DEPARTMENT OF STATISTICS AND DATA SCIENCES

UT Summer Statistics Institute

The University of Texas at Austin
May 21–24, 2018

UT's Summer Statistics Institute (SSI) offers intensive four-day workshops on diverse topics from introductory data sciences to advanced statistics. Whether you are new to data analysis or a seasoned statistician, SSI provides a unique hands-on opportunity to acquire valuable skills directly from experts in the field.

The UT Summer Statistics Institute (SSI) is open to 700 participants.

stat.utexas.edu/training/ssi

Table of Contents

INTRODUCTION

COURSE DESCRIPTIONS: **MORNING**

COURSE DESCRIPTIONS: **AFTERNOON**

THE DEPARTMENT OF STATISTICS AND DATA SCIENCES

at The University of Texas at Austin is proud to host the 11th Annual 2018 UT Summer Statistics Institute (SSI).

The three main purposes of SSI are:

- To provide participants with access to new statistical knowledge and skills
- To give participants hands-on experience with data analysis
- To prepare participants to interpret studies employing statistical methods

All 2018 SSI courses will be held in the UT Campus in the College of Liberal Arts (CLA) building, the Flawn Academic Center (FAC), as well as Robert A Welch (WEL) Hall.

COURSES:

The 2018 Summer Statistics Institute will offer 27 courses covering introductory statistics, statistical software, and statistical methods and applications. Each course will meet for four half-days, mornings or afternoons, for a total of twelve hours. There will be no examinations or tests, and participants will receive a certificate upon completion of each course. Academic credit will not be issued.

The following table lists the courses offered. An outline of the material to be covered in each course can be found on the SSI website at stat.utexas.edu/training/ssi. Participants are encouraged to carefully check the prerequisite knowledge specified for each course.

RETURN TO TABLE OF CONTENTS

Category	Morning (9:00 a.m.–noon)	Afternoon (1:30–4:30 p.m.)
Software and Database	Data Analysis using SPSS	Data Analysis using SAS
	Introduction to Causal Inference	Introduction to Data Science in Python
	Introduction to Data Analysis and Graphics Using R	Introduction to GIS
		Introduction to SQL and Relational Database Design
		Introduction to Stata
Statistical Methods	Analysis of Variance (Anova)	Geospatial Data Analysis in R
	Big Data Analytics: Theory and Methods	Introduction to Statistics
	Introduction to Regression	Introduction to Bayesian Statistics
	Introduction to Statistics	Large Scale Data Analysis with Hadoop and Spark
	Structural Equation Modeling	Time Series Forecasting and Modeling
Design and Application	Applied Hierarchical Linear Modeling	Non Parametric Statistical Methods for Small Datasets
	Common Mistakes in using Statistics: Spotting Them and Avoiding Them	The Power and Pleasure of Probability
	Data Science in Industry with R	Questionnaire Design and Survey Analysis
	Missing Data Analysis using Mplus	
	Power Analysis for Proposal Writing	
	Statistics for the Dissertation	

RETURN TO TABLE OF CONTENTS

NEW FOR 2018!

Analysis of Variance (ANOVA): The purpose of this course is to familiarize participants with the use and interpretation of the In this course, participants will learn the theory, use, and application of ANOVA) statistical test. ANOVA is used to analyze group differences on numeric response variables; it has applications across a wide variety of domains including science and business.

Applied Hierarchical Linear Modeling: This applied, hands-on course provides an introduction to the basic concepts and applications of hierarchical linear models. The course will cover applications in social science research (e.g. neighborhood effects research, school effect research), growth curve modeling (e.g., repeated measures on individuals), as well as introduce models for dichotomous outcomes.

Introduction to Causal Inference: This course covers contemporary statistical approaches to questions about causality. It introduces an important framework for thinking about cause-and-effect (the potential outcomes framework) both in the context of randomized experiments and in observational studies. Techniques covered in the course include blocking/stratification, instrumental variables estimation, matching methods (including propensity scores), and regression-discontinuity designs.

Missing Data Analysis Using Mplus: This workshop covers the problem of missing data that is common to social science research. Topics include patterns and mechanisms of missing data as well as conventional and modern missing data treatments, focusing particularly on the use of maximum likelihood and multiple imputation. Missing data treatments will be applied to various statistical models, such as multiple regression and factor analysis. Workshop participants will learn when a given missing data treatment is suitable and how such methods can be implemented using Mplus software.

Statistics for the Dissertation: A comprehensive review of common statistical techniques for PhD students in non-mathematically leaning fields. We will cover methods that may be useful as they design their dissertations such as t-tests, linear and multiple regression, various correlation equations (Pearson, Spearman, point-biserial), logistic regression, ANOVA, and ways to apply these in combination with qualitative research. An emphasis will be place on learning how to interpret the terms associated with these methods.

The Power and Pleasure of Probability: Participants will learn fundamental rules for computing probabilities, including the explanations behind some famous paradoxical puzzles, gain insight into statistical practice (including the frequentist vs. Bayesian debate) through a deeper understanding of connections with probability theory, dispel misconceptions and cognitive biases surrounding randomness, and explore simulation as a tool for problem solving and as a means to understand limit theorems.

RETURN TO TABLE OF CONTENTS

REGISTRATION AND COST

To register, visit the following website: stat.utexas.edu/training/ssi. *(A UT EID is required. See below for information on how to obtain an EID.)*

Registration dates and fees are as follows:

Dates	Category	Registration Fees (per course)
January 8, 2018 through May 7, 2018	UT-Austin Students	$200
	UT-Austin Faculty/Staff	$300
	UT-System Faculty/Staff	$300
	Non-UT-Austin Students	$250
	Non-UT Participants	$600
	Groups of five or more from the same institution or agency (**Contact SDS to register groups)	$480 per person per course

***Please contact SDS at (512)-232-0693 to register groups of five or more.*

Contact our office at (512)-232-0693 for questions or more information.

OBTAINING A UT EID

You must have a current UT EID to register for SSI. To obtain a UT EID, visit idmanager.its.utexas.edu/eid_self_help and select "Get a UT EID." If you already have a UT EID, but you do not know your password, select "Find/Reset My Password." Your UT EID will allow you access to registration, your course website, and software applications during SSI. Contact our office at (512)-232-0693 with any questions.

METHODS OF PAYMENT ACCEPTED

Registration fees can be paid by credit card (MasterCard, Visa, Discover, or American Express) or by IDT (UT-Austin employees/students only). To pay by IDT please enter your appropriate discount code and provide the IDT number at checkout.

REFUND AND CANCELLATION POLICY

A full refund of registration fees, less a $25 cancellation fee, will be available if requested in writing to the Department of Statistics and Data Sciences and received by March 30, 2018. No refunds will be made after that date.

RETURN TO TABLE OF CONTENTS

Please note that course substitutions cannot be made. If you fail to cancel by the deadline and do not attend, you are still responsible for full payment. UT-Austin reserves the right to cancel SSI courses and to return all fees in the event of insufficient registration.

WAITLIST POLICY

SSI does not maintain a priority waiting list. However, if you are unable to register for a course because it is full, contact us at stat.admin@utexas.edu or (512) 232-0693 and provide us with the name of the course and your email address. If there is sufficient demand for a course and resources allow, we might open additional seats or sections in those courses beginning April 1. You will be notified by email if additional seats will be opened. Registration will continue to be first-come, first-serve for these additional seats.

