reliability and eventually regulatory acceptance. A general overview of the concept and statistical methodologies and designs related to precision medicine will be presented. Specifically, we will discuss a variety of designs including adaptive designs available at our disposal and its merits and limitations.

About the instructor:

Dr. Sandeep Menon is currently the Vice President and the Head of Early Clinical Development Statistics at Pfizer Inc. and holds adjunct faculty positions at Boston University and Tufts University School of Medicine. He is an elected fellow of American Statistical Association. He leads an organization of early development statisticians, clinicians and pharmacologists globally and is the executive member of the early clinical development leadership team and global clinical leadership team. He is internationally known for his technical expertise especially in adaptive designs, precision medicine, multi-regional trials, and small populations. He has participated in the core review of draft version of the regulatory guidance documents. He has co-authored and co-edited numerous books and contributed to influential papers in this area. He has taught short courses internationally and is a regular invited speaker in academia, FDA, Industry forums and Business Management institutes. He is the co-author and co-editor of the books titled "*Clinical and Statistical Considerations in Personalized Medicine*", "*Modern Approaches to Clinical Trials Using SAS: Classical, Adaptive, and Bayesian Methods*" and "*Biosimilars – Clinical Development.*" Dr. Menon completed his master's and Ph.D. in Biostatistics at Boston University and research

assistantship at Harvard Clinical Research Institute. He has received several awards for academic excellence.

Dr. Weidong Zhang is a Senior Director in the Statistical Research and Innovation department at Pfizer. His responsibilities include statistical methodology development for precision medicine and providing scientific leadership and consultation in clinical biomarker strategy to senior Pfizer management and portfolios spanning from target discovery through proof-of-concept, and late phase studies across multiple therapeutic areas including oncology, immunology and inflammation, rare disease, and cardiovascular and metabolism. His research interest focuses on developing new statistical methods in biomarker discovery and precision medicine studies using high throughput omics data generated from cutting edge technologies including next-generation sequencing technology. He has taught short courses for ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop, the Biopharmaceutical Applied Statistics Symposium (BASS), and the International Chinese Statistics Association (ICSA) Applied Statistics Symposium. He obtained his PhD degree in Statistical Genetics and MS degree in Statistics from the University of Wisconsin-Madison.