MS&E 125: Intro to Applied Statistics

Introduction

Professor Udell

Management Science and Engineering
Stanford

April 17, 2023

MS&E 125: Intro to Applied Statistics

want to take this class?

- **ASAP:**
    - enroll (or drop)
    - sign up for Ed
    - sign up for polleverywhere
      (visit `pollev.stanford.edu` and log in with your SUNetID)

links on course website:
`https://stanford-mse-125.github.io/website-2023/`

Outline

Meet and greet

Course topics and objectives

Logistics

Course staff

- ▶ Prof. Madeleine Udell
- ▶ CA: Mike Van Ness (MS&E PhD)
- ▶ CA: Josh Grossman (MS&E PhD)
- ▶ CA: Samita Kamath (Statistics MS)

Who am I?

academic

- ▶ B.S. in Mathematics and Physics at Yale
- ▶ Ph.D. in Computational and Mathematical Engineering at Stanford
- ▶ postdoctoral fellow at the Center for the Mathematics of Information at Caltech
- ▶ professor OR at Cornell / in MS&E at Stanford

Who am I?

academic

- ▶ B.S. in Mathematics and Physics at Yale
- ▶ Ph.D. in Computational and Mathematical Engineering at Stanford
- ▶ postdoctoral fellow at the Center for the Mathematics of Information at Caltech
- ▶ professor OR at Cornell / in MS&E at Stanford

applied work

- ▶ finance: Goldman Sachs, BlackRock, Capital One, Schonfeld, Two Sigma, . . .
- ▶ tech: Google, Retina.ai, Marketing Attribution
- ▶ cybersecurity: DARPA, Expanse (formerly Qadium)
- ▶ clean energy: Aurora Solar
- ▶ politics: Obama 2012
- ▶ also healthcare, supply chain, . . .

Who are you?

poll:

- Majors: MS&E? Other majors?
- Future jobs: PMs, finance analysts, consultants, founders, data scientists, ???
- Year: 1st, sophomore, junior, senior, grad student?

Outline

What is this class about? Intro

prereqs:

- basic coding
- basic probability
- basic linear algebra
- basic calculus

prereq means: if it's your first time seeing it, it will go very fast!

What is this class about? Applied

the most useful things

- ▶ preprocessing data
- ▶ assessing data
- ▶ selecting data
- ▶ making sense of results

but also: there is nothing more useful than a good theory

- ▶ how do we know? can we be sure?

What is this class about? Statistics

- quantitative assessment for important questions
    - hypothesis testing, confidence intervals, . . .
- models
- data
- inference
- prediction
- uncertainty
- causality

Application: risk factors for heart failure

- collect data
- handle missing values
- encode features
- choose a model
- assess model performance
- assess causality
- test whether decision support tool improves outcomes

Applications

- healthcare (e.g., identify risk factors for heart failure)
- finance (e.g., forecast stock prices)
- marketing (e.g., understand customer behavior)
- manufacturing (e.g., quality control)
- social science (e.g., understand voting patterns)
- natural sciences (e.g., analyze climate data)
- sports analytics (e.g., predict game outcomes)
- . . .

Applications

- ▶ healthcare (e.g., identify risk factors for heart failure)
- ▶ finance (e.g., forecast stock prices)
- ▶ marketing (e.g., understand customer behavior)
- ▶ manufacturing (e.g., quality control)
- ▶ social science (e.g., understand voting patterns)
- ▶ natural sciences (e.g., analyze climate data)
- ▶ sports analytics (e.g., predict game outcomes)
- ▶ . . .

talk to your neighbor: what applications are you excited about?

Course topics

course website:
https://github.com/stanford-mse-125/website-2023

Course objectives (I)

- plot
- predict
- choose
- understand

Course objectives (II)

this course is about

- ▶ learning to ask the right questions
- ▶ learning to understand the answers

at the end of the course, you should know

- ▶ at least one method to solve any problem
- ▶ when **not** to trust your solution

Course objectives (II)

this course is about

- ▶ learning to ask the right questions
- ▶ learning to understand the answers

at the end of the course, you should know

- ▶ at least one method to solve any problem
- ▶ when **not** to trust your solution

the rest you can learn online. . .

Outline

Tech stack

- ▶ In person or Zoom for lectures, section, and office hours
- ▶ Course website for course materials
  (syllabus, schedule, homework, project, etc)
- ▶ poll everywhere for polls `pollev.com/madeleineudell824`
- ▶ Ed for Q&A and announcements
- ▶ Gradescope for quizzes, submitting homework, grades, solutions
- ▶ Github for warehousing code (demos, projects, and hw starter code)
- ▶ Google colab for developing code

Our programming language policy

- we'll do demos and provide homework starter code in python
- you're welcome to use any language you like (that your TAs can read) for homework or project
- TAs will only support python

Course requirements and grading

course website:
(grading, course requirements, lectures, homework, etc)
https://github.com/stanford-mse-125/website-2023

Course requirements and grading

Poll: should we allow chatGPT and friends

▶ on homework?

▶ on quizzes?

take home quizzes? how many quizzes?

Questions

during lecture:

- ask out loud

outside of lecture:

- ask at office hours
- ask on Ed
- don't send email

MS&E 125: Intro to Applied Statistics

want to take this class?

- **ASAP:**
  - enroll (or drop)
  - sign up for Ed
  - sign up for polleverywhere

links on course website:
https://github.com/stanford-mse-125/website-2023