SOFTWARE

Statistical software will be used in many courses. Participants are provided with access to this software at no additional cost. In some courses, participants might be expected to bring a laptop and install freeware. Please see course information posted at stat.utexas.edu/training/ssi for detailed information on computer requirements.

MISCELLANEOUS

Beverages and snacks will be available for morning and afternoon breaks. Vending machines selling sodas and snacks can be found on the first and second floors of the College of Liberal Arts (CLA) building, the Flawn Academic Center (FAC), as well as Robert A. Welch (WEL) Hall.

CONTACT

Department of Statistics and Data Sciences

Tel: (512) 232-0693

Fax: (512) 232-1045

Email: stat.admin@austin.utexas.edu

Website: stat.utexas.edu

RETURN TO TABLE OF CONTENTS

Analysis of Variance (ANOVA)

Category	Statistical Methods
Prerequisite Knowledge	Participants should understand basic descriptive statistics (mean, standard deviation, variance) and research design (collecting data). Additionally, students should be comfortable managing data in MS Excel or a similar program.
Description	In this course, participants will learn the theory, use, and application of the Analysis of Variance (ANOVA) statistical test. ANOVA is used to analyze group differences on numeric response variables; it has applications across a wide variety of domains including science and business. Instruction will begin with basic one-way ANOVAs and continue through two-way ANOVAs with interactions. Additionally, students will learn how to analyze multiple response variables by using a multivariate analysis of variance (MANOVA). Data will be analyzed by hand, when possible, and through the use of the computer program R.
Intended Audience	This course is designed for a wide variety of participants including graduate students, researchers, and business practitioners. Anyone who plans on using or analyzing data will benefit from this course.
Computer Requirements	Participants should bring a personal laptop. Installation of R and RStudio should be completed prior to the start of class; instructions will be provided.
Time	9:00 AM – 12:00 Noon
Instructor	Lauren Blondeau
Department	Statistics & Data Sciences
Title	Lecturer
Bio	Lauren Blondeau received her Ph.D. in Educational Psychology from The University of Texas at Austin where she currently teaches in the Department of Statistics & Data Sciences. Her research interests include gender differences in undergraduate STEM education, the impostor phenomenon, and self-efficacy.

RETURN TO TABLE OF CONTENTS

Applied Hierarchical Linear Modeling

Category	Design and Application
Prerequisite Knowledge	Knowledge of multiple regression methods and working knowledge of SAS software (reading in data, recoding variables, descriptive statistics, regression modeling.)
Description	This applied, hands-on course provides an introduction to the basic concepts and applications of hierarchical linear models. The course will cover applications in social science research (e.g. neighborhood effects research, school effect research), growth curve modeling (e.g., repeated measures on individuals), as well as introduce models for dichotomous outcomes. Topics will include multilevel data structures, model building and testing, fixed random effects, and interpretation of results. At the end of the course, participants should be able to specify a social science research question requiring hierarchical linear modeling, understand when and why hierarchical linear models should be used, apply hierarchical linear models to nested data, and correctly interpret analysis results from hierarchical linear models.
Intended Audience	Graduate students and faculty in the social sciences who want to learn to apply hierarchical linear modeling to nested data.
Computer Requirements	Applied Hierarchical Linear Modeling will be held in a computer classroom where participants will have access to SAS.
Time	9:00 AM – 12:00 Noon
Instructor	Catherine Cubbin
Department	School of Social Work
Title	Professor & Associate Dean for Research
Bio 	Dr. Catherine Cubbin is Professor & Associate Dean for Research in the Steve Hicks School of Social Work and a Faculty Research Associate at the Population Research Center, at The University of Texas at Austin. Dr. Cubbin's research focuses on using epidemiological methods to better understand socioeconomic and racial/ethnic inequalities in health for the purpose of informing policy. Specific areas of her research include using contextual analysis to investigate how neighborhood environments might explain social inequalities in health, and the measurement of socioeconomic status/position in studies of racial/ethnic disparities in health. She teaches the hierarchical linear modeling (HLM) course in the Steve Hicks School of Social Work.

RETURN TO TABLE OF CONTENTS

Big Data Analytics: Theory and Methods

Category	Statistical Methods
Prerequisite Knowledge	Elementary knowledge of probability, statistics, and calculus, plus familiarity using computers, R and SAS.
Description	This course will cover theory and methods based on structured, semi-structured, and unstructured data based on real-world scenarios. Examples will include application of mathematical statistics, machine learning, stochastic processes, and mathematical methods to numeric, click-stream, and text data from the real world. The range of algorithms will span outlier detection, projections, principal component analysis, factor analysis, independent component analysis, spectral analysis, regression analysis, neural networks, statistical clustering, discriminant analysis, Markov chains (discrete and continuous), and methods from information theory. We will use R and SAS programming languages for analyzing the data.
Intended Audience	Graduate and undergraduate students, faculty, and practitioners in industry.
Computer Requirements	Big Data Analytics: Theory and Methods will be held in a computer classroom where students will have access to SAS and R.
Time	9:00 AM – 12:00 Noon
Instructor	Choudur Lakshminarayan
Department	HP Labs
Title	Principal Research Scientist
Bio	Choudur K. Lakshminarayan specializes in the areas of mathematical statistics, applied mathematics, machine learning and data mining with applications in digital marketing, sensors and sensing in healthcare, energy, large-scale data centers, semiconductor manufacturing, and histogram statistics in query Optimization. He contributed to developing novel algorithms for statistical clustering, time series, and classification using structured, semi-structured, and unstructured data. He is widely published in peer-reviewed international conferences and journals, and his name appears as an inventor in over 50 patents; granted, published, or pending. He has conducted workshops in Data Mining and Analytics in India, Hong Kong, China, the Middle East and the USA. He taught as a visiting professor at the Indian Institute of Technology, Hyderabad, and the Indian Institute of Information Technology, Bangalore. He speaks regularly at international conferences, symposia, and universities. He served as a consultant to government, and private industry in the US and India. He holds a PhD in mathematical sciences, and lives in Austin, Texas.

RETURN TO TABLE OF CONTENTS

Common Mistakes in Using Statistics: Spotting Them and Avoiding Them

Category	Design and Application
Prerequisite Knowledge	This is an intermediate level course, but is also appropriate for people who have taken advanced statistics courses that have been weak on discussion of limitations of techniques. Familiarity with random variables, sampling distributions, hypothesis testing, and confidence intervals are the only statistical prerequisites. These concepts will be reviewed in the course, providing more depth than is given in most introductory courses. Willingness to engage in "minds-on" learning is an important prerequisite.
Description	In 2005, medical researcher John P. Ioannidis asserted that most claimed research findings are false. Since then, this concern has spread to other fields, and is sometimes referred to as "the replication crisis". For example, in 2011, psychologists Simmons, Nelson and Simonsohn brought further attention to this topic by using practices common in their field to "show" that people were almost 1.5 years younger after listening to one piece of music than after listening to another. In 2015, the Open Science Collaboration published the results of replicating 100 studies that had been published in three psychology journals. They concluded that, "A large portion of replications produced weaker evidence for the original findings," despite efforts to make the replication studies sound. These articles highlight the frequency and consequences of misunderstandings and misuses of statistical inference techniques. These misunderstandings and misuses are often passed down from teacher to student or from colleague to colleague, and some practices based on these misunderstandings have become institutionalized. This course will discuss some of these misunderstandings and misuses. Topics covered include the File Drawer Problem (AKA Publication Bias), Multiple Inference (AKA Multiple Testing, Multiple Comparisons, Multiplicities, or The Curse of Multiplicity), Data Snooping, the Statistical Significance Filter, the Replicability Crisis, and ignoring model assumptions. To aid understanding of these mistakes, about half the course time will be spent deepening understanding of the basics of statistical inference beyond what is typically covered in an introductory statistics course. Participants will have online access to downloadable slides used for class presentation, plus downloadable supplemental materials. The latter will elaborate on some points discussed briefly in class; give specific suggestions for teachers, readers, researchers, referees, reviewers, and editors to deal with and reduce the high incidence of mistakes in using statistics; and provide references. Thus, students in this course should gain understanding of these common mistakes, how to spot them when they occur in the literature, and how to avoid them in their own work. Many students will also gain deeper understanding of basic statistical concepts such as p-values, confidence intervals, sampling distributions, robustness, model assumptions, Type I and II errors, and statistical power.

