

**DEPARTMENT OF STATISTICS AND DATA SCIENCES**

# ***UT Summer Statistics Institute***

---

The University of Texas at Austin  
May 26–29, 2015

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.

<http://stat.utexas.edu/training/ssi>

# Table of Contents

## INTRODUCTION

---

Overview | **3**

Classes Overview | **4**

New Classes | **5**

Registration and Cost, Obtaining a UT EID,  
Methods of Payment Accepted,  
Refund and Cancellation Policy | **6**

Waitlist Policy, Software, Miscellaneous,  
Contact | **7**

## COURSE DESCRIPTIONS: MORNING

---

Advanced Regression & Predictive Models | **8**

Bayesian Biostatistics | **9**

Common Mistakes in Using Statistics:  
Spotting Them and Avoiding Them | **10**

Event History Analysis | **12**

Hierarchical Linear Modeling | **13**

Introduction to Big Data Analytics | **14**

Introduction to MATLAB | **15**

Introduction to Mixed Models with  
Applications | **16**

Introduction to Regression | **17**

Introduction to SPSS | **18**

Introduction to Statistics | **19**

Introduction to Visual Analytics | **20**

Large Scale Predictive Modeling | **21**

Making Sense of Multivariate Data: Principal  
Component Analysis, Factor Analysis and  
Cluster Analysis | **22**

Multivariate Data Analysis in R | **23**

Power Analysis for Proposal Writing | **24**

## COURSE DESCRIPTIONS: AFTERNOON

---

Big Data Analytics: Structured, Semi-  
Structured and Unstructured | **25**

Data Analysis Using SAS | **26**

Geospatial Data Analysis in R | **27**

Introduction to Bayesian Statistics | **28**

Introduction to Data Analysis & Graphics  
Using R | **29**

Introduction to Data Mining: Methods and  
Tools | **30**

Introduction to GIS | **32**

Introduction to Microsoft Access | **33**

Introduction to Statistics | **34**

Introduction to Psychoinformatics | **35**

Introduction to Stata (sponsored by  
Stata) | **36**

Questionnaire Design and Survey  
Analysis | **37**

Structural Equation Modeling | **38**

Time Series Modeling | **39**

Writing Competitive Federal Grant  
Proposals | **41**

## THE DEPARTMENT OF STATISTICS AND DATA SCIENCES

*at The University of Texas at Austin is proud to host the 8th Annual 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 2015 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 2015 SSI will offer 31 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. Participants will have access to a course website where instructors will post lectures, datasets, exercises, and other useful course information. 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 <http://stat.utexas.edu/training/ssi>. Participants are encouraged to carefully check the prerequisite knowledge specified for each course.

Category	Morning (9:00 a.m.–noon)	Afternoon (1:30–4:30 p.m.)
<b>Introductory Statistics</b>	Introduction to Statistics	Introduction to Statistics
<b>Software</b>	Introduction to MATLAB	Data Analysis Using SAS
	Introduction to SPSS	Introduction to Data Analysis, & Graphics using R
	Multivariate Data Analysis Using R	Introduction to Microsoft Access Introduction to Stata
<b>Statistical Methods &amp; Applications</b>	Advanced Regression & Predictive Models	Geospatial Data Analysis in R
	Bayesian Biostatistics	Introduction to Bayesian Statistics
	Common Mistakes in Using Statistics: Spotting Them & Avoiding Them	Introduction to GIS
	Event History Analysis	Introduction to Psychoinformatics
	Hierarchical Linear Modeling	Questionnaire Design and Survey Analysis
	Introduction to Regression	Time Series Modeling
	Introduction to Visual Analytics	Writing Competitive Federal Grant Proposals
	Making Sense of Multivariate Data: Principle Component Analysis, Factor Analysis, & Clustering Technique	
Power Analysis for Proposal Writing	Time Series Modeling	
<b>Big Data</b>	Introduction to Big Data Analytics	Big Data Analytics: Structured, Semi-Structured, and Unstructured
	Large Scale Predictive Modeling	Introduction to Data Mining: Methods and Tools

# NEW FOR 2015!

## **Bayesian Biostatistics**

This short course is an introduction to Bayesian inference methods that are commonly used in biomedical applications. The course includes a brief review of basic Bayesian inference. We then discuss inference for binary and count data, early phase clinical trial design, including Bayesian adaptive designs, and survival analysis. The emphasis will be on data analysis and practical implementation issues.

