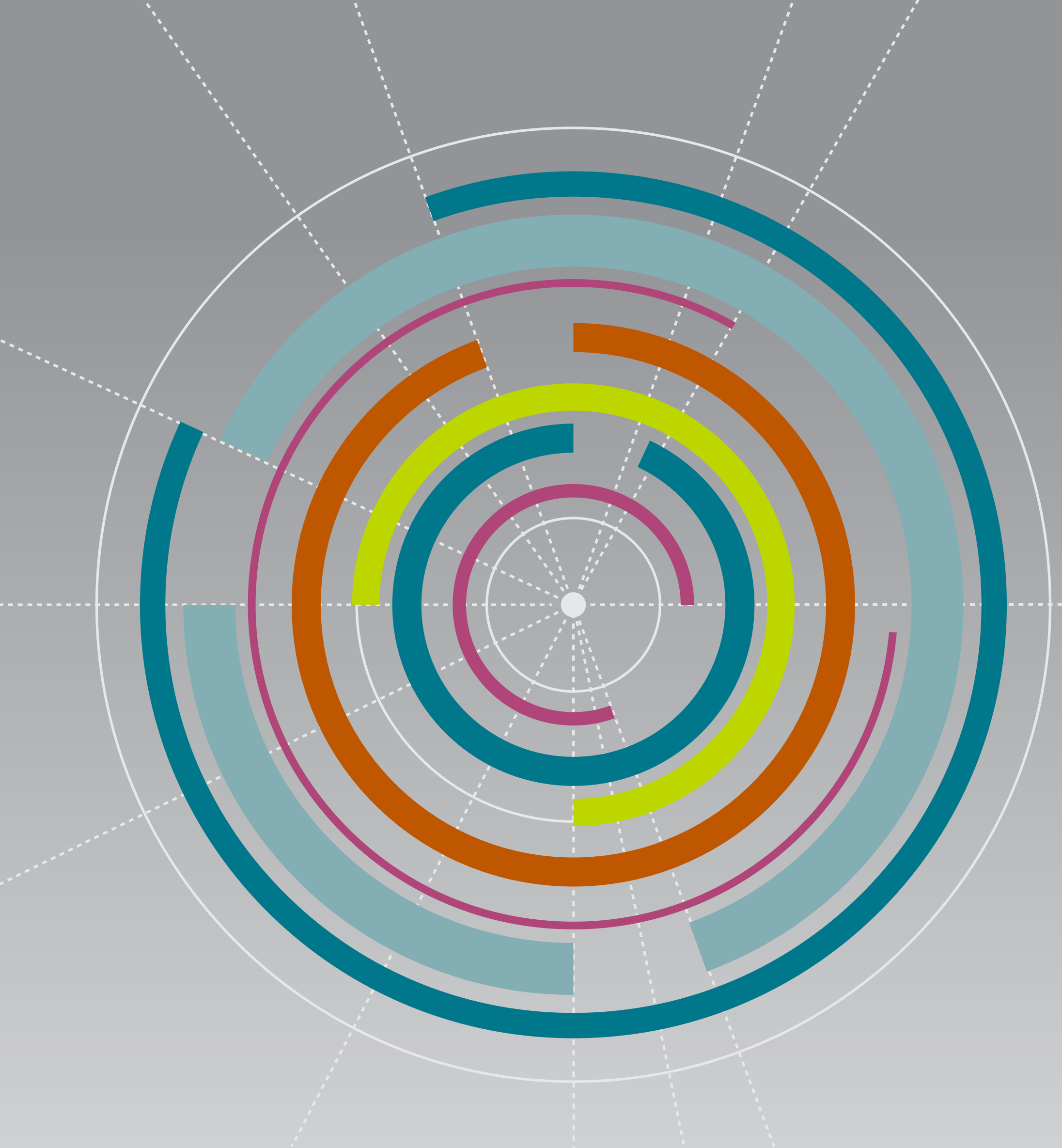


DEPARTMENT OF STATISTICS AND DATA SCIENCES

UT Summer Statistics Institute

The University of Texas at Austin
May 23–26, 2016

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 9th Annual 2016 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 2016 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 2016 Summer Statistics Institute will offer 26 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.)
Introductory Statistics	Introduction to Statistics	
Software	Introduction to MATLAB	Data Analysis Using SAS
	Introduction to SPSS	Introduction to Data Analysis, & Graphics using R
		Introduction to Stata
Statistical Methods & Applications	Common Mistakes in Using Statistics: Spotting Them & Avoiding Them	Geospatial Data Analysis in R
	Data Science in Industry with R	Introduction to Bayesian Statistics
	Event History Analysis	Introduction to Data Science in Python
	Hierarchical Linear Modeling	Introduction to GIS
	Introduction to Mixed Models with Applications	Introduction to SQL and Relational Database Design
	Introduction to Regression	Non-Parametric Statistical Methods for Small Datasets
	Making Sense of Multivariate Data: Principal Component Analysis, Factor Analysis, & Clustering Analysis	Questionnaire Design and Survey Analysis
	Power Analysis for Proposal Writing	Time Series Modeling
	Structural Equation Modeling	
Big Data	Introduction to Big Data Analytics	Big Data Analytics: Structured, Semi-Structured, and Unstructured
		Introduction to Methods and Tools for Large Scale Data Driven Analysis

RETURN TO TABLE OF CONTENTS

NEW FOR 2016!

Data Science in Industry with R

This course will cover some practical data science tasks found in industry. Topics will include: connecting to databases, parsing XML and JSON data, data wrangling, building web applications with shiny, and making predictive models. Participants will be introduced to several commonly used R packages.

Introduction to Data Science in Python

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 and visualization; and machine learning and predictive modeling.

Introduction to SQL and Relational Database Design

This course will teach the basics of relational database design and Structured Query Language (SQL). Participants will design their own database, and learn how to input and extract data using SQL. The course will also focus on best practices of relational database design as well as 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.

Non-Parametric Statistical Methods for Small Datasets

The objective of this course is to discuss the non-parametric equivalents for most of the common statistical tests. Topics will include the non-parametric equivalents to the t-tests for means, chi-square tests, correlation, regression, and ANOVA, with examples using R. 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.

RECENT ADDITIONS TO SSI

Geospatial Data Analysis in R

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 hands-on experience loading, displaying, manipulating, and analyzing these data in R.

Introduction to MATLAB