RETURN TO TABLE OF CONTENTS

Intended Audience	This course is intended for a wide audience, including: graduate students who read or do research involving statistical analysis; workers in a variety of fields (e.g., public health, social sciences, biological sciences, public policy) who read or do research involving statistical analysis; faculty members who teach statistics, read or do research involving statistical analysis, supervise graduate students who use statistical analysis in their research, peer review research articles involving statistical analysis, review grant proposals for research involving statistical analysis, or are editors of journals that publish research involving statistical analysis; and people with basic statistical background who would like to improve their ability to evaluate research relevant to medical treatments for themselves or family members.
Computer Requirements	None.
Time	9:00 AM – 12:00 Noon
Instructor	Mary Parker
Department	Mathematics, Department of Statistics & Data Sciences
Title	Senior Lecturer
Bio	 Mary Parker has been a Lecturer and Senior Lecturer in the UT Mathematics Department and UT Statistics & Data Sciences Department since 1989. She received her PhD in 1988 from the UT Mathematics Department, working under Professor Carl Morris on Empirical Bayes Estimation. She has taught Mathematical Statistics at the undergraduate and graduate level and occasionally other statistics courses. In her courses, she emphasizes careful attention to the assumptions needed for the various statistical techniques and the implications of those assumptions for the use of the technique. She also teaches courses in Elementary Statistics and various other courses at Austin Community College, and is active in the statistics education communities of the American Statistical Association, the Mathematical Association of America, and the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE.) During her early years of teaching in the UT Mathematics Department she frequently talked with Professor Martha Smith as Dr. Smith shifted her teaching emphasis more to statistics. Dr. Smith found that her students needed, and were interested in, discussions of how statistics techniques can be misunderstood and misapplied, so she developed materials on that. She shared those with students and others in various ways, including a successful short course in the UT Summer Statistics Institute between 2010 and 2016, and Dr. Parker took over and adapted the course in 2017.

RETURN TO TABLE OF CONTENTS

Data Analysis using SPSS

Category	Software and Database
Prerequisite Knowledge	Participants should be familiar with basic descriptive and inferential statistics (topics covered in an introductory statistics course).
Description	This course is designed to teach participants how to use SPSS for data manipulation and analysis. The course will begin with an overview of the software, data handling and manipulation, descriptive statistics, and data visualization. The remainder of the course will focus on inferential analyses including correlation, simple and multiple linear regression, chi-square tests, t-tests, and ANOVA. As the inferential analyses are conducted, the basic theory behind each analysis will be reviewed and instruction about how to check each of the associated assumptions will be addressed.
Intended Audience	Individuals with an interest in using SPSS for data analysis
Computer Requirements	Data Analysis Using SPSS will be held in a computer classroom where participants will have access to SPSS software.
Time	9:00 AM – 12:00 Noon
Instructor	Lindsey Smith
Department	Department of Statistics & Data Sciences
Title	Lecturer
Bio	Lindsey Smith received her Ph.D. from The University of Texas at Austin where she now teaches undergraduate and graduate statistics courses. Her primary research interest is the evaluation of multilevel models, specifically its use with multiple membership data structures.

RETURN TO TABLE OF CONTENTS

Data Science in Industry with R

Category	Design and Application
Prerequisite Knowledge	A basic familiarity with R and RStudio is the only prerequisite. Students should know how to install/load packages, use RStudio to create/edit/run script files, and some familiarity with data frames. There is NO need for web development skills or machine learning skills. There also is no prerequisite of statistical knowledge.
Description	This course will cover some practical data science tasks found in industry. Topics will include: software development practices, connecting to databases and web APIs, parsing JSON data, data wrangling, building web applications with shiny, and making predictive models. Participants will be introduced to several commonly used R packages.
Intended Audience	Any students interested in aspects of R that are a) pertinent to data scientists in industry, and b) not necessarily introduced in an academic setting
Computer Requirements	Participants should bring a personal laptop. Installation of R and RStudio should be completed prior to the first day of class; instructions will be provided.
Time	9:00 AM – 12:00 Noon
Instructor	Richard Leu
Department	Dropoff
Title	Data Scientist
Bio	Richard Leu received a PhD in Physics and an MS in Statistics from The University of Texas at Austin. Richard currently works as a data scientist for Dropoff, applying statistics, machine learning, and operations research to same day logistics. After working in the statistics department for a year and a half as a lecturer, he moved into the data science industry. Prior to Dropoff, Richard was a principal data scientist with Clockwork Solutions performing reliability analysis, data mining, and predictive analytics in support of asset life cycle management for aviation, oil/gas, and military.

RETURN TO TABLE OF CONTENTS

Introduction to Causal Inference

Category	Software and Database
Prerequisite Knowledge	This course presupposes good numeracy, some knowledge of experimental design, and a working familiarity with regression (i.e., you should be able to run a multiple regression and interpret the output). This course will be taught entirely using R (and RStudio), so any prior experience with this software (data manipulation, scripting) will serve you well. R experience is not required; however, there are tons of good, free resources on the web for learning R, so it would be to your advantage to acquaint yourself with the basics (RStudio, specifically).
Description	This course covers contemporary statistical approaches to questions about causality. It introduces an important framework for thinking about cause-and-effect (the potential outcomes framework) both in the context of randomized experiments and in observational studies. Techniques covered in the course include blocking/stratification, instrumental variables estimation, matching methods (including propensity scores), and regression-discontinuity designs. Additional topics might include probabilistic graphical models, attrition/missing data, and principal stratification. In addition to many new techniques, you will learn easy ways to add statistical rigor to your favorite analysis procedures (e.g., matching during preprocessing, bootstrapping/randomization tests, robust estimators). After taking this course, you will have surveyed the modern approaches to causal inference and gotten your feet wet with in-class examples of each. Since R is open source software, you will have free access to all of the packages we use in the course and will be able to easily apply the techniques you have learned to your own data analysis. That being said, the topic of causal inference is enormous and many techniques are quite involved. You will gain a working knowledge of many topics, which can be developed to proficiency as you continue to study and use them in your work.
Intended Audience	This course is intended for those who want to be able to conduct their own statistical defensible causal analysis of observational data and to be able to critically review and interpret research addressing causal questions or making causal claims; for those who want to be introduced to the modern statistical framework for posing and answering causal questions; for those who wish to survey commonly used methods for causal inference in both experimental and observational settings. These techniques are increasingly important in academic research, which seeks to discover and describe cause-and-effect relationships, but are especially relevant for people in government, economists, policy makers, marketing/advertising agencies, and epidemiologists.
Computer Requirements	Participants should bring a personal laptop (Recent Windows or Mac). Installation of latest versions of R and RStudio should be completed prior to the first day of the course; instructions will be provided.
Time	9:00 AM – 12:00 Noon

RETURN TO TABLE OF CONTENTS

Instructor	Nathaniel Raley
Department	Department of Statistics & Data Sciences
Title	PhD Candidate
Bio	This is Nathaniel's fourth SSI as a participant and an assistant. He is a researcher here at The University of Texas at Austin, where he earned a MS in Statistics. Nathaniel has worked as an Instructor and Statistical Consultant for SDS, and he is currently a PhD Candidate in the department of Educational Psychology, where he has routinely used modern approaches to causal inference in his work.