## **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.

# 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 Psychoinformatics**

This short course surveys tools and techniques for acquiring, organizing, synthesizing, and reporting psychological data. Topics include Big Data and machine learning applications in psychology; exploratory data visualization; web- and smartphone-based data collection methods; the Python ecosystem for scientific computing; and new approaches to publishing and evaluating psychological research. The course will provide researchers in psychology and cognate disciplines with the background and skills needed to apply informatics tools and techniques to research problems in their fields.

## REGISTRATION AND COST

To register, visit the following website: <http://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	
January 9, 2015 through May 1, 2015	UT-Austin	Students *	\$175 per course
		Faculty/Staff	\$275 per course
	UT-System	Faculty/Staff	\$275 per course
	Non-UT-Austin	Students **	\$225 per course
		Participants	\$550 per course
		Groups of five or more from the same institution or agency: \$440 per person per course	

\*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 [https://idmanager.its.utexas.edu/eid\\_self\\_help](https://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, 2015. 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.



## 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](mailto: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 <http://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](mailto:stat.admin@austin.utexas.edu)

Website: [stat.utexas.edu](http://stat.utexas.edu)

# Course Descriptions

## Morning (9:00AM – 12:00 NOON)

### Advanced Regression & Predictive Models

**Prerequisite Knowledge** Knowledge of introductory statistics and basic regression including sampling distributions, confidence intervals, hypothesis testing, and simple linear regression.

**Description** This course is an advanced take on the use of regression models. Strong emphasis will be placed on the implementation of the methods presented in the statistical computing package R. The course has an applied flavor where participants will be constantly faced with real data analysis examples from many areas of applications.

**Intended Audience** Participants familiar with regression at an introductory level who are looking to extend their knowledge to more advanced uses of regression and other, more modern predictive modeling tools.

**Computer Requirements** Participants should bring a personal laptop. Installation of R will be covered on the first day of the course.

**Time** 9:00 AM – 12:00 noon

**Instructor** Carlos M. Carvalho

**Department** Information, Risk, and Operations Management; Department of Statistics and Data Sciences

**Title** Associate Professor

#### Bio



Carlos M. Carvalho is an associate professor of statistics at The University of Texas at Austin with a joint appointment in the McCombs School of Business and the Department of Statistics and Data Sciences. Dr. Carvalho received his Ph.D. in Statistics from Duke University in 2006. His research focuses on Bayesian statistics in complex, high-dimensional problems with applications ranging from finance to genetics. Some of his current projects include work on large-scale factor models, graphical models, Bayesian model selection, and particle filtering and stochastic volatility models. Before moving to Texas, Dr. Carvalho was part of the faculty at The University of Chicago Booth School of Business and, in 2009 he was awarded The Donald D. Harrington Fellowship by The University of Texas at Austin.



## Bayesian Biostatistics

**Prerequisite Knowledge** Familiarity with statistical inference at the level of an introductory graduate level class in statistical inference is assumed. Some prior background or interest in Bayesian inference is useful but not strictly necessary. Knowledge of R is desired. Recommended Review Reading: Hoff, "A First Course in Bayesian Stat Methods", Springer, 2009. The book is available as an e-book from UT libraries. Personal hardcopies can be purchased at: <http://www.springer.com/gp/eproducts/springer-ebooks/mycopy>

**Description** The short course is an introduction to Bayesian inference methods that are commonly used in biomedical applications. The course includes a brief review of basic Bayesian inference. We then discuss inference for binary and count data, early phase clinical trial design, including Bayesian adaptive designs, and survival analysis. The emphasis will be on data analysis and practical implementation issues.