This course will provide an introduction to MATLAB as well as a survey of intermediate topics including data manipulation, analysis, and visualization, and MATLAB programming. Working with example datasets from different disciplines, students will learn to use MATLAB through hands on tutorials. The course will begin with an introduction to the MATLAB desktop interface and command line tools, importing/exporting data, and elementary descriptive statistics and data visualization. Day two will cover more advanced methods of working with data including sorting and restructuring data, creating a variety of 2-D and 3-D plots, and performing basic inferential statistics. Day three will focus on using MATLAB as a computational tool.

Introduction to Mixed Models with Applications

This course would be beneficial to people who work in medicine, public health, clinical trials, or other medical fields as well as for researchers in the pharmaceutical industry. People who are interested in longitudinal data analysis are particularly encouraged to sign up for the course. Researchers in social sciences who are interested in multilevel models will also find the course relevant.

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 is ***February 2, 2016 – May 6, 2016*** and fees are as follows:

Category	Registration Fees (per course)	
UT-Austin	Students *	$175
	Faculty/Staff *	$275
UT-System	Faculty/Staff *	$275
Non-UT-Austin	Students **	$225
	Participants	$550
	Groups of five or more from the same institution or agency (fee is per person per course)	$440

**Our staff will confirm affiliations with UT.*
***Non-UT students must send us a copy of their current student identification.*

Contact our office at (512) 232-0693 for 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.

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 contact our office at (512) 232-0693. When paying by IDT, do not enter a discount code. The discount code will be entered manually through our office. You will receive a receipt via email with the discounted price.

REFUND AND CANCELLATION POLICY

A full refund of registration fees, less a $25 cancellation fee, will be available if requested in writing and received by March 31, 2016. No refunds will be made after that date.

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.

RETURN TO TABLE OF CONTENTS

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

Morning (9:00AM – 12:00 NOON)

Common Mistakes in Using Statistics: Spotting Them and Avoiding Them

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. Some acquaintance with transformations of random variables (especially the logarithm), Analysis of Variance, and multiple regression might be helpful, but is not necessary. However, 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 Required.
Time	9:00 AM – 12:00 noon
Instructor	Martha K. Smith
Department	Mathematics
Title	Professor Emerita
Bio 	Martha Smith was a professor in The University of Texas at Austin math department from 1973 to 2009. In her last ten years before retiring as Professor Emerita, her teaching focused mainly on statistics, both graduate and undergraduate. As a mathematician, she had the habit of asking "Why is this so?" and looking for errors in reasoning. She became aware that statistical techniques are often misunderstood and misapplied. Since retiring, she has kept active professionally, including creating a website on the topic of this SSI course, blogging on uses and misuses of statistics, serving on Ph.D. committees, attending seminars in statistics and biology, trying to keep up to date on statistics books and journals as they come into the library, and writing the occasional article or review.