RETURN TO TABLE OF CONTENTS

Introduction to Data Analysis and Graphics using R

Category	Software and Database
Prerequisite Knowledge	Absolutely no prior knowledge of R is necessary. Participants should be comfortable working with data in .xls, .csv, or similar file formats. A basic understanding of common statistical methods is recommended but not required.
Description	This hands-on course is intended to provide first-time users the ability to analyze data using R. We will start by covering basic programming skills in R and interacting with the user-friendly interface RStudio. Participants will practice using example datasets from a variety of disciplines to run statistical analyses and create graphical displays of the data. Those with some prior R experience will benefit from the more advanced statistical methods (multiple linear regression, generalized linear models, multi-factor ANOVA, mixed models) and programming topics (user-written functions and simulations) covered in the second half of the course.
Intended Audience	This course is designed for those interested in using R to manage, analyze, and display data. Whether coming from academia, industry, or government, this free and open-source software is a great tool for any researcher or analyst.
Computer Requirements	Participants should bring a personal laptop (recent Windows or Mac). Installation of latest versions of R and RStudio should be completed prior to the first day of the course; instructions will be provided.
Time	9:00 AM – 12:00 Noon
Instructor	Sally Ragsdale
Department	Department of Statistics & Data Sciences
Title	Lecturer, Consultant
Bio	Sally received her M. S. in Statistics from The University of Texas at Austin in May 2012 and has been a statistical consultant for SDS since July 2012. As a consultant, she provides one-on-one assistance to researchers with questions about study design, data management, running appropriate statistical analyses, and interpreting results. In addition to teaching SDS 328M Biostatistics, an undergraduate introductory stats course where students use R in a weekly lab, she also teaches various software and topic short courses each semester.

RETURN TO TABLE OF CONTENTS

Introduction to Regression

Category	Statistical Methods
Prerequisite Knowledge	Familiarity with the basics of statistical inference is required. For example, participants should know the basics of random variables, probability distributions, sample statistics, hypothesis testing, and confidence intervals.
Description	The objective of this course is to provide participants with a broad base of understanding in the application of regression analysis. We will begin with basic fundamentals and move to simple regression. We will continue with discussions of multiple regression (including diagnostics, correct application, and interpretation), dummy coding, the use of regression in mediation and moderation, and finish up with logistic regression. The class will use R and RStudio to run and save our work in RMarkdown for easy reproducibility.
Intended Audience	The intended audience is anyone who wants to learn the fundamentals of regression analysis to apply to their own research questions or to serve as a background for learning more advanced techniques.
Computer Requirements	Participants should bring a personal laptop (recent Windows or Mac). Installation of R and RStudio should be completed prior to the first day of the course; instructions will be provided.
Time	9:00 AM – 12:00 Noon
Instructor	Michael Mahometa
Department	Department of Statistics & Data Sciences
Title	Manager of Statistical Consulting and Lecturer
Bio	 Michael J. Mahometa is the manager of Consulting Services at the Department of Statistics & Data Sciences (SDS) at The University of Texas at Austin. He received his Ph.D. in Psychology from The University of Texas at Austin in 2006. His major course work was completed in Behavioral Neuroscience, with a minor in Statistics. His background in animal models of learning makes him familiar with full factorial designs—which he quickly expanded into a love of all things regression. Dr. Mahometa has been a statistical consultant for the SDS department since its inception and enjoys helping not only students from his class, but also faculty and staff in their research endeavors.

RETURN TO TABLE OF CONTENTS

Introduction to Statistics (AM)

Category	Statistical Methods
Prerequisite Knowledge	Absolutely no previous knowledge of statistics is necessary or expected. However, participants should be comfortable working with spreadsheets in Microsoft Excel (either the Mac or PC version). Those who have never used Excel should prepare before coming to SSI, as a basic familiarity with the program will be assumed.
Description	This hands-on course will introduce participants to common descriptive and inferential statistical analyses. In addition to covering the concepts behind each method, participants will also practice applying them on real datasets using Microsoft Excel. Sufficient time will be spent on understanding relevant assumptions and how to correctly interpret the results of each analysis. The specific topics covered in this course include: describing and visualizing data, t-tests, ANOVA, chi-squared test of independence, correlation, and linear regression. Optional "homework" will be offered after each class day for those who want additional practice applying the techniques discussed.
Intended Audience	This course is designed for those with little to no experience in statistics and who want use descriptive and inferential methods to analyze data. Whether coming from academia, industry, or government, participants in this course will learn the skills needed to help them better understand the data that they work with.
Computer Requirements	All participants will need a version of Excel from 2013 or newer. For PC, version 2013 or 2016 is ok, Mac users MUST have Excel 2016 (most recent version). The University of Texas at Austin students and staff can download Excel 2016 for free through campus resources.
Time	9:00 AM – 12:00 Noon
Instructor	Kristin Harvey
Department	Department of Statistics & Data Sciences
Title	Lecturer
Bio 	Kristin Harvey is a lecturer for the Department of Statistics & Data Sciences at The University of Texas at Austin. She has a Master's degree in Educational Psychology specializing in Program Evaluation and a Ph.D. in Educational Psychology specializing in Human Development, Culture, and Learning Sciences. She teaches and coordinates a large introductory statistics course for health science and pre-nursing students.

RETURN TO TABLE OF CONTENTS

Missing Data Analysis Using Mplus

Category	Design and Application
Prerequisite Knowledge	Participants should have a good working knowledge of multiple regression. Applications in the workshop will primarily involve multiple regression, but might also include factor analysis, analysis of covariance, and logistic regression. No previous experience with Mplus is necessary.
Description	This workshop covers the problem of missing data that is common to social science research. Topics include patterns and mechanisms of missing data as well as conventional and modern missing data treatments, focusing particularly on the use of maximum likelihood and multiple imputation. Missing data treatments will be applied to various statistical models, such as multiple regression and factor analysis. Workshop participants will learn when a given missing data treatment is suitable and how such methods can be implemented using Mplus software.
Intended Audience	The workshop is designed for graduate students, applied researchers, and faculty who wish to learn about the proper treatment of missing data, particularly as it is used in applied research studies in the fields of education, psychology, and the social sciences (although the methods are applicable to virtually any data analyses). Missing data are commonplace, and this workshop is intended to help applied researchers select and apply appropriate missing data treatments.
Computer Requirements	Missing Data Analysis Using Mplus will be held in a computer classroom where participants will have access to Mplus software.
Time	9:00 AM – 12:00 Noon
Instructor	Keenan Pituch
Department	Educational Psychology
Title	Associate Professor
Bio	Keenan Pituch (Ph.D., Florida State University) is Associate Professor of Quantitative Methods in the Department of Educational Psychology at The University of Texas at Austin. His research interests include missing data analysis, multilevel modeling, mediation analysis, intensive longitudinal modeling, and multivariate analysis of variance. Dr. Pituch has published over 40 peer-reviewed articles and is an author of Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS (2016, 6th edition). He has taught a variety of quantitative methods courses, including Missing Data Analysis, Survey of Multivariate Methods, Hierarchical Linear Modeling, and Factorial Analysis of Variance.