**Intended Audience** Anyone working on statistics for biomedical inference problems including graduate students in related fields, researchers, and practitioners.

**Computer Requirements** Participants should bring a personal laptop. Installation of R and RStudio should be installed prior to the first day of class; instructions will be provided.

**Time** 9:00 AM – 12:00 noon

**Instructor** Peter Mueller

**Department** Department of Statistics and Data Sciences, Department of Mathematics

**Title** Professor

**Bio**



Dr. Mueller is Professor in the Department of Statistics and Data Sciences and in the Department of Mathematics at The University of Texas at Austin. Before coming to Austin he served on the faculty in the Institute of Statistics and Decision Science at Duke University, and in the Department of Biostatistics at M.D. Anderson Cancer Center. He received his Ph.D. degree in statistics from the Purdue University in 1991. Dr. Mueller's current major area of interest is the theory and application of statistics to biomedical problems. In particular, he has proposed new methods for non-parametric data analysis, semi-parametric statistical methods for repeated measurement data, simulation based approaches to optimal design, innovative clinical trial designs, model based smoothing methods, and simulation based methods for posterior inference.

## 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** We often hear results of research studies that contradict earlier studies. In 2005, medical researcher John P. Ioannidis asserted that most claimed research findings are false. In 2011, psychologists Simmons, Nelson and Simonsohn brought further attention to this topic by using methods 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. Both 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. 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 (a.k.a. Publication Bias), Multiple Inference (a.k.a. Multiple Testing, Multiple Comparisons, Multiplicities, or The Curse of Multiplicity), Data Snooping, 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 participants 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 participants will also gain deeper understanding of basic statistical concepts such as p-values, confidence intervals, sampling distributions, robustness, and model assumptions.

**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**                    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.

## 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.

## 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.

**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** Keenan A. Pituch is an associate professor and former Chair of the Quantitative Methods Area in the Department of Educational Psychology at The University of Texas at Austin. His research interests include multilevel modeling, mediation analysis, and the analysis of data arising from large-scale experimental designs.



## 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.



## Introduction to MATLAB

<b>Prerequisite Knowledge</b>	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.
<b>Description</b>	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.
<b>Intended Audience</b>	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.
<b>Computer Requirements</b>	“Introduction to MATLAB” will be held in a computer classroom where participants will have access to the software.
<b>Time</b>	9:00 AM – 12:00 noon
<b>Instructor</b>	Michael Mack
<b>Department</b>	Center for Learning and Memory, Department of Psychology
<b>Title</b>	Research Affiliate
<b>Bio</b>	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.



## 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.

**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. 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.

**Computer Requirements** None

**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. 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 Stata12, but corresponding examples will be shown from other software packages such as SAS and SPSS.

**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**



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 Department of Statistics and Data Sciences since its inception and enjoys helping not only students from his class, but also faculty and staff in their research endeavors.

## 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** No prerequisite knowledge necessary, but familiarity with Microsoft Excel would be helpful.

**Description** This course will provide an introduction to descriptive and inferential statistics using Excel. After this course, participants will be able to conduct many fundamental statistical analyses and better interpret key statistical concepts.

**Intended Audience** This course is intended for a general audience; no background or previous experience in statistics is required

**Computer Requirements** Participants should bring a personal laptop equipped with Microsoft Excel 2010 and the Data Analysis Toolpak. Excel and the Data Analysis Toolpak should be installed prior to the first day of class; instructions for the Toolpak will be made available via Canvas. Excel 2010 for Macintosh has many of the same functionalities, but is presented differently. Please be advised that it might be difficult to follow along in class if you have this version of Excel.