RETURN TO TABLE OF CONTENTS

Data Science in Industry with R

Prerequisite Knowledge	There is no prerequisite. Content will emphasize practical usage of R.
Description	This course will cover some practical data science tasks found in industry. Topics will include: connecting to databases, parsing XML and 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	Anyone interested with using R for industrial problems or in a commercial setting.
Computer Requirements	Participants should bring a personal laptop. Installation of R and RStudio should be installed prior to the first day of class.
Time	9:00 AM – 12:00 noon
Instructor	Richard Leu
Company	Dropoff
Title	Data Scientist
Bio	Richard Leu currently works as a data scientist for Dropoff applying statistics, machine learning, and operations research to same day logistics. 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. Richard received a PhD in physics and an MS in statistics from The University of Texas at Austin.

RETURN TO TABLE OF CONTENTS

Event History Analysis

Prerequisite Knowledge

Participants should have a background in linear regression and should have experience with at least one computer package such as Stata, SAS or R.

Description

This course will introduce statistical methods for the analysis of event history data, where events are outcomes of interest that occur over the lifetimes of the individuals or objects being studied. The course will progress from methods for exploratory analysis to statistical models that introduce explanatory variables and unmeasured heterogeneity. The class will provide clear explanations of the basic statistical underpinnings of the methods and models and will illustrate example analyses in Stata and R through a set of daily hands-on tutorials using data from several substantive research areas.

Intended Audience

Graduate students, faculty, and others seeking to learn more about or apply event history modeling in their work.

Computer Requirements

Participants should bring a personal laptop. Instructions for installation of Stata and R will be sent before the first day of the course.

Time

9:00 AM – 12:00 noon

Instructor

Dan Powers

Department

Sociology

Title

Professor

Bio

Dan Powers is a professor in the Department of Sociology at The University of Texas at Austin with interests in social demography and statistics. He is also research associate with the Population Research Center where he has participated in funded research over the past 21 years. He has interests in statistics and methods, fertility, mortality, social inequality, and health disparities. His research examines issues relating to the Hispanic epidemiological paradox in infant mortality, temporal change in infant mortality, multivariate decomposition methodology for hazard rates, and statistical methods for adjusting life tables and survivor functions. Dr. Powers plays a key role in statistics and methods training in the Department of Sociology and at the Population Research Center by teaching graduate courses in categorical data analysis and longitudinal data analysis. He has served on over 100 MS and Ph.D. committees, and is former graduate advisor and GSC chair of the Master's of Science in Statistics program in the Department of Statistics and Data Sciences at The University of Texas at Austin.

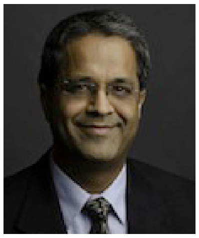
RETURN TO TABLE OF CONTENTS

Hierarchical Linear Modeling

Prerequisite Knowledge	Participants should be comfortable with the use and interpretation of multiple regression. In particular, participants should be familiar with the use of dummy-coding for binary independent variables, the use of product variables to incorporate interactions, and the use of polynomial regression to model nonlinear relationships. Prior exposure to logistic regression is helpful, but not necessary.
Description	The purpose of the workshop is to help participants to begin to learn how to analyze multilevel data sets and interpret results of multilevel modeling analyses. Organizational analysis and growth curve modeling, the most common multilevel modeling applications, are featured in the workshop. Further, using data sets provided in the workshop, workshop participants will learn how to use the HLM software program to obtain analysis results. Additionally, the workshop will emphasize proper interpretation of analysis results and illustrate procedures that can be used to specify multilevel models. Coverage of multilevel models for binary outcomes will also be included.
Intended Audience	Graduate students, applied researchers, and faculty who wish to learn about HLM, particularly as it is used in social science research.
Computer Requirements	"Hierarchical Linear Modeling" will be held in a computer classroom with HLM and SPSS software available for access.
Time	9:00 AM – 12:00 noon
Instructor	Keenan Pituch
Department	Educational Psychology
Title	Associate Professor
Bio 	Dr. Keenan Pituch is an Associate Professor of Quantitative Methods in the Department of Educational Psychology at the University of Texas at Austin. His research interests include multilevel modeling, multivariate analysis, and mediation analysis

