DAT 401: Statistical Modeling and Inference for Data Science

Subject to change: All in-class announcements are official addenda to the syllabus

Title/Section: DAT 401 (32946): Statistical Modeling and Inference for Data Science
Instructor: Steffen Eikenberry
Office hours: TTh 3:30 PM-5:00 PM, WXLR 441 in person, or via Zoom by request only:
https://asu.zoom.us/j/6342106756
Email: seikenbe@asu.edu, steffen.eikenberry@asu.edu (aliases of same email)
Logistics: Time: TTh 12:00-1:15 PM, Location: Tempe WXLR A103

Course Description and Topics: This course covers the basic statistical concepts underlying data science as well as some of the major methods. Statistical inference underlines much of the data science methodology, which includes both frequentist and Bayesian approaches to inference and how they are used in data science. The course gives a review of frequentist and Bayesian interpretation of probability, as well as conditional probability, independence and Bayes formula. It covers topics such as maximum likelihood estimator, kernel density estimate, confidence intervals, bootstrap intervals and Bayesian posterior intervals. Various methods and notions from standard statistical courses are also covered: hypothesis testing, p-value, type I and II errors, A/B testing, linear, polynomial and multiple linear regression and variable selection. The course also discusses bias-variance tradeoff, one of the key phenomena in data science modeling.

Prerequisites: DAT 301 with C or better; Credit is allowed for DAT 401 or STP 494 (Statistical Modeling and Inference for Data Sci) OR Visiting University Student

Credits: 3

Textbook: Lecture notes courtesy of Dr. Marko Samara are available at:
https://math.la.asu.edu/~samara/StatMod-lectures/index.html. These will cover the majority of the topics in the class; additional lectures notes will be made available as needed.

You may also wish to consult the following (optional) supplementary texts:

1. OpenIntro Statistics, 4th edition, by D. Diez, M. Çetinkaya-Rundel & C. Barr, free at https://leanpub.com/openintro-statistics

2. A First Course in Probability, 10th edition, by Sheldon Ross; publisher: Pearson, 2020.

3. An Introduction to Statistical Learning with Applications in R, by G. James, D. Witten, T. Hastie & R. Tibshirani, free at https://www.statlearning.com/

4. Statistical Inference for Data Science, by Brian Caffo, free at https://leanpub.com/LittleInferenceBook

Learning Outcomes: By the end of the course students should be able to: At the completion of this course, students should be able to:

1. Understand fundamentals of probability
2. Understand discrete and continuous random variables
3. Understand Bayes' Theorem
4. Understand the significance of the Central Limit Theorem and the normal distribution
5. Estimate parameters using Maximum Likelihood Estimation (MLE)
6. Use MLE for frequentist and Bayesian inference
7. Construct confidence intervals for means and proportions
8. Understand basic frequentist hypothesis testing; understand the notion of p-value
9. Explain the difference between type I and type II errors
10. Compute bootstrap confidence intervals
11. Use Bayes rule to calculate posterior probabilities
12. Compute Bayesian credible intervals
13. Perform simple linear and polynomial regression
14. Understand bias-variance tradeoff
15. Perform subset selection method to fit a multiple linear regression model

Evaluation: Grade will be based upon ~7 homeworks, two midterm exams, and a final exam, with weighting as follows:

30% Homework
20% Exam 1
20% Exam 2
30% Final Exam

Grading is on a "soft curve," with final grade cutoffs to be determined at the end of semester, but the standard grading scale is guaranteed:

A ≥ 90%   B ≥ 80%   C ≥ 70%   D ≥ 60%   E <60%

Cutoffs will only be adjusted in student favor, and plus or minus grades may be assigned in borderline cases (again, only in student favor). A+s may be awarded for exceptional performance. Additionally, attendance and class participation are expected, and may be factored into your grade in borderline cases.

Homework: There will be ~7 HW assignments, which will generally be done in Jupyter notebook via the nbgrader system, available at mathds.asu.edu. You will have about one week for each homework; *no late homework will be accepted.*

In-Class Announcements: The following schedule is tentative only (including homework due and exam dates). Students are responsible for following all in-class announcements, which are considered official addenda to the syllabus.

Tentative Schedule (Subject to Change)

Week	Lecture Dates	Topics	Comments
1	1/10, 1/12	1. Intro to course; Probability definitions, axioms, set laws 2. Properties of probability; permutations, combinations, binomial coefficent	
2	1/17, 1/19	3. Conditional probability, Total probability formula, Bayes Theorem 4. Bayes Theorem, counting problems, classifiers?	HW 1 Due Sun 1/22
3	1/24, 1/26	5. Independence, Random variables, binomial distribution 6. Expected value and variance	
4	1/31, 2/2	7. Discrete & continuous random variables; Normal distribution and DeMoivre-Laplace 8. Sample histograms, KDE, population and sample means	HW2 Due Sun 2/5
5	2/7, 2/9	9. Distributions in R, quantiles, parameters vs. statistics; Central Limit Theorem 10. Central Limit Theorem	
6	2/14, 2/16	11. Exam 1 Review **12. Exam 1, 2/