**Time** 9:00 AM – 12:00 noon

**Instructor** René Dailey

**Department** Communication Studies

**Title** Associate Professor

**Bio** René Dailey is an associate professor in the Communication Studies Department at The University of Texas at Austin. Her research focuses on communication in dating and family relationships. She uses statistical analyses such as structural equation modeling and hierarchical linear modeling in her research.



## Introduction to Visual Analytics

**Prerequisite Knowledge** Basic mathematics skills, basic statistics skills.

**Description** In this course, attendees will learn about the visualization process, specifically targeted at information visualization, scientific visualization, and visual analytics. Attendees will be given an overview of information visualization, scientific visualization, and visual analytics concepts, constructs and applications. Attendees will also learn the basics of the Processing language as it applies to creating and developing visualizations. A significant portion of the class will involve hands-on interaction with Processing, and attendees are encouraged to bring data for use in the class (sample data will be provided as well).

**Intended Audience** Anyone needing visualization skills.

**Computer Requirements** Participants should bring a personal laptop. Instructions for installation of Processing will be covered on the first day of the course.

**Time** 9:00 AM – 12:00 noon

**Instructor** Kelly Gaither

**Department** Texas Advanced Computing Center

**Title** Director of Visualization and Senior Research Scientist

**Bio**



Kelly Gaither is the Director of Visualization and a Senior Research Scientist at the Texas Advanced Computing Center (TACC). Dr. Gaither leads the visualization activities while conducting research in scientific visualization. She received her doctoral degree in Computational Engineering from Mississippi State University in May 2000, and received her masters and bachelor degree in Computer Science from Texas A&M University in 1992 and 1988 respectively. Dr. Gaither has over thirty refereed publications in fields ranging from Computational Mechanics to Supercomputing Applications to Scientific Visualization. She has given a number of invited talks. Over the past ten years, she has actively participated in conferences related to her field, specifically acting as general chair in 2004 of IEEE Visualization.



## Large Scale Predictive Modeling

**Prerequisite Knowledge** Basic mathematical maturity; knowledge of random variables.

**Description** The course will provide an introduction to large-scale predictive analytics: covering both the modeling methodology, as well as computational algorithms to learn large-scale statistical models over a large number of variables.

**Intended Audience** The intended audience is graduate students, staff, and technology professionals.

**Computer Requirements** Participants should bring a personal laptop. Installation of R should be completed prior to the first day of the course.

**Time** 9:00 AM – 12:00 noon

**Instructor** Pradeep Ravikumar

**Department** Computer Science; Department of Statistics and Data Sciences

**Title** Assistant Professor

**Bio**



Pradeep Ravikumar received his B.Tech. in Computer Science and Engineering from the Indian Institute of Technology, Bombay, and his PhD in Machine Learning from the School of Computer Science at Carnegie Mellon University. He was then a postdoctoral scholar at the Department of Statistics at the University of California, Berkeley. He is now an Assistant Professor in the Department of Computer Science, at the University of Texas at Austin. He is also affiliated with the Department of Statistics and Data Sciences, and the Institute for Computational Engineering and Sciences at The University of Texas at Austin. His thesis has received honorable mentions in the ACM SIGKDD Dissertation award and the CMU School of Computer Science Distinguished Dissertation award. He is a Sloan Fellow, a Siebel Scholar, a recipient of the NSF CAREER Award; and was Program Chair for AISTATS 2013.

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

**Prerequisite Knowledge** Participants should have some knowledge of introductory statistics, including 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:

- a) 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.
- b) Acquiring experience in analyzing and interpreting selected multivariate datasets using the following fundamental multivariate techniques: principal component analysis, factor analysis, and cluster analysis.
- c) Learning or reviewing how to use SAS for data analysis.
- d) 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.