RETURN TO TABLE OF CONTENTS

Introduction to Big Data Analytics

Prerequisite Knowledge	None required, but introductory courses in statistics (or data mining) and computing would be helpful.
Description	Data is being generated at a tremendous rate in modern applications that are as diverse as social network analysis, genomics, health care, energy management, and computer vision. It is now widely accepted that very large-scale data sets will be collected and will need to be analyzed in the context of these applications. Indeed, the recent moniker of Big Data emphasizes that massive volumes of data are ubiquitous. This short course will cover analysis tasks that can be performed on data sets, such as regression, classification, clustering and dimensionality reduction, and discuss ways of scaling these methods so that very large data sets can be analyzed. The course will also discuss systems aspects that arise in such large-scale analysis, for example, multi-core versus multi-machine, Hadoop/ MapReduce versus MPI and synchronous versus asynchronous computations. The vibrant application areas of social networks analysis, text analysis, recommender systems and genomics will be used to illustrate uses of big data analytics.
Intended Audience	The intended audience is anyone interested in learning how to analyze very large data sets.
Computer Requirements	Participants should bring a personal laptop.
Time	9:00 AM – 12:00 noon
Instructor	Inderjit S. Dhillon
Department	Computer Science
Title	Professor
Bio 	Inderjit Dhillon is a Professor of Computer Science and Mathematics at The University of Texas at Austin, where he is the Director of the ICES Center for Big Data Analytics. His main research interests are in big data, machine learning, network analysis, linear algebra and optimization. He received his B.Tech. Degree from IIT Bombay, and Ph.D. from UC Berkeley. Inderjit has received several prestigious awards, including the ICES Distinguished Research Award in 2013, the SIAM Outstanding Paper Prize in 2011, the Moncrieff Grand Challenge Award in 2010, the SIAM Linear Algebra Prize in 2006, the University Research Excellence Award in 2005, and the NSF Career Award in 2001. He has published over 100 journal and conference papers, and has served on the Editorial Board of the Journal of Machine Learning Research, the IEEE Transactions of Pattern Analysis and Machine Intelligence, Foundations and Trends in Machine Learning and the SIAM Journal for Matrix Analysis and Applications. Inderjit is an IEEE Fellow and a SIAM Fellow.

RETURN TO TABLE OF CONTENTS

Introduction to MATLAB

Prerequisite Knowledge

No prior programming knowledge or experience with MATLAB is required. Participants should be comfortable working with data in .xls, .csv, or other text file formats, and a basic understanding of common statistical concepts and methods is recommended. Those with prior experience in MATLAB will benefit from the course and are encouraged to attend.

Description

This course will provide an introduction to MATLAB as well as a survey of intermediate topics including data manipulation, analysis, and visualization, and MATLAB programming. Working with example datasets from different disciplines, students will learn to use MATLAB through hands on tutorials. The course will begin with an introduction to the MATLAB desktop interface and command line tools, importing/exporting data, and elementary descriptive statistics and data visualization. Day two will cover more advanced methods of working with data including sorting and restructuring data, creating a variety of 2-D and 3-D plots, and performing basic inferential statistics. Day three will focus on using MATLAB as a computational tool. Topics covered will include linear algebra with matrices, Fourier analyses, and using probability distributions, as well as writing MATLAB scripts and functions for complex computations. Day four will feature more advanced topics of scientific computation in MATLAB including simulations, bootstrapping techniques, and pattern analytics.

Intended Audience

This course is designed for those interested in using MATLAB for data management and analysis, visualizing data and analytic results, and more advanced computational processing of large datasets.

Computer Requirements

"Introduction to MATLAB" will be held in a computer classroom where participants will have access to the software.

Time

9:00 AM – 12:00 noon

Instructor

Michael Mack

Department

Center for Learning and Memory, Department of Psychology

Title

Research Affiliate

Bio

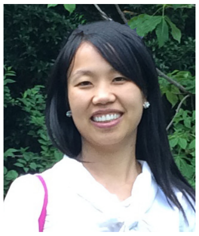
Dr. Mack is a postdoctoral researcher in the Department of Psychology and Center for Learning and Memory at The University of Texas at Austin. He received his Ph.D. in Psychology from Vanderbilt University and a B.S. and M.S. in Computer Science from Michigan State University. His research is focused on computational modeling of learning and memory processes underlying human cognition utilizing computational and neuroimaging methods.

