

STATS 191: Introduction to Applied Statistics

Instructor

Pratheepa Jeganathan

- E-mail: pjeganat [at] stanford [dot] edu; **Write in the subject line STATS 191.**
- Office: 390 Serra Mall, Sequoia Hall Room 116.
- Office hours: Wednesday, Thursday 3.00 p.m. - 4 p.m. in 105 Sequoia or by an email appointment.
- Course website: Canvas @ Stanford University, this will contain a bulletin board, lecture slides, homework assignments, reading list, R, R Studio, R Markdown resources etc.

Teaching Assistants & Office Hours

Course Information

- Term: Autumn 2019
- Units: 3
- Time: Mon, Wed, Fri 1:30 PM - 2:20 PM
- Location: Gates B3
- LEC: 09/23/2019 - 12/06/2019 (10 Weeks - 30 hours)

Course Overview

Statistical tools for modern data analysis. Topics include regression and prediction, elements of the analysis of variance, bootstrap, and cross-validation. Emphasis is on conceptual rather than theoretical understanding. Applications to social/biological sciences. Student assignments require use of the software package R.

Expected outcomes

By the end of the course, students should be able to:

- Enter tabular data using R.
- Plot data using R, to help in exploratory data analysis.
- Formulate regression models for the data, while understanding some of the limitations and assumptions implicit in using these models.
- Fit models using R and interpret the output.
- Test for associations in a given model.
- Use diagnostic plots and tests to assess the adequacy of a particular model.
- Find confidence intervals for the effects of different explanatory variables in the model.
- Use some basic model selection procedures, as found in R, to find a *best* model in a class of models.
- Fit simple ANOVA models in R, treating them as special cases of multiple regression models.
- Fit simple logistic and Poisson regression models.

Prerequisites

An introductory statistics course, such as - STATS 60 or STATS 110 or STATS 141, and R.

Texts

- Required:
 - **(CH)** Regression Analysis by Example.
 - * Authors: Samprit Chatterjee, Ali S. Hadi
 - * Edition: 5th Edition
 - * Print ISBN:978-0-470-90584-05

Software

- In this course, we will use R for computing and R Markdown for producing lecture slides, solutions for homework assignments. R Markdown is highly recommended to write the solutions for homework assignments. Install the following software:
 - R (required): <https://www.r-project.org/>.
 - R Studio is highly recommended for syntax highlighting, package management, document generation, and more: <https://www.rstudio.com/>.
 - * The newest version of R Studio is highly recommended (v1.1.463).
 - Latex, which will enable you to create PDFs directly from the R Markdown in RStudio.
 - * Mac users should download macTeX <http://www.tug.org/mactex/downloading.html> from Safari (not Chrome).
 - * Windows users should install MiKTeX <https://miktex.org/download>.

Grading

The final letter grade for this course will be determined by each method of assessment weighted as follows:

- 6-7 weekly homework assignments (55%)
- Midterm examination (15%, Wednesday, 10/23/2019)
- Final examination (30%, according to Stanford calendar: Wednesday, 12/11/2019 @ 3:30 PM, location TBD)

Policies

- Class Participation
 - This includes in-class participation and Canvas Discussion.
 - When homework involves simulations and data analysis, you will use R statistical computing software. Please post your R or R Markdown questions to Canvas Discussion.
 - Instructor or TA or other students in your class can answer your questions.
 - When asking questions about code, be specific (copy and paste the exact error, relevant code, and describe what you are attempting to do). We will not answer questions that are too similar to the problem sets or that would be better answered in office hours with a whiteboard. Needless to say, you should conduct yourself in a courteous and respectful manner on Canvas Discussion.
 - The class-participation credit is based on the instructor's subjective judgment of the effort involved.
- Weekly homework assignments
 - Homework assignment will be assigned every Friday on Canvas Assignments.
 - Homework assignment will be due every Friday at the beginning of class.
 - * Turn-in your completed homework assignment in PDF format and R Markdown file to TA: XXXX [at] stanford [dot] edu.
 - Use R markdown in R Studio and render it to PDF
 - * See the following link for further outline of using R markdown for reporting.
 - Each question in the homework assignment will be graded as follows: $scale \in \{0, 1, 2, 3, 4\}$
 - * 4: submitted on time and more or less completely correct