## Multivariate Data Analysis in R

<b>Prerequisite Knowledge</b>	Participants should be familiar with introductory statistical topics, such as descriptives and hypothesis testing, and should also be comfortable with importing data, assigning objects, and indexing data frames in R (no advanced programming skills are expected).
<b>Description</b>	This is a very hands-on course intended to introduce R users to different multivariate statistical methods by working through various datasets and answering relevant research questions. We will concentrate mainly on methods related to testing the effects of multiple independent variables on different types of outcome variables, but will also cover factor analysis and other dimension reduction techniques, as time allows. The specific methods covered in this course will include: ANOVA, MANOVA, simple and multiple linear regression, logistic regression, mixed/multi-level/hierarchical models, and factor analysis. For each of these topics, we will learn how to test assumptions, run the analyses, and interpret the results. Although some time will be spent discussing the theory behind these methods, we will mostly focus on applying them using R. Optional “homework” assignments will be assigned after each day for those who want extra practice with applying these techniques to real data.
<b>Intended Audience</b>	This course is designed for anyone who has used R before and is interested in learning how to analyze multivariate data.
<b>Computer Requirements</b>	Participants should bring a personal laptop. Installation of R and RStudio should be installed prior to the first day of class; instructions will be provided.
<b>Time</b>	9:00 AM – 12:00 noon
<b>Instructor</b>	Sally Amen
<b>Department</b>	Department of Statistics and Data Sciences
<b>Title</b>	Consultant, Instructor
<b>Bio</b>	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 the Department of Statistics and Data Sciences since July 2012. In this role, she helps graduate students, faculty, and staff with analyzing their research data and has experience using a variety of statistical software packages (R being her favorite). She is also an instructor for SDS 328M, an undergraduate biostatistics course that uses R for weekly lab assignments and projects, and she teaches software short courses in R each semester.



## 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 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.

# Course Descriptions

## Afternoon (1:30PM – 4:30PM)

### 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
Department	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.

## Data Analysis Using SAS

<b>Prerequisite Knowledge</b>	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.
<b>Description</b>	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.
<b>Intended Audience</b>	Anyone who is interested in using SAS for data analysis.
<b>Computer Requirements</b>	“Data Analysis using SAS” will be held in a computer classroom where participants will have access to SAS.
<b>Time</b>	1:30 PM – 4:30 PM
<b>Instructor</b>	Matt Hersh
<b>Department</b>	Department of Statistics and Data Sciences
<b>Title</b>	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.



## 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** A preconfigured virtual-machine environment will be provided. “Geospatial Data Analysis in R” will be held in a computer classroom where students will have access to R.

**Time** 1:30 PM – 4:30 PM

**Instructor** Tim Keitt

**Department** Department of Integrative Biology, Keittlab.org

**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.



## Introduction to Bayesian Statistics

**Prerequisite Knowledge** Knowledge of basic probability statistics including estimation and hypothesis testing, and 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

**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 beginning with regression, ANOVA, and the ggplot2 package. The scope and depth of topics covered on day four will depend on the interests expressed by the participants as well as time constraints. Therefore, topics listed under the fourth day are tentative, and participants 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** Participants should bring a personal laptop. R and RStudio should be installed prior to the first day of class; instructions will be provided.

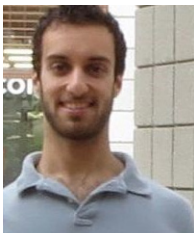
**Time** 1:30 PM – 4:30 PM

**Instructor** Novin Ghaffari

**Department** Department of Statistics and Data Sciences

**Title** Doctoral Student

**Bio** Novin Ghaffari is a current PhD student in statistics in the Department of Statistics and Data Sciences at The University of Texas at Austin. His undergraduate studies were in finance and mathematics, and he holds a M.S. in Statistics, from the Department of Statistics and Data Sciences. His interests include probability theory, Bayesian statistics, time series, and dynamic state space models.



