STATS 191: Introduction to Applied Statistics

Instructor

Pratheepa Jeganathan

- Office: 390 Serra Mall, Sequoia Hall 116.
- E-mail: pjeganat [at] stanford [dot] edu; **Write in the subject line STATS 191**.
- Office hours: Wednesday 2:30 PM - 4:30 PM 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

Benjamin Seiler

- Zoom office hours for SCPD students: Thursday 4:30 PM - 6:30 PM.
- Zoom meeting ID: https://stanford.zoom.us/s/793447924 .
- All contacts about SCPD .

Jayoon Jang

- Office hours: Thursday 1:00 PM - 3:00 PM
- Location: Sequoia 207 (Bowker)

Samir Anwar Khan

- Office hours: Tuesday 1:00 PM - 3:00 PM
- Location: Sequoia 207 (Bowker)

Email list

The course has an email list that reaches all TAs as well as the instructor: stats191-aut1920-staff@lists.stanford.edu

As a general rule, you should send course related questions to this email list.

Questions can also be posted on Canvas Discussion.

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)

Prerequisites

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

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. 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.

Textbook

- 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.
    - Latex, which will enable you to create PDFs directly from the R Markdown in RStudio.
        * Install TinyTex package: `install.packages("tinytex")`.
        * After installing TinyTex, close RStudio.
        * Reopen RStudio.
        * Run the following: `tinytex::install_tinytex()`.

Evaluation

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

- 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)

The final percentage to letter grade conversion:

A+ = 97-105	A = 96-94	A- = 90-93
B+ = 87-89	B = 84-86	B- = 80-83
C+ = 77-79	C = 74-76	C- = 70-73
D+ = 67-69	D = 64-66	D- = 60-63

Policies

- Class Participation
  - **Bonus points 5%**. Pop quizzes will be given in class (SCPD students can complete in 24 hours). Scan your hand-written solutions as a PDF file/write down your answer in R markdown and render your answer to PDF and upload your answers to gradescope.
  - In-class participation and Piazza discussion are encouraged.
    * Find our class page at: https://piazza.com/stanford/fall2019/stats191/home
    * This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.
  - When homework involves simulations and data analysis, you will use R statistical computing software. Please post your R or R Markdown questions to Piazza.
  - 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.
- Weekly homework assignments
  - Homework assignment will be assigned every Friday on Canvas Assignments .
  - Homework assignment will be due every Friday.
    * Prepare your completed homework assignment in PDF format and submit a copy to gradescope.
    * Write the solution for each question on a new page (use \newpage).
  - 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: $scale \in \{0, 1, 2\}$
    * 2: submitted on time and more or less correct answer
    * 1: submitted on time and partially correct answer
    * 0: submitted with a completely incorrect answer or late submission (any day after the due date for more than one homework assignment).
  - Each student can hand in only one homework late (within three 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: Wednesday, October 23, 2019 @ 1:30PM - 2.20 PM, Gates B3.

- – Students **are not allowed** to take midterm examinations other than the scheduled date and time (except for the event of extraordinary circumstance that is determined solely by me.)
- 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 date and time (except for the event of extraordinary circumstance that is determined solely by me.).
  - – Students **are not allowed** to take this course with another conflicting final examination schedule.
- 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 **09/30/2019**.
- 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 (SCPD)
  - – Stats 191 is listed as one of the SCPD courses.
  - – Lecture recordings are being made and might be shared with others at Stanford beyond those currently enrolled in the class.
  - – SCPD 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.**

Topics

1. Course introduction and review.

2. Some tips on R.

3. Simple linear regression.

4. Diagnostics for simple linear regression.

5. Multiple linear regression.

6. Diagnostics for multiple linear regression.

7. Interactions and qualitative variables.

8. Analysis of variance.

9. Transformations and Weighted Least Squares.

10. Correlated errors.

11. Bootstrapping regression.

12. Selection.

13. Penalized regression.

14. Logistic regression.

15. Poisson regression.

16. Wrap-up and Final Review.

Course Schedule (Tentative)