- * 3: submitted on time and partially correct answer
- * 2: late submission and more or less completely correct
- * 1: late submission and partially correct answer
- * 0: late submission (any day after the due date for more than one homework assignment) or submitted with a completely incorrect answer
- Each student can hand in only one homework late (within two days after the deadline).
- After attempting homework problems on an individual basis, students may discuss a homework assignment with their classmates. However, students must write up their own solutions individually and explicitly indicate from whom (if anyone) or resources students received help at the top of their homework solutions.
- Midterm examination
 - In-class examination.
 - Midterm examination: **Wednesday, October 25, 2019 @ 1:30PM - 2.20 PM, Gates B3.**
 - Students **are not allowed** to take midterm examinations other than scheduled date (except the event of extraordinary circumstances that solely determined by the instructor)
- Final examination
 - In-class examination.
 - Following the Stanford calendar: **Wednesday, December 11, 2019 @ 3:30PM-6:30 PM, location TBD.**
 - Students **are not allowed** to take final examinations earlier than the scheduled time (except the event of extraordinary circumstances that solely determined by the instructor).
 - Students **are not allowed** to take this course with another conflicting final examination.
- Accessible Education
 - Students with Documented Disabilities: Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty. Unless the student has a temporary disability, Accommodation letters are issued for the entire academic year. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 723-1066, URL:<https://oae.stanford.edu/>.)
 - Provide me an accommodation letter on or before **09/27/2019.**
- Honor Code
 - Students are bound by the Stanford Honor Code. Violation of the honor code will result in a failing grade among other penalties.
- Stanford Center for Professional Development (SCDP)
 - Stats 191 is listed as one of the SCDP courses.
 - Lecture recordings are being made and might be shared with others at Stanford beyond those currently enrolled in the class.
 - SGDP policies on student privacy: **Video cameras located in the back of the room will capture the instructor presentations in this course. For your convenience, you can access these recordings by logging into the course Canvas site. These recordings might be reused in other Stanford courses, viewed by other Stanford students, faculty, or staff, or used for other education and research purposes. Note that while the cameras are positioned with the intention of recording only the instructor, occasionally a part of your image or voice might be incidentally captured. If you have questions, please contact a member of the teaching team.**

Course Schedule

Date	Week	Topic	Reading	Notes
09/23/2019	Week 1	–	–	–

Date	Week	Topic	Reading	Notes
09/25/2019	Week 1	–	–	–
09/27/2019	Week 1	–	–	–
09/30/2019	Week 2	–	–	–
10/02/2019	Week 2	–	–	–
10/04/2019	Week 2	–	–	–
10/07/2019	Week 3	–	–	–
10/09/2019	Week 3	–	–	–
10/11/2019	Week 3	–	–	–
10/14/2019	Week 4	–	–	–
10/16/2019	Week 4	–	–	–
10/18/2019	Week 4	–	–	–
10/21/2019	–	–	–	–
10/23/2019	–	–	–	Midterm Examinations
10/25/2019	–	–	–	–
10/28/2019	–	–	–	–
10/30/2019	–	–	–	–
11/01/2019	–	–	–	–
11/04/2019	–	–	–	–
11/06/2019	–	–	–	–
11/08/2019	–	–	–	–
11/11/2019	–	–	–	–
11/13/2019	–	–	–	–
11/15/2019	–	–	–	–
11/18/2019	–	–	–	–
11/20/2019	–	–	–	–
11/22/2019	–	–	–	–
11/25/2019	–	–	–	Thanksgiving Recess (no classes)
11/27/2019	–	–	–	Thanksgiving Recess (no classes)
11/29/2019	–	–	–	Thanksgiving Recess (no classes)
12/02/2019	–	–	–	–
12/04/2019	–	–	–	–
12/06/2019	–	–	–	–
12/11/2019	–	–	–	End-Quarter examinations

Date	Week	Topic	Reading	Notes
04/01/2019	Week 1	Overview of current research in nonparametric and adequate initiation to R and R Markdown	ASA Nonparametric statistics section news gives an overview of some current research topics; install R; install RStudio; TryR, R Markdown webinar, R Markdown provide adequate initiation to R and R Markdown.	
04/03/2019	Week 1	Logistics and Preliminaries	Syllabus, HWC: Chapter 1	