## Introduction to Data Mining: Methods and Tools

<b>Prerequisite Knowledge</b>	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.
<b>Description</b>	This course introduces participants to the basic concepts in data mining, common methods for data preprocessing, and techniques for different data mining tasks. Three types of data mining tasks will be discussed: association analysis, cluster analysis, and classification. The purpose of this course is to teach participants broad and applicable knowledge about data mining and available tools. The broad topics covered in class will help participants grasp the basics of the field. At the same time, participants will be able to choose and apply appropriate mining methods and tools to suit their own data sets. The class will introduce open source data mining tools with WEKA and/or R for demonstration in class. For large scale data analysis, we will introduce the Map Reduce framework and how to use Hadoop to conduct data mining tasks. 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 the Apache Hadoop project and how it can be used for large scale data mining.
<b>Intended Audience</b>	This course is intended for people 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.
<b>Computer Requirements</b>	Participants should bring a personal laptop. Installation of R 3.01 and WEKA 3.6 should be completed prior to the first day of the course.
<b>Time</b>	1:30 PM – 4:30 PM
<b>Instructor</b>	Weijia Xu
<b>Department</b>	Texas Advanced Computing Center
<b>Title</b>	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.

## 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 Microsoft Access

**Prerequisite Knowledge** Participants should be comfortable working with data contained in spreadsheets (like in Excel or other programs). Those who have some knowledge of Access will benefit from the course and are encouraged to attend, but absolutely no prior knowledge is necessary. No familiarity with statistical concepts is required.

**Description** This course is intended to provide first-time users with the ability to manage, summarize, and transform datasets in Access. Participants will work with example datasets from various areas of research to learn the fundamentals of data cleaning, aggregation, and manipulation. Those with some Access experience will benefit from a detailed look into advanced query design, automating procedures with macros, designing reports, and creating macro-powered forms.

**Intended Audience** This course is designed for people who are interested in using Access to manage and manipulate survey, experimental, or business-related data.

**Computer Requirements** "Introduction to Microsoft Access" will be held in a computer classroom where participants will have access to Microsoft Access.

**Time** 1:30 PM – 4:30 PM

**Instructor** Andrea Hall

**Department** Department of Statistics and Data Sciences

**Title** Graduate Research Assistant

**Bio** Andrea holds a Master's of Science in Community and Regional Planning and a Master's of Science in Engineering. She is currently pursuing a PhD in Civil Engineering at The University of Texas at Austin. Andrea has worked in with Microsoft Access for many years, mainly in urban planning projects and to support Geographic Information Systems (GIS).



## Introduction to Statistics

**Prerequisite Knowledge** None. All analyses will be conducted using Microsoft Excel or free websites.

**Description** This course is designed for the true beginner in statistics. It requires no previous knowledge of or exposure to statistics. Simple datasets will be used and all calculations will be performed using Microsoft Excel or free websites.

**Intended Audience** No prerequisites are required. This course is designed for people who have either not had introductory statistics or have had a course before but need a refresher.

**Computer Requirements** Participants should bring a personal laptop equipped with Microsoft Excel and the Data Analysis Toolpak. This add-in comes with all versions of Microsoft Excel except for Excel 2008 for Macintosh. Excel 2008 for Macintosh has many of the same functionalities, but is presented differently. Please be advised that it may be difficult to follow along in class if you have this version of Excel.

**Time** 1:30 PM – 4:30 PM

**Instructor** Dan Robinson

**Department** Director of the School of Education at Colorado State University

**Title** Professor

**Bio**



Dan Robinson is a Professor in the School of Education at Colorado State University. Dr. Robinson serves as the editor of Educational Psychology Review and Associate Editor of the Journal of Educational Psychology. He has published over 100 articles, books, and book chapters, presented over 100 papers at research conferences, and taught over 100 college courses. His research interests include educational technology innovations that may facilitate learning, team-based approaches to learning, and examining trends in articles published in various educational journals and societies.