RETURN TO TABLE OF CONTENTS

Power Analysis for Proposal Writing

Category	Design and Application
Prerequisite Knowledge	Familiarity with regression models.
Description	Power analysis is a critical component of research planning that conveys the feasibility of achieving research goals with finite amounts of time and resources. This course will begin with estimating effect sizes and power analysis for conventional research designs. Next, the course will cover simulation-based methods for power analyses that can be used for virtually any data structure and research design, extending power analysis beyond the limited designs available in traditional power analysis software. The course will begin with strategies for research synthesis and effect size conversions that will form the basis of estimating power. We will use GPower to cover comparisons of means, comparisons of proportions, correlation, analysis of variance (ANOVA), repeated measures ANOVA, and regression models. Next, the course will cover simulation-based power analysis methods, using examples that might include nested data, auto-correlated data, and missing data. The presentation of power analyses in the context of proposal writing will be covered throughout the course. The course will also be useful for applications in meta-analysis and simulation studies.
Intended Audience	Anyone planning or involved with planning a research project. The course will be of interest to graduate students planning a proposal for a thesis or dissertation, faculty and research staff who are writing grant proposals, and consultants who assist with the development of research and grant proposals.
Computer Requirements	Power Analysis for Proposal Writing will be held in a computer classroom where participants will have access to the following software: R, Mplus, and GPower.
Time	9:00 AM – 12:00 Noon
Instructor	Nate Marti
Department	Psychology
Title	Research Associate
Bio 	Dr. Marti served as the manager of the statistical and mathematical consulting services with the Division of Statistics and Scientific Computation (DSSC) for 3.5 years and is the principal in a research consulting practice. His research and research collaboration has included topics in student engagement, persistence patterns in community college students, eating disorder prevention, and meta-analysis of program effectiveness. He has consulted on numerous grant proposals as an analytic consultant in which he has developed analytical plans and conducted power analyses.

RETURN TO TABLE OF CONTENTS

Statistics for the Dissertation

Category	Design and Application
Prerequisite Knowledge	None
Description	A comprehensive review of common statistical techniques for PhD students in non-mathematically leaning fields. We will cover methods that might be useful as they design their dissertations, such as t-tests, linear and multiple regression, various correlation equations (Pearson, Spearman, point-biserial), logistic regression, ANOVA, and ways to apply these in combination with qualitative research. An emphasis will be place on learning how to interpret the terms associated with these methods. The expected learning outcomes would be an increased awareness of and comfort with the mentioned statistical techniques, the ability to both read and comprehend studies using these methods, and knowledge of how to apply them to data relevant to their own areas of research.
Intended Audience	PhD students or candidates with a limited statistical background who want to enhance their statistical understanding of common techniques before or during the design of their thesis or dissertation.
Computer Requirements	None
Time	9:00 AM – 12:00 Noon
Instructor	Sarah Collins
Department	Educational Psychology, Department of Statistics & Data Sciences
Title	Statistics and Program Evaluation Lecturer
Bio	Sarah Collins is a Statistics and Program Evaluation lecturer in the Department of Educational Psychology and Statistics and the Department of Statistics & Data Sciences. She received her Ph.D. in Educational Psychology, Quantitative Methods, in 2010 at The University of Texas at Austin. She also serves as a statistical consultant for non-profit organizations around Austin.

RETURN TO TABLE OF CONTENTS

Structural Equation Modeling

Category	Statistical Methods
Prerequisite Knowledge	Knowledge of correlation and multiple regression methods.
Description	This course will build upon participants' previous knowledge of multiple linear regression by expanding to allow for correlated and causally related latent variables. This course assumes no prior experience with Structural Equation Modeling and is intended as both a theoretical and practical introduction. Topics covered in the course will include path analysis with measured variables, confirmatory factor analysis, structural equation models with latent variables, and a preview of more advanced models. The software package Mplus will be used for exploring and providing support for structural models. Participants will conduct hands-on practice exercises using Mplus software throughout the course.
Intended Audience	The intended audience includes graduate students, faculty, staff, and applied researchers in various disciplines, research consultants, and private industry researchers.
Computer Requirements	Participants should bring a personal laptop with basic Excel installed. Participants should also download and install a free Mplus demo version (or purchase a Mplus license) prior to the first day of the course; instructions will be provided.
Time	9:00 AM – 12:00 Noon
Instructor	Tiffany Whittaker
Department	Educational Psychology
Title	Assistant Professor
Bio	Tiffany Whittaker received her Ph.D. in Educational Psychology with a specialization in Quantitative Methods from The University of Texas at Austin in May 2003. She is an Associate Professor in the Department of Educational Psychology at The University of Texas at Austin. She teaches courses in quantitative methods, including statistical analysis for experimental data, data analysis using SAS, and structural equation modeling. Her research interests include structural equation modeling, multilevel modeling, and item response theory with a particular emphasis on model comparison/selection methods.

RETURN TO TABLE OF CONTENTS

Data Analysis using SAS

Category	Software and Database
Prerequisite Knowledge	Ability to navigate in a Windows environment and have taken an introductory statistics course that covered the following concepts: mean, standard deviation, normal distribution, t-tests, chi-square, regression, and ANOVA.
Description	The purpose of the course is to provide instruction in the use of SAS for conducting statistical analyses. Day one will cover opening and creating datasets, data manipulation, and t-tests. Days two and three will cover basic statistical analyses, including categorical analyses, two-sample tests, ANOVA, correlation and regression, and repeated measures analyses. Appropriate graphs will be taught along with the analyses. The basic statistics behind each type of analysis will be reviewed. Day four will cover special topics such as programming in SAS and working with sample data.
Intended Audience	Anyone who is interested in using SAS for data analysis.
Computer Requirements	Data Analysis using SAS will be held in a computer classroom where participants will have access to SAS.
Time	1:30 PM – 4:30 PM
Instructor	Matt Hersh
Department	Department of Statistics & Data Sciences
Title	Lecturer
Bio	Matt Hersh is a Lecturer in the Department of Statistics & Data Sciences at The University of Texas at Austin. He received his Ph.D. in Statistics from the University of Kentucky in 2007. While obtaining his degree, he was in the microarray core facility where he worked with researchers from various medical fields to help design and analyze their experiments. He also received a Master's degree from the LBJ School of Public Affairs at The University of Texas at Austin in 2000. As part of SDS's Graduate Fellows Program, Dr. Hersh assisted graduate students in analyzing data, preparing the results, and presenting conclusions for faculty members around campus. The statistical software packages he is most familiar with are SAS and R.