Date	Week	Topic	Reading	Notes
04/05/2019	Week 1	The One-sample problem I (testing procedure)	HWC: Chapter 3.4-3.6, 3.8, 3.1-3.3, 3.7	Homework 1 posted
04/08/2019	Week 2	The One-sample problem II (estimator associated with the statistic, confidence interval, example)	HWC: Chapter 3.4-3.6, 3.8, 3.1-3.3, 3.7	
04/10/2019	Week 2	Statistical functionals and Influence functions	Notes will be posted on W: Chapter 2, ET: Chapter 4, 5, 21.3	
04/12/2019	Week 2	Jackknife and Bootstrap I	HWC: Chapter 8.4 and notes will be posted based on W: Chapter 3, DH, ET: Chapter 6, 11	Homework 2 Posted, Homework 1 Due
04/15/2019	Week 3	Bootstrap II	Notes will be posted based on ET: Chapter 23, Re:DH1994, Re:DH1995	
04/17/2019	Week 3	Discrete data problems I	HWC: Chapter 2	
04/19/2019	Week 3	Discrete data problems II	HWC: Chapter 10	Homework 3 Posted, Homework 2 Due
04/22/2019	Week 4	Two-sample problem I	HWC: Chapter 4	
04/24/2019	Week 4	Two-sample problem II	HWC: Chapter 5	
04/26/2019	Week 4	Permutation Test I		Homework 4 Posted, Homework 3 Due
04/29/2019	Week 5	Permutation Test II		
05/01/2019	Week 5	Ranked-based linear regression	HWC: Chapter 9	
05/03/2019	Week 5	Smoothing I	W: Chapter 4	Midterm project proposal due
05/06/2019	Week 6	Nonparametric regression I	HWC: Chapter 9.7, 14, W: Chapter 5	
05/08/2019	Week 6	Nonparametric regression II	HWC: Chapter 9.7, 14, W: Chapter 5	
05/10/2019	Week 6	Wavelets	HWC: Chapter 13, W: Chapter 9	Homework 5 Posted, Homework 4 Due
05/13/2019	Week 7	ANOVA I	HWC: Chapter 6, 7	
05/15/2019	Week 7	ANOVA II , multiple comparison	HWC: Chapter 6, 7	
05/17/2019	Week 7	Survival analysis I	HWC: Chapter 10	Homework 6 Posted, Homework 5 Due
05/20/2019	Week 8	Survival analysis II	HWC: Chapter 10	

Date	Week	Topic	Reading	Notes
05/22/2019	Week 8	Ranked set sampling	HWC: Chapter 15	Homework 7 Posted, Homework 6 Due Memorial Day
05/24/2019	Week 8	Bayesian nonparametric I	HWC: Chapter 16, Li:W2016	
05/27/2019	Week 9	(Holiday, no classes)		
05/29/2019	Week 9	Bayesian nonparametric II	HWC: Chapter 16, Li:W2016	Homework 7 Due (no late submission allowed, End-Quarter Period starts)
05/31/2019	Week 9	Inference for data visualization	Re:BHLLSW2009, Re:D1983, Re:JWH2014	
06/03/2019	Week 10	Multivariate nonparametric tests, Bootstrap III	ET: Chapter 12, 14	
06/05/2019	Week 10	Wrap-up		Final project due

Important Dates

Date	Day	Description
10/11/2019	Friday, 5:00 p.m.	Last day to add or drop a class
10/23/2019	Wednesday, 1:30-2:20 p.m.	Midterm examination
11/04/2019	Monday, 5:00 p.m.	Term withdrawal deadline with a partial refund
11/25/2019- 11/29/2019	Monday - Friday	Thanksgiving Recess (no classes)
12/02/2019 - 12/08/2019	Monday - Sunday	End-Quarter Period
12/06/2019	Friday	Last day of classes
12/11/2019	Wednesday	Final Examinations @ 3:30 p.m. - 6.30 p.m.
12/17/2019	Tuesday, 11.59 p.m.	Grades due

Disclaimer: This syllabus provides a general plan; deviations may be necessary.