RETURN TO TABLE OF CONTENTS

Introduction to Mixed Models with Applications

Prerequisite Knowledge	Some basic knowledge of statistical inference such as linear regression, Anova and hypothesis testing is desired. A basic understanding of multivariate statistics would be a plus although not required.
Description	Mixed modes are models that can incorporate both fixed and random effects. The last few decades have witnessed a vast growth of mixed modeling in many fields of study in particular in biomedical fields and public health sciences such as in medicine. Mixed models are powerful tools in analyzing correlated data such as, repeated measurements, longitudinal data, hierarchical and clustered data etc. The ability to incorporate random effects in addition to fixed effects allows mix models to model different sources of variation such as within- and between-subject variation, thus providing more accurate and broader scope of inference in general. This course provides an introduction and overview to mixed modeling with a focus on linear mixed models and generalized linear mixed models. The main goal of the course is to equip participants with a basic understanding of the historical development of mixed models as well as basic knowledge, theory, estimation methods and algorithms necessary for making practical inference. The content of the course includes: (1) An overview to linear regression (linear models) (2) An introduction to linear mixed models. (3) An introduction to generalized linear mixed models (4) Applications of mixed models in clinical trials, medicine and other fields (illustrated through SAS procedures).
Intended Audience	This course would be beneficial to people who work in the field of medicine, public health, clinical trials, or other medical fields as well as for researchers in the pharmaceutical industry. Participants who are interested in longitudinal data analysis are particularly encouraged to sign up for the course. Researchers in social sciences who are interested in multilevel models will also find the course relevant.
Computer Requirements	None Required.
Time	9:00 AM – 12:00 noon
Instructor	Lizhen Lin
Department	Department of Statistics and Data Sciences
Title	Assistant Professor

Bio



Dr. Lizhen Lin is an assistant professor in the Department of Statistics and Data Sciences at University of Texas Austin and has extensive experience teaching math and statistics courses. Dr. Lin received a PhD in Mathematics in 2012 from University of Arizona under the guidance of Rabi Bhattacharya. Prior to that, she was an undergraduate studying Mathematics and Statistics in Sichuan University from 2002-2006. Dr. Lin spent two years as a two-year postdoc from 2012 to 2014 at Duke University mainly working with David Dunson and has also been a member of the Laboratory for Psychiatric Neuroengineering in Duke University medical center since early 2013 working on neuro-psychiatric research.

Introduction to Regression

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 primarily use Stata14, but corresponding examples will at times be shown from other software packages such as SAS, SPSS, and R.
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. Instructions for installation of Stata will be sent prior to the first day of the course.
Time	9:00 AM – 12:00 noon
Instructor	Michael J. Mahometa
Department	Department of Statistics and 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 SDS 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 SPSS

Prerequisite Knowledge	Participants should have completed an introductory statistics course within the last two years that included t-tests, ANOVA, and ideally correlation and regression. Participants should know what the following terms mean: mean, standard deviation, p-value, and frequency. If students do not meet the prerequisite, they should read up on the above topics so as to be familiar with them.
Description	This course will teach participants how to perform descriptive and inferential statistics on data in SPSS. Participants will also learn how to perform basic data manipulations within SPSS.
Intended Audience	Anyone who is interested in using SPSS for data analysis.
Computer Requirements	"Introduction to SPSS" will be held in a computer classroom with SPSS software available for access.
Time	9:00 AM – 12:00 noon
Instructor	Lindsey Smith
Department	Department of Statistics and 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.

Introduction to Statistics

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 I will assume a basic familiarity with the program.
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	Participants should bring a personal laptop, Mac or PC with Excel 2016 installed.
Time	9:00 – 12:00 noon
Instructor	Sally Amen
Department	Statistics and Data Sciences
Title	Instructor, Consultant
Bio 	Sally received her Master of Science degree 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 The University of Texas at Austin graduate students, faculty, and staff who need help with study design, data management, running appropriate statistical analyses, and interpreting the results. She is also an instructor for SDS 328M Biostatistics, an undergraduate introductory stats course, and she teaches various software and topic short courses each semester.

RETURN TO TABLE OF CONTENTS

Making Sense of Multivariate Data: Principal Component Analysis, Factor Analysis and Cluster Analysis

Prerequisite Knowledge

Introduction to statistics: Participants should have some knowledge of variance, correlation, regression, and hypothesis testing

Description

This workshop is designed to provide participants experience using statistical methods that can help them make sense of data when there are a large number of variables and/or cases. The workshop will first cover the basic principles of constructing and testing multivariate statistical models. Next, participants will be introduced to three fundamental multivariate methods: principal component analysis, factor analysis, and cluster analysis. Besides having practical utility, the three methods provide an essential background for learning other multivariate techniques in the future. Participants will gain experience applying each of the three methods on real datasets with SAS statistical software.

