STATS 191: Introduction to Applied Statistics

Summer Quarter 2024

COURSE SYLLABUS

Instructor: Cyrus DiCiccio
Office: TBD
E-mail: cyrusd@alumni.stanford.edu
Office hours: TBD

TAs:
Rahul Kanekar
Office: TBD
E-mail: rkanekar@stanford.edu
Office hours: TBD

Puheng Li
Office: TBD
E-mail: rkanekar@stanford.edu
Office hours: TBD

Viet Vu
Office: TBD
E-mail: rkanekar@stanford.edu
Office hours: TBD

Times and Rooms:
Lectures: Monday, Wednesday, Friday 4:30 – 5:50 pm (NVIDIA Auditorium, Huang Engineering Center)

Learning Goals: 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/projects require use of the software package R.

Textbook:
   Regression Analysis by Example Using R.
   Authors: Samprit Chatterjee, Ali S. Hadi
   Edition: *6th*

Course Materials:
Class notes, the syllabus, and announcements will be available on Canvas (canvas.stanford.edu).

Grading:
Your final course grade will be based on performance on exams, homework, and class participation, weighted as follows
- Attendance/Participation: 10%
- Homework: 30%
- Midterm Exam 1: 20%
- Midterm Exam 2: 20%
- Final Exam: 20%

Your lowest homework assignment will be dropped. Your lower midterm grade will be replaced with that of the higher. Letter grades will be assigned as follows: If the leading digit of your score is a "9", you will receive some form of "A", "8" some form of "B", "7" some form of "C", and below "7" will be a failing grade. Your grade will receive a "+" if the trailing digit is "7" or above, or a "-" if it is below "3."

Midterm Schedule and Policies:
- Midterm 1 will be held on Friday, June 12th during the usual lecture time.
- Midterm 2 will be held on Friday, August 2nd during the usual lecture time.

Exams will be open book/notes. To access these materials, you may use a laptop, tablet, or phone; however, you may not use the internet.

Final Exam: Saturday, August 17th, 7-10pm. Location to be announced via Canvas.

Homeworks: Homeworks will be due Mondays at midnight. Late homeworks will not be accepted or graded. The lowest homework grade will be dropped to accommodate for unforeseen circumstances preventing you from submitting an assignment on time. Homework assignments will be made available on Canvas, and are to be submitted digitally as a single file through Canvas.

Academic Conduct: You are expected to abide by Stanford's Honor Code (https://communitystandards.stanford.edu/policies-guidance/honor-code).

Prerequisites: An introductory statistics course such as STATS 60, 110, or 141 which provides familiarity with concepts such as confidence intervals, hypothesis testing, and simple linear regression.

Computing: Throughout this class, we will perform statistical computations using R. We will be using RStudio, available for free at www.rstudio.com.

Tentative Schedule:

Week 1 (6/24-6/28): Review of Introductory Statistics, Simple Linear Regression

Week 2 (7/1-7/5): Computing with R, Multiple Linear Regression

Week 3 (7/8-7/12): Diagnostics, Interaction Variables, Qualitative Variables

Midterm 1

Week 4 (7/15-7/19): Transformations, ANOVA, Weighted Least Squares

Week 5 (7/22-7/26): Colinear Data, Autocorrelation, Correlated errors

Week 6 (7/29-8/2): The Bootstrap, Penalized Regression

Midterm 2

Week 7 (7/5-8/8): Penalized Regression, Variable Selection

Week 8 (8/12 – 8/14): Logistic Regression, Poisson Regression (models for binomial and count data)