RETURN TO TABLE OF CONTENTS

Geospatial Data Analysis in R

Category	Statistical Methods
Prerequisite Knowledge	The main prerequisite is general ability to work with computers including running software and working with files and directories. Participants will progress more quickly if they have some experience with R or a similar environment like MATLAB. Some programming or scripting experience will also help but is not essential. Participants might wish to study basic concepts of Geographic Information Systems and complete one or more R tutorials. These resources are widely available on the World Wide Web.
Description	This course will cover how to use R as a GIS. Participants will gain a conceptual understanding of the different types of spatial data used in GIS and hand-on experience loading, displaying, manipulating and analyzing these data in R.
Intended Audience	Students and researchers interested in mapping and modeling spatial data using R, especially those who are initiating or have ongoing projects involving spatial analysis. Beginning graduate students will benefit by gaining a sound understanding of techniques for manipulating and analysis spatial data. Established researchers might also find the course valuable if they are making the transition from other spatial analysis platforms to R.
Computer Requirements	Geospatial Data Analysis in R will be held in a computer classroom where participants will have access to R. A preconfigured virtual-machine environment will be provided.
Time	1:30 PM – 4:30 PM
Instructor	Tim Keitt
Department	Department of Integrative Biology, Keittlab
Title	Associate Professor, Principal Investigator
Bio	Tim Keitt, Ph.D. is an Associate Professor in the Department of Integrative Biology within the College of Natural Sciences at The University of Texas at Austin. He studies complexity in the environment and works at the interfaces of landscape, population, community and ecosystem ecology. A major theme of his work is the influence of spatial heterogeneity on ecological processes. He is also a software developer and expert in R, C++ and SQL. Dr. Keitt authored the "rgdal" package exposing functions from the Geospatial Data Abstraction Library to the R language. This package is the top downloaded R package and is the basis of a large collection of dependent spatial data analysis packages for the R system.

RETURN TO TABLE OF CONTENTS

Introduction to Bayesian Statistics

Category	Statistical Methods
Prerequisite Knowledge	Knowledge of basic probability and statistics including estimation and hypothesis testing, plus some familiarity with maximum likelihood.
Description	This course will introduce participants to Bayesian statistics including the basic differences between Bayesian and Frequentist approaches as well as simple models, linear regression and generalized linear models, and hierarchical modeling. It will also cover modern simulation-based methods such as Gibbs sampling and briefly introduce participants to tools such as JAGS and STAN for the estimation of a wide array of models.
Intended Audience	Participants who have a basic understanding of introductory statistics including estimation and hypothesis testing as well as some exposure to maximum likelihood who are interested in learning about Bayesian methods.
Computer Requirements	None.
Time	1:30 PM – 4:30 PM
Instructor	Stephen Jessee
Department	Government
Title	Associate Professor
Bio	Dr. Stephen Jessee is an Associate Professor of Government in the College of Liberal Arts at The University of Texas at Austin. He received his Ph.D. from Stanford University and his B.A. and B.S. degrees from The University of Texas at Austin. Stephen teaches classes in American politics and statistical methodology, focusing on the measurement of ideology and voting behavior. His work has appeared in the American Political Science Review, American Journal of Political Science, Journal of Politics, and other outlets. Dr. Jessee takes interest in Bayesian statistics, ideal point estimation, and hierarchical models.

RETURN TO TABLE OF CONTENTS

Introduction to Data Science in Python

Category	Software and Database
Prerequisite Knowledge	There are no hard prerequisites. However, participants are likely to get more out of the course if they have (a) passing familiarity with basic statistical concepts and techniques (e.g., linear regression), and (b) minimal prior experience analyzing data in a command-line or scripting environment (e.g., R, Matlab, SAS, etc.).
Description	Modern data scientists have a bewildering array of tools at their disposal. In recent years, Python has emerged as a language of choice for many data scientists' due to its appealing combination of flexibility, power, and extensive community support. This short course surveys the Python software ecosystem and familiarizes participants with cutting-edge data science tools. Topics include: interactive computing basics, data preprocessing and cleaning, exploratory data analysis, visualization, and machine learning and predictive modeling. Participants will explore core concepts in data science and Python via hands-on, interactive exploration and analysis of sample datasets.
Intended Audience	This course is geared towards researchers and analysts who have had prior exposure to basic statistics or data science concepts and are interested in learning how to conduct state-of-the-art data analysis using open-source Python tools.
Computer Requirements	Participants should bring a personal laptop. A working installation of Python (version 2.7+ or 3+) is required. Course participants should make sure that they have a working Python installation on their laptop in advance of the course. Participants are strongly encouraged to install Python via the free Anaconda distribution, which has one-click installers for all major platforms (www.continuum.io/downloads), and includes most of the data science packages the course will cover.
Time	1:30 PM – 4:30 PM
Instructor	Tal Yarkoni
Department	Department of Psychology
Title	Research Assistant Professor
Bio 	Tal Yarkoni is a Research Assistant Professor in the Department of Psychology at The University of Texas at Austin and the director of the Psychoinformatics Lab. His research centers on the development of novel methods for the large-scale acquisition, organization, and analysis of psychological and neuroimaging data. He has over a decade of experience writing and applying Python code for data analysis, and has previously taught a thematically related and well-reviewed course (Introduction to Psychoinformatics) at the Summer Statistics Institute in 2014.

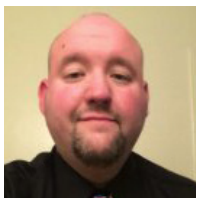
RETURN TO TABLE OF CONTENTS

Introduction to GIS

Category	Software and Database
Prerequisite Knowledge	Some statistics recommended. Familiarity with computers required.
Description	This course describes basic concepts underlying geographic information systems and science (GIS) and introduces participants to spatial analysis with GIS. Although the course will include hands-on laboratory exercises using ArcGIS software, the focus is on the "science behind the software" (e.g., types and implications of functions and analysis, rather than just how to do the analysis).
Intended Audience	This course should be of interest to anyone who uses spatial data and would like to learn about GIS and the types of analyses that can be done with it. In the past, employees of government agencies and organizations such as the health department, school boards, and city planning have attended.
Computer Requirements	Introduction to GIS will be held in a computer classroom where participants will have access to the required software available.
Time	1:30 PM – 4:30 PM
Instructor	Jennifer Miller
Department	Department of Geography and the Environment
Title	Associate Professor
Bio 	Dr. Miller is an associate professor in the Department of Geography and the Environment. She received a Ph.D. from a joint program at San Diego State University and UC-Santa Barbara. Her research focuses on GIScience and spatial analysis in general, and modeling biogeographical distributions and movements in particular.

RETURN TO TABLE OF CONTENTS

Introduction to SQL and Relational Database Design

Category	Software and Database
Prerequisite Knowledge	Knowledge of computer use.
Description	This course will teach interested parties the basics of relational database design and Structured Query Language (SQL). Participants will have the opportunity to design their own database, as well as learn how to input and extract data using SQL. The course will focus on best practices of relational database design as well as provide a broad overview of the different types of queries used to retrieve data from a relational database. Technology used will include Microsoft Access and Microsoft SQL Server; however, the material taught in this course can be applied to many different technology platforms.
Intended Audience	People who are interested in learning about relational database: how to use them, and how to input, retrieve, and analyze data using Structured Query Language (SQL).
Computer Requirements	Introduction to GIS will be held in a computer classroom where participants will have access to the required software available.
Time	1:30 PM – 4:30 PM
Instructor	Chris Golubski
Department	Department of Statistics & Data Sciences
Title	Lecturer
Bio 	Chris Golubski is a doctoral student in mathematics education at The University of Texas at Austin, specializing in statistics education. He is also simultaneously pursuing a master's degree in statistics. He currently holds a Master of Science in Mathematics and teaches at several local colleges in Austin with over 15 years of educational and professional experience in mathematics and computer science. Chris also does IT consulting and software development in the area.