## Introduction to Psychoinformatics

<b>Prerequisite Knowledge</b>	No prior background in programming or statistics is assumed, though a familiarity with basic programming concepts and/or command line-based statistical analysis packages (e.g., R) is strongly recommended. Prior experience programming in Python is also helpful.
<b>Description</b>	An introduction to the development and application of informatics methods in psychology. This short course surveys tools and techniques for acquiring, organizing, synthesizing, and reporting psychological data. Topics include Big Data and machine learning applications in psychology; exploratory data visualization; web- and smartphone-based data collection methods; the Python ecosystem for scientific computing; and new approaches to publishing and evaluating psychological research. The course will provide researchers in psychology and cognate disciplines with the background and skills needed to apply informatics tools and techniques to research problems in their fields.
<b>Intended Audience</b>	This course is geared towards new and experienced researchers in psychology and cognate disciplines interested in using computational/informatics tools to acquire, analyze, and interpret psychological data.
<b>Computer Requirements</b>	A working installation of Python (version 2.7 or above) is strongly recommended. Participants should bring a personal laptop.
<b>Time</b>	1:30 PM – 4:30 PM
<b>Instructor</b>	Tal Yarkoni
<b>Department</b>	Department of Psychology
<b>Title</b>	Research Associate

### Bio



Tal Yarkoni is a Research Associate 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.

## 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.

## 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.

## 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, research consultants, and private industry researchers, and applied researchers in various disciplines.

**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** 1:30 PM – 4:30 PM

**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.



## Time Series Modeling

<b>Prerequisite Knowledge</b>	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.) Participants 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 participants quickly productive in SAS. Calculus is not required. Appropriate readings will be provided before the course.
<b>Description</b>	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.
<b>Intended Audience</b>	The course is intended to be immediately useful for anyone (The University of Texas at Austin 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 participants, 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.
<b>Computer Requirements</b>	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.
<b>Time</b>	1:30 PM – 4:30 PM
<b>Instructor</b>	Tom Sager
<b>Department</b>	Information, Risk, and Operations Management
<b>Title</b>	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 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 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.

## Writing Competitive Federal Grant Proposals

Prerequisite Knowledge None

### Description

This course will provide strategies for navigating the increasingly competitive federal funding landscape, with specific emphasis on the National Science Foundation (NSF) and National Institutes of Health (NIH). We will examine the prevailing attitudes that govern funding decisions at these federal agencies and walk section-by-section through the process of crafting a research proposal that is responsive to each agency's unique needs. A golden rule of effective writing is to know your audience; therefore we will dedicate time in each seminar to explaining how reviewers evaluate sections of proposals. Attendees will leave this course with an enhanced understanding of how to transform their research findings into written requests for funding, and will enjoy receiving meaningful grant development advice in an environment that's not reminiscent of a freshman composition class.

### Intended Audience

Anyone with an interest in learning about strategically presenting their research to a federal agency for funding. This course will be of particular interest to early career faculty, research staff who are responsible for writing grants and graduate students or postdocs who are considering careers as tenure-track faculty.

Computer Requirements None

Time 1:30 PM – 4:30 PM

Instructor Jennifer Lyon

Department College of Natural Sciences

Title Director of Strategic Research Initiatives

### Bio



Jennifer Lyon is the Director of Strategic Research Initiatives for the University of Texas at Austin's College of Natural Sciences. In this role she co-writes and develops multi-investigator, major research proposals with teams of faculty. She is also a board-certified Editor in the Life Sciences and regularly leads grant writing workshops for faculty and student audiences. Dr. Lyon serves as a review panelist for federal agencies including the National Science Foundation, Department of Energy and Department of Defense, and she draws from her experiences as a proposal writer and reviewer to educate those new to grant writing in how to most effectively convey their ideas on paper. Dr. Lyon received her Ph.D. in Analytical Chemistry from The University of Texas at Austin in 2007 and has research interests in energy storage and electrochemical methods development.