Date	Week	Topic	Reading	Notes
09/23/2019	Week 1 (Lecture 1)	Course introduction and review	Syllabus, Lecture notes	
09/25/2019	Week 1 (Lecture 2)	Review	**CH**: Chapter 1	
09/27/2019	Week 1 (Lecture 3)	Some tips on R	Lecture notes	**Homework 1 posted**
09/30/2019	Week 2 (Lecture 4)	Simple linear regression 1 (introduction, correlation, model, estimation)	**CH**: Chapter 2.1-2.4	–
10/02/2019	Week 2 (Lecture 5)	Simple linear regression 2 (inference and prediction)	**CH**: Chapter 2.5-2.8	–
10/04/2019	Week 2 (Lecture 6)	Diagnostics for simple linear regression	**CH**: Chapter 2.9	**Homework 2 posted**, **Homework 1 Due**
10/07/2019	Week 3 (Lecture 7)	Multiple linear regression 1 (introduction, model, estimation, geometry of least squares)	**CH**: Chapter 3.1-3.5	–
10/09/2019	Week 3 (Lecture 8)	Multiple linear regression 2 (interpretation, matrix formulation, estimation, inference)	**CH**: Chapter 3.6-3.9	–
10/11/2019	Week 3 (Lecture 9)	Multiple linear regression 3 (prediction, contrasts, testing)	**CH**: Chapter 3.10-3.11	**Homework 3 posted**, **Homework 2 Due**
10/14/2019	Week 4 (Lecture 10)	Diagnostics in multiple linear regression (types of residuals, influence)	**CH**: Chapter 4	–
10/16/2019	Week 4 (Lecture 11)	Diagnostics in multiple linear regression (outlier detection, residual plots)	**CH**: Chapter 4	–

Date	Week	Topic	Reading	Notes
10/18/2019	Week 4 (Lecture 12)	Interactions and qualitative variables (interactions)	**CH**: Chapter 5	**Homework 4 posted, Homework 3 Due**
10/21/2019	Week 5 (Lecture 13)	Interactions and qualitative variables (visualization, ANOVA)	**CH**: Chapter 5	–
10/23/2019	Week 5	–	–	Midterm Examinations
10/25/2019	Week 5 (Lecture 15)	ANOVA models (one-way ANOVA, testing, contrasts)	**CH**: Chapter 5	–
10/28/2019	Week 6 (Lecture 16)	ANOVA models (two-way ANOVA, testing, contrasts, mixed effects model)	**CH**: Chapter 5	–
10/30/2019	Week 6 (Lecture 17)	Transformations and Weighted Least Squares	**CH**: Chapter 6,7	–
11/01/2019	Week 6 (Lecture 18)	Correlated errors	**CH**: Chapter 8,9	**Homework 5 posted, Homework 4 Due**
11/04/2019	Week 7 (Lecture 19)	Correlated errors	**CH**: Chapter 8,9	–
11/06/2019	Week 7 (Lecture 20)	Bootstrapping regression	Lecture notes will be provided	–
11/08/2019	Week 7 (Lecture 21)	Selection	**CH**: Chapter 11	**Homework 6 posted, Homework 5 Due**
11/11/2019	Week 8 (Lecture 22)	Selection	**CH**: Chapter 11	–
11/13/2019	Week 8 (Lecture 23)	Selection	**CH**: Chapter 11	–
11/15/2019	Week 8 (Lecture 24)	Penalized regression	**CH**: Chapter 10	**Homework 7 posted, Homework 6 Due**
11/18/2019	Week 9 (Lecture 25)	Penalized regression	**CH**: Chapter 10	–
11/20/2019	Week 9 (Lecture 26)	Penalized regression	**CH**: Chapter 10	–
11/22/2019	Week 9 (Lecture 27)	Logistic regression	**CH**: Chapter 12	**Homework 7 Due**
11/25/2019	–	–	–	Thanksgiving Recess (no classes)
11/27/2019	–	–	–	Thanksgiving Recess (no classes)

Date	Week	Topic	Reading	Notes
11/29/2019	–	–	–	Thanksgiving Recess (no classes)
12/02/2019	Week 10 (Lecture 28)	Logistic regression	**CH**: Chapter 12	–
12/04/2019	Week 10 (Lecture 29)	Poisson regression	**CH**: Chapter 13.3	–
12/06/2019	Week 10 (Lecture 30)	Final Review	Review will be posted	–
12/11/2019		–	–	End-Quarter examinations

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.