RETURN TO TABLE OF CONTENTS

Introduction to Stata

Category	Software and Database
Prerequisite Knowledge	Participants should have the ability to navigate in the operating system environment of their choice (Windows, Mac, or Linux) and knowledge equivalent to that from an introductory statistics course covering p-values, confidence intervals, t-tests, ANOVA, and correlations.
Description	The purpose of the course is to provide instruction in the use of Stata for data handling and for conducting statistical analyses. Day one will provide an overview of the software, information on basic data handling and manipulation, and exploratory descriptive analyses. Days two and three will cover basic inferential analyses including chi-square tests, t-tests and ANOVA, and regression including the use of bootstrapping. Also covered in this section are principal components/factor analysis and related techniques used in scale construction. Throughout, the use of appropriate graphical techniques will be addressed and the basic theory behind each type of analysis will be reviewed. Day four will feature more advanced categorical analysis via binary and multinomial logistic regression. Coverage in this area will include the implementation of likelihood ratio testing in Stata. There will also be a brief introduction to Stata's programming capabilities for custom needs, and coverage of Stata's capabilities in structural equation modeling. After taking this class, participants will have excellent foundational knowledge of this software tool, and should have no trouble building on that foundation as needed by learning how to use Stata for other basic analyses not directly covered in the class and/or learning how to use Stata for more advanced or specialized techniques.
Intended Audience	The intended audience is anyone with knowledge of basic inferential statistics who wants to learn about Stata's capabilities and about how to use Stata to perform a wide variety of common analyses.
Computer Requirements	Participants should bring a personal laptop. Installation of Stata should be completed prior to the first day of class; instructions will be provided.
Time	1:30 PM – 4:30 PM
Instructor	Greg Hixon
Department	Psychology
Title	Professor

RETURN TO TABLE OF CONTENTS

Bio

Dr. Hixon received his Ph.D. from The University of Texas in 1991. In the more than two decades since, he has served on the faculties of the University of Connecticut and the University of Texas at Austin, and has worked with a variety of governmental agencies and corporations in the areas of statistics, applied mathematics, and computational analytics. He currently teaches four Ph.D. courses at the University of Texas at Austin, spanning the range from basic approaches like ANOVA and linear regression to more advanced techniques such as multivariate non-parametric modeling, simulation methods, and structural equations.

RETURN TO TABLE OF CONTENTS

Introduction to Statistics (PM)

Category	Statistical Methods
Prerequisite Knowledge	Absolutely no previous knowledge of statistics is necessary or expected. However, participants should be comfortable working with spreadsheets in Microsoft Excel (either the Mac or PC version). Those who have never used Excel should prepare before coming to SSI, as a basic familiarity with the program will be assumed.
Description	This hands-on course will introduce participants to common descriptive and inferential statistical analyses. In addition to covering the concepts behind each method, we will also practice applying them on real datasets using Microsoft Excel. Sufficient time will be spent on understanding relevant assumptions and how to correctly interpret the results of each analysis. The specific topics covered in this course include: describing and visualizing data, t-tests, ANOVA, chi-squared test of independence, correlation, and linear regression. Optional "homework" will be offered after each class day for those who want additional practice applying the techniques discussed.
Intended Audience	This course is designed for those with little to no experience in statistics and who want use descriptive and inferential methods to analyze data. Whether coming from academia, industry, or government, participants in this course will learn the skills needed to help them better understand the data that they work with.
Computer Requirements	All participants will need a version of Excel from 2013 or newer. For PC, version 2013 or 2016 is ok, Mac users MUST have Excel 2016 (most recent version). The University of Texas at Austin students and staff can download Excel 2016 for free through campus resources.
Time	1:30 PM – 4:30 PM
Instructor	Steven Hernandez
Department	Department of Statistics & Data Sciences
Title	Lecturer
Bio	Steven Hernandez is a native Austinite. He received a B.A. in Mathematics from The University of Texas at Austin in 2008, and Master's in Statistics in 2015. He is a former high school math teacher and currently a lecturer for Introduction to Market Analysis and Biostatistics at the University of Texas at Austin.

RETURN TO TABLE OF CONTENTS

Large Scale Data Analysis with Hadoop and Spark

Category	Statistical Methods
Prerequisite Knowledge	Participants should have basic working knowledge of the Linux operating system and of using a command line interface. Participants are also expected to have at least introductory level of education in computer programming, such as knowledge on data structure, and control flow. Experience and working knowledge on at least ONE of the following Java, Scala, Python, R, SQL are preferred.
Description	This course will introduce participants to using the two most popular big data processing frameworks, Hadoop and Spark, for big data analysis tasks. The course will introduce basic system architecture and core components of each system in order to give beginners a clear picture on basics of the two systems. The course will feature clear instructions and a test system access for participants to get started on using those systems from day one. The course will give a grand tour of the data analysis capability to show how common data analysis needs for large data can be met with those platforms. Useful libraries and existing tools will also be introduced including Mahout, MLib, GraphX and SparkSQL. Those tools and libraries include a set of implementations of a wide range of analysis algorithms. Finally, the course will also introduce components and applications that enable utilization of the Hadoop and Spark through other programming language and interface including Hadoop Streaming, Spark-Shell and Hive. The course materials will include exemplar problems, hands-on exercises, and demonstrations.
Intended Audience	This course is intended for people who are interested in learning more about available tools and solutions to support large scale data analysis. Students and professionals who are facing the scalability issue with data driven problems are welcome in this course.
Computer Requirements	Participants should bring a personal laptop. Installation of Java 1.8 and Secure Shell Client should be completed prior to the first day of class.
Time	1:30 PM – 4:30 PM
Instructor	Weijia Xu
Department	Texas Advanced Computing Center (TACC)
Title	Research Engineer / Scientist Associate Manager

RETURN TO TABLE OF CONTENTS

Bio

Dr. Weijia Xu is a research scientist and the group manager for Data Mining & Statistics group at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin. He has a Ph.D. in Computer Science and a M.S. degree in Life Science from The University of Texas at Austin. Dr. Xu's main research interest is to enable data-driven discoveries through developing new computational methods and applications that facilitate the data-to-knowledge transfer process. Dr. Xu has over 50 peer-reviewed conference and journal publications in similarity-based data retrieval, data analysis, and information visualization with data from various scientific domains. He has served on program committees for several workshops and conferences in big data and high-performance computing area, most recently, co-chair for IEEE Conference on Big Data in 2015 and 2016. He also has been a guest editor for Journal of Big Data Research since 2015. Dr. Xu's group is also responsible in support two other computing resources dedicated to support data intensive workflow such as those requires Hadoop and Spark programming paradigm.

