

The 37th New England Statistics Symposium

SHORT COURSES

Essentials of Data Science: Project Management, Reproducibility, Communication, and Ethics

Half day - PM Session; Hybrid, May 20, 2024

Instructor: Dr. Jun Yan is a Professor in the Department of Statistics at the University of Connecticut and a Research Fellow at the Center for Population Health at UConn Health.

Outline: This course is designed to arm participants with the fundamental skills essential for success in the data science field. Integrating theoretical insights with practical exercises, the course covers key areas including efficient project management using Git, creating reproducible analytics with Quarto, and the nuances of ethical considerations in data science.



Dr. Jun Yan

Informative Prior Elicitation Using Historical Data, Expert Opinion, and Other Sources

Full day; Hybrid, May 21, 2024

Instructor: Dr. Joseph G. Ibrahim, Alumni Distinguished Professor of Biostatistics at the University of North Carolina. Dr. Ibrahim's areas of research focus are Bayesian inference, missing data problems, cancer, and clinical trials.

Outline: This full-day short course is designed to give biostatisticians and data scientists a comprehensive overview of informative prior elicitation from historical data, expert opinion, and other data sources, such as real-world data, prior predictions, estimates, and summary statistics.



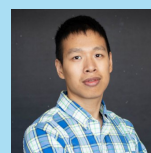
Dr. Joseph G. Ibrahim

Introduction to Causal Inference: From Theory to Practice

Full day; Hybrid, May 21, 2024

Instructor: Dr. Linbo Wang is an assistant professor in the Department of Statistical Sciences and the Department of Computer and Mathematical Sciences at the University of Toronto.

Outline: This course offers a quick overview of causal inference concepts and methods, designed for those new to the field but familiar with basic statistical tools such as regression models and R programming proficiency. It covers the mathematical underpinnings and cutting-edge methods of causal analysis, aiming to squeeze as much evidence as possible from imperfect studies about the causal effects of interest.



Dr. Linbo Wang

Tutorial on Deep Learning and Generative AI

Full Day, Hybrid, May 21, 2024

Instructor: Dr. Haoda Fu is an Associate Vice President and an Enterprise Lead for Machine Learning, Artificial Intelligence, and Digital Connected Care from Eli Lilly and Company.

Outline: Designed specifically for individuals possessing a strong foundation in statistics and biostatistics, this course seeks to bridge the gap into the realm of deep learning and generative AI. Beginning with fundamental knowledge of deep learning, participants will be guided through hands-on implementations using the PyTorch framework.



Dr. Haoda Fu

The 37th New England Statistics Symposium

SHORT COURSES

An Introduction to the Statistical Foundations of Transfer Learning

Full day; Hybrid, May 20, 2024

Instructor: Dr. Yang Feng is a professor of Biostatistics at New York University. Ye Tian is a fifth-year PhD student in Statistics at Columbia University.

Outline: This course offers a comprehensive introduction to the statistical foundations underpinning the prevalent machine learning technique: transfer learning. We delve into how transfer learning effectively transfers knowledge from one task to another in an adaptive and robust fashion, thereby enhancing model performance across both supervised and unsupervised learning frameworks.



Dr. Yang Feng



Ye Tian

Large Scale Data Science

Full day; Hybrid, May 20, 2024

Instructor: Dr. Marc G. Genton is a Distinguished Professor of Statistics at the Spatio-Temporal Statistics and Data Science (STSDS) research group, King Abdullah University of Science and Technology (KAUST), Saudi Arabia. Dr. Sameh Abdulah is a research scientist at the Extreme Computing Research Center (ECRC), King Abdullah University of Science and Technology, Saudi Arabia. Dr. Mary Lai Salvaña is an Assistant Professor in Statistics at the University of Connecticut (UConn).

Outline: The course content will cover the basic concepts of large-scale spatial statistics on parallel systems through synthetic and real data examples using both exact and approximation methods. The course will also provide a comprehensive comparison between existing Geostatistics packages (fields and GeoR) with the cutting-edge HPC packages (ExaGeoStatR and MPCR) to show the main contribution and benefits of using HPC techniques on leading-edge parallel hardware architectures such as GPUs and supercomputers.



Dr. Marc G. Genton



Dr. Sameh Abdulah



Dr. Mary Lai Salvaña

Statistical Network Analysis in R

Half day - AM Session; Hybrid, May 20, 2024

Instructor: Dr. Eric Kolaczyk is a professor in the Department of Mathematics and Statistics, and inaugural director of the McGill Computational and Data Systems Initiative (CDSI).

Outline: A gentle introduction to the statistical analysis of network data, largely through the lens of the R package igraph. Topics to be covered include basic definitions and concepts in networks, manipulation and visualization of network data, and tools for describing network characteristics, as well as a brief look at select inferential topics such as node clustering (aka 'community detection') and network modeling.



Dr. Eric Kolaczyk