Intended Audience

Individuals that are interested in:

- Learning about basic principles of multivariate data analysis to apply in their own data analysis efforts as well as to foster their learning other techniques.
- Acquiring experience in analyzing and interpreting selected multivariate datasets using the following fundamental multivariate techniques: principal component analysis, factor analysis, and cluster analysis.
- Learning or reviewing how to use SAS for data analysis.
- Hands-on experience using SAS to carry out multivariate analyses.

Computer Requirements

"Making Sense of Multivariate Data" will be held in a computer classroom with SAS software available for use.

Time

9:00 AM – 12:00 noon

Instructor

Jerry Manheimer

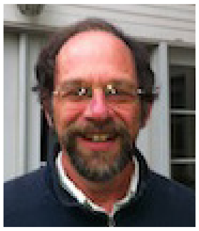
Department

Human Development and Family Science

Title

Lecturer

Bio



Jerry Manheimer has taught courses in statistics and human development for several years at The University of Texas at Austin. His educational background is in cognitive psychology, applied statistics, and research methods. Besides working in academia, Dr. Manheimer worked for many years in the high-tech industry where he used multivariate classification techniques in a number of practical applications and developed special-purpose software for multivariate classification applications.

RETURN TO TABLE OF CONTENTS

Power Analysis for Proposal Writing

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 may 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	The audience is anybody that is planning or is in 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, that are writing grant proposals, and consultants that 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	C. Nathan Marti
Department	Psychology
Title	Research Associate
Bio 	Dr. Marti served as the manager of the statistical and mathematical consulting services with the Department of Statistics and Data Sciences for 3.5 years and 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.

Structural Equation Modeling

Prerequisite Knowledge	Knowledge of correlation and multiple regression methods
Description	This course will build upon participants' previous knowledge of multiple linear regression and expand 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.
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 multivariate methods, correlation/regression, 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

Afternoon (1:30PM – 4:30 PM)

Big Data Analytics: Structured, Semi-Structured and Unstructured

Prerequisite Knowledge	Elementary knowledge of Probability, Statistics, and Calculus, but not essential and 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	Students (graduate and undergraduate), faculty, and practitioners in industry.
Computer Requirements	"Big Data Analytics: Structured, Semi-Structured and Unstructured" will be held in a computer classroom where students will have access to SAS and R.
Time	1:30 PM – 4:30 PM
Instructor	Choudur K. Lakshminarayan
Company	HP Labs
Title	Principal Research Scientist
Bio 	Choudur K. Lakshminarayan is a Principal Research Scientist in the Advanced R&D Centre at HP Software research, USA. He 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

Data Analysis Using SAS

Prerequisite Knowledge	Ability to navigate in a Windows environment and have taken an introductory statistics course that covered that 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 1 will cover opening and creating datasets, data manipulation, and t-tests. Days 2 and 3 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 4 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 and Data Sciences
Title	Specialist
Bio 	Matt Hersh is a Specialist in the Department of Statistics and 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, The University of Texas at Austin, in 2000. As part of SSC's Graduate Fellows Program, Dr. Hersh assists 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

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 may 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 hands-on experience loading, displaying, manipulating, and analyzing these data in R.
Intended Audience	People 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 analyzing 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 students 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 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. He authored the "rgdal" package exposing functions from the Geospatial Data Abstraction Library to the R language. This package is #62 of the top 100 downloaded R packages 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

Prerequisite Knowledge	Knowledge of basic probability statistics including estimation and hypothesis testing, 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 students to tools such as JAGS, WinBUGS or STAN for the estimation of a wide array of models.
Intended Audience	Those with a basic understanding of introductory statistics including estimation and hypothesis testing as well as some exposure to maximum likelihood.
Computer Requirements	None Required.
Time	1:30 PM – 4:30 PM
Instructor	Stephen Jessee
Department	Government
Title	Associate Professor
Bio	Stephen Jessee teaches courses on American politics and statistical methodology in the Department of Government at The University of Texas at Austin. He received his Ph.D. in political science from Stanford University in 2007. Interests include American politics, ideology, and voting behavior as well as Bayesian statistics and latent traits modeling.

Introduction to Data Analysis & Graphics Using R

Prerequisite Knowledge

A familiarity with introductory concepts in statistics (mean, median, mode, IQR, barplots, histograms, etc.) is preferable. In-depth knowledge, even at a basic level, is not necessary, but some prior exposure is desirable.