RETURN TO TABLE OF CONTENTS

Non-Parametric Statistical Methods for Small Datasets

Category	Statistical Methods
Prerequisite Knowledge	Familiarity with basic statistical concepts will be useful. For example, students should know the basics of probability, random variables, descriptive statistics, and hypothesis testing. Prior knowledge of parametric statistical tests and probability distributions is not required, but that knowledge will enable participants to compare and contrast the non-parametric methods they will learn in this course.
Description	The objective of this course is to discuss the non-parametric equivalents for most of the common statistical tests that are typically taught in introductory statistics courses. These tests come into play either when the assumptions of the parametric tests don't hold, or when sample sizes are too small to assess validity of assumptions. Topics will include the non-parametric equivalents to the t-tests for means, chi-square tests, correlation, regression, and ANOVA, with examples using R. Bootstrapping, kernel smoothing, and spline regression will be discussed if there is time. Guidelines and decision tables will be provided to facilitate the selection of the appropriate test for each scenario, and advantages and drawbacks of each method will be discussed. Problem sets will be provided for practice.
Intended Audience	Anyone dealing with small data sets or data sets that don't typically adhere to the assumptions needed for parametric methods to be applicable will find this course useful. These types of methods are quite often relevant in the fields mathematics, statistics, engineering, manufacturing, business, education, social, biological, and environmental sciences.
Computer Requirements	Participants should bring a personal laptop. The instructor will provide examples primarily in R, but may also provide examples in SAS and SPSS. Installation of R should be completed prior to the class; instructions will be provided.
Time	1:30 PM – 4:30 PM
Instructor	Bindu Viswanathan
Department	Statistics & Data Sciences
Title	Lecturer

RETURN TO TABLE OF CONTENTS

Bio

Dr. Viswanathan is a lecturer in the Department of Statistics & Data Sciences. Before coming to The University of Texas at Austin, she worked as research faculty at Emory University, as the statistical lead on numerous research projects in the schools of Nursing, Medicine, and Public Health, as well as at the CDC and VA Hospital. She has also worked as a Biostatistician at Merck & Co. and Novartis Ophthalmics, designing and overseeing Phase III clinical trials. She received her Ph.D. in Biostatistics from Emory University in 1999, and also has a Master's degree in Conservation Biology from TX State University. At UT Austin, she primarily teaches Biostatistics, where she draws from her experiences to motivate students to see the practical applications of concepts taught in class.

RETURN TO TABLE OF CONTENTS

Questionnaire Design and Survey Analysis

Category	Design and Application
Prerequisite Knowledge	An introductory social research class would be helpful but is not necessary.
Description	The goal of this course is to introduce participants to the construction and analysis of social surveys. In the first part of the course, participants will be taught the tools needed to create effective and reliable questions, craft questionnaires that could be used in multiple settings (e.g., telephone, written, web-based), test questionnaires to ensure their effectiveness, design implementation strategies that will increase the likelihood of good response rates. By the end of the course participants will know the basics of designing and fielding a survey that could be used for research or other purposes.
Intended Audience	The course is primarily oriented towards graduate students, faculty, and others in the community who want a comprehensive introduction to survey design and implementation.
Computer Requirements	None Required.
Time	1:30 PM – 4:30 PM
Instructor	Marc Musick
Department	Sociology
Title	Professor and Associate Dean in the College of Liberal Arts
Bio	Marc Musick received his Ph.D. in Sociology from Duke University, and then trained for two years as a postdoctoral fellow in the NIMH Postdoctoral Training Program on Psychosocial Factors and Mental Health at the Survey Research Center. His research examines the social production of pro-social activity and the consequences of that activity.

RETURN TO TABLE OF CONTENTS

The Power and Pleasure of Probability

Category	Design and Application
Prerequisite Knowledge	No prior knowledge of probability is necessary. Some demonstrations will be done using R, but participants are not required to have any coding knowledge.
Description	Participants will learn fundamental rules for computing probabilities, including the explanations behind some famous paradoxical puzzles, gain insight into statistical practice (including the frequentist vs. Bayesian debate) through a deeper understanding of connections with probability theory, dispel misconceptions and cognitive biases surrounding randomness, and explore simulation as a tool for problem solving and as a means to understand limit theorems.
Intended Audience	This course is for everyone! Humans are not born well-equipped to understand random phenomena. But with some mathematical ground rules, and a bit of practice, we can attain a deeper understanding and appreciation of our unpredictable world. Recommended for researchers, data analysts, statisticians, gamblers, doctors, lawyers, journalists, judges, politicians, policy wonks, conspiracy theorists, athletes, actuaries, poets, philosophers, etc.
Computer Requirements	None
Time	1:30 PM – 4:30 PM
Instructor	Joel Nibert
Department	Mathematics
Title	Lecturer
Bio	Joel Nibert received his Ph.D. from the University of Southern California in 2012 for research in probability and stochastic processes. He joined the faculty of The University of Texas at Austin in 2013. He teaches a variety of math courses including probability, statistics, calculus, introduction to mathematics, and actuarial mathematics. Joel enjoys jazz music and games of strategy.

RETURN TO TABLE OF CONTENTS

Time Series Forecasting and Modeling

Category	Statistical Methods
Prerequisite Knowledge	Students should be comfortable with the use and interpretation of linear regression with multiple predictor variables: Plugging values into the regression equation, interpretation of coefficients, significance of predictors, R-square, root mean-squared error, correlation, etc. Students should also be familiar with Excel. Some use of logarithms and exponentials will be made. Some familiarity with SAS would be desirable, but I will include a tutorial to make students quickly productive in SAS. Calculus is not necessary. Appropriate readings will be provided before the course.
Description	This course will teach you a practical approach to modeling time series data. The goal of modeling is to explain and to predict: to account for why a phenomenon varies over time and to predict its future. The course focus is on empirical modeling, rather than theoretical properties. You will learn how to propose models, estimate them with data, diagnose whether they fit, and interpret their meanings. Models covered include random samples, random walks, regression, autoregression, moving averages, and related structures. Computer demonstrations with both real and simulated data will be used extensively.
Intended Audience	The course is intended to be immediately useful for anyone (UT students, faculty, administrative staff, state agency employees, private company employees, consultants, etc.) who has a time series dataset sitting on his/her desk that he/she needs to understand and/or forecast. The course will provide a general-purpose method that the student, on his/her own, can use to fit a model to the data, diagnose whether the model fits, and use the model to understand the data and forecast future values. The course is not intended to provide exposure to a wide variety of specialized models, but rather to provide a few widely applicable general-purpose tools.
Computer Requirements	Participants should bring a personal laptop that runs Microsoft Excel. (Mac users should note that the Excel files used in the course are prepared in Excel for Windows. Such files often run most smoothly on a Mac that is running a Windows emulator.) Participants will have access to SAS-on-Demand, the free, cloud version of SAS. There is no software to download for any part of this course, including SAS-on-Demand, but you do need to register with SAS. Details will be provided.
Time	1:30 PM – 4:30 PM
Instructor	Tom Sager
Department	McCombs

RETURN TO TABLE OF CONTENTS

Title	Professor

Bio

Tom Sager was raised and educated in Iowa. He served in the Army as a trumpet player during the Vietnam War. After getting his Ph.D. in Statistics from the University of Iowa, he practiced the art of professing at Stanford University and The University of Texas at Austin, and someday might get it right. Attracted to statistics because he thought it would allow him to avoid specializing, he has published articles in leading statistics and applied journals that span the gamut from very applied to very theoretical. He has dabbled in statistics in insurance companies, mathematics, air pollution, law, auditing, and quality. Tom's current research interests focus on econometric analysis of insurance companies. He has just completed a three-year project to develop models for forecasting financial crises and stress-testing European banks. Tom has consulted extensively for insurance and re-insurance companies, lawyers, government agencies, large and small corporations, and consulting firms. His primary teaching responsibilities include the core statistics course in the MBA curriculum and econometrics for doctoral students. Tom has won the Joe D. Beasley Award for teaching excellence in the MBA program and recently was selected by students as outstanding professor in the Masters in Business Analytics program. Currently Professor of Statistics in the IROM Department, Tom just loves statistics in all its ubiquity.

RETURN TO TABLE OF CONTENTS