Description

The objective of this course is to cover basic functionality, elementary statistics, and base-package graphics in R during the first three days. This includes a basic understanding of how R operates, of data types and control structures, of graph types and graph-formatting parameters, of tabulating statistical descriptors, and of conducting statistical tests.

The fourth course day will cover more advanced topics. The scope and depth of topics covered on day four will depend on the interests expressed by the students as well as time constraints. Therefore, topics listed under the fourth day are tentative, and students are encouraged to make suggestions as to preferred subjects.

Intended Audience

Individuals with a basic knowledge of statistics and data analysis who wish to learn the fundamentals of R for analysis of data. Prospective participants may include (but certainly not limited to) individuals with a background in business, management, life sciences, social sciences, or engineering. More broadly, any individual interested in learning a new tool for analyzing and visualizing data may find the course of interest.

Computer Requirements

"Introduction to Data Analysis & Graphics Using R" will be held in a computer classroom where students will have access to R.

Time

1:30 PM – 4:30 PM

Instructor

Novin Ghaffari

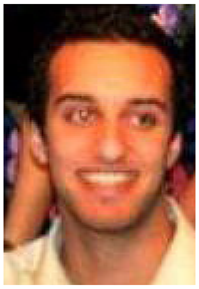
Department

Department of Statistics and Data Sciences

Title

Doctoral Student

Bio



Novin Ghaffari is a second year PhD student in the Statistics program at The University of Texas at Austin. His undergraduate studies were in finance and mathematics, and he holds an M.S. in Statistics, also from the University of Texas at Austin. His interests include probability theory, Bayesian statistics, and stochastic analysis.

RETURN TO TABLE OF CONTENTS

Introduction to Methods and Tools for Large Scale Data Driven Analysis

Prerequisite Knowledge

Participants should have taken introductory level courses in statistics and linear algebra. Participants should have knowledge of executing programs from command line interface (such as list content of directory from command line) and a working knowledge of computer programming language (such as understanding of basic loop and branching statements). Familiarity with programming in R and Java is recommended and required if the participants desire to implement their own algorithms.

Description

This course introduces students to the basic concepts and methods in data mining and software tools that can be used for practical data analysis. Three types of data mining tasks will be discussed: association analysis, cluster analysis, and classification. The purpose of this course is to teach students broad and applicable knowledge about data mining and how to use open source tools to carry out those analyses in practice. The class will focus on discussing the pro and cons among different methods and focus on how existing tools can be used rather than teaching students how to implement a particular method from scratch.
The class will introduce open source data mining tools with WEKA, R for in class demonstration as well as introduce large scale analysis tools, Hadoop and Spark for conducting data mining tasks at large scale. Examples of using tools will be showed during the class. The course will start with an overall introduction on data mining methods and open source tools in the first day. Then we will cover concepts and methods can be used in association, cluster and classification analysis in the following two sessions. In the last day, we will focus on introducing Apache Hadoop and Spark project and how it can be used for large-scale data mining.

Intended Audience

This course is intended for participants with practical needs of conducting analysis tasks but without formal training in data mining and analysis methods. We anticipate participants can have an overall understanding of common data mining and analysis methods covered by this course and be able to choose appropriate methods for their own analysis problem based on knowledge they gained in the course. The course will also introduce general open source tools that can be used with theoretical data mining method. The course will especially benefit people who are interested in large-scale data mining in practice.

Computer Requirements

Participants should bring a personal laptop. Installation of R 3.01, Java 1.6+ and WEKA 3.6 should be completed prior to the first day of the course.

Time

1:30 PM – 4:30 PM

Instructor

Weijia Xu

Department

Texas Advanced Computing Center

Title

Research Associate and Lecturer

Bio

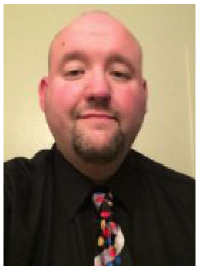
Dr. Weijia Xu leads the Data Mining & Statistics group at the Texas Advanced Computing Center. He received his Ph.D. from the Computer Science Department at The University of Texas at Austin. He has led and collaborated in a number of projects, which applies large-scale data management and mining techniques to practical research problems across various domain fields. He is also a co-instructor for the course, entitled "Visualization and Data Analysis for Science and Engineering" offered since Fall 2009 at The University of Texas at Austin.

RETURN TO TABLE OF CONTENTS

Introduction to GIS

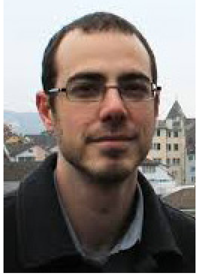
Prerequisite Knowledge	Some statistics are recommended.
Description	This course describes basic concepts underlying geographic information systems and science (GIS), and introduces participants to spatial analysis with GIS. Although the course will include hands-on laboratory exercises using ArcGIS software, 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.
Computer Requirements	"Introduction to GIS" will be held in a computer classroom with the required software available for access.
Time	1:30 PM – 4:30 PM
Instructor	Jennifer Miller
Department	Geography and the Environment
Title	Associate Professor
Bio 	Jennifer Miller is an associate professor in the Department of Geography and the Environment at The University of Texas at Austin. Her research focuses on GIScience and spatial analysis in general, and modeling biogeographical distributions and movements in particular.

Introduction to SQL and Relational Database Design

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 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 databases, how to use them, and how to input, retrieve, and analyze data using Structured Query Language (SQL).
Computer Requirements	Participants should bring a personal Windows laptop (32 or 64bit with Windows 7, 8, or 10) with MS Access – recent version OR SQL Server 2014 Express Edition – installed prior to the first day of class.
Time	1:30 PM – 4:30 PM
Instructor	Chris Golubski
Department	Department of Statistics and Data Sciences
Title	Instructor
Bio 	Chris 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 Data Science in Python

Prerequisite Knowledge	There are no hard prerequisites. However, participants are likely to get more out of the course if they have (a) 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 and 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 (instructions will be emailed to students ahead of time). 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. My research centers on the development of novel methods for the large-scale acquisition, organization, and analysis of psychological and neuroimaging data. I have over a decade of experience writing and applying Python code for data analysis, and have 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 Stata (sponsored by Stata)

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 correlation.

Description

The purpose of the course is to provide instruction in the use of Stata for data handling and for conducting statistical analyses. Day 1 will provide an overview of the software, information on basic data handling and manipulation, and exploratory descriptive analyses. Days 2 and 3 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 4 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

Lecturer

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

Non-Parametric Statistical Methods for Small Datasets

Prerequisite Knowledge	Familiarity with basic statistical concepts will be useful. For example, participants 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	None required.
Time	1:30 PM – 4:30 PM
Instructor	Bindu Viswanathan
Department	Department of Statistics and Data Sciences
Title	Lecturer
Bio 	Dr. Viswanathan is a lecturer in the Department of Statistics and 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 Masters degree in Conservation Biology from TX State University. At The University of Texas at 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

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: 1) create effective and reliable questions; 2) craft questionnaires that could be used in multiple settings (e.g., telephone, written, web-based); 3) test questionnaires to ensure their effectiveness, and 4) 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
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.

Time Series Modeling

Prerequisite Knowledge	Participants should be very comfortable with the use and interpretation of multiple regression (including calculating plug-in estimates from the regression equation and their confidence intervals, hypothesis testing on coefficients, R-square, root mean-squared error, correlation, etc.) Students should also be familiar with logarithms and exponentials, and with Excel. Some familiarity with SAS would be desirable, but I will include a short tutorial to make students quickly productive in SAS. Calculus is not required. Appropriate readings will be provided before the course.
Description	This course will teach participants 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. Participants 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. 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 variety of specialized models, but rather to provide widely applicable general-purpose tools.
Computer Requirements	Participants should bring a personal Windows OS laptop with basic Excel installed and up to date internet browser. Participants will be instructed to download SAS OnDemand prior to the first class session.
Time	1:30 PM – 4:30 PM
Instructor	Tom Sager
Department	Information, Risk, and Operations Management
Title	Professor

RETURN TO TABLE OF CONTENTS

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 PhD 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 may 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 to very applied to very theoretical. He has dabbled in statistics in insurance companies, air pollution, law, auditing, quality, and mathematics to name a few research interests. Tom has consulted extensively for insurance and re-insurance companies, lawyers, government agencies, large and small corporations, and consulting firms in the roles of quantitative analyst, expert witness, audit designer, data miner, interpreter and adviser. Last year, he testified in the Texas voter ID trial in Washington, D.C. He has taught numerous times as a visiting professor in the Masters in Finance program at Thammasat University in Thailand and recently led a group of MBA students on a trip to China. Currently Professor of Statistics in the IROM Department, Tom just loves statistics in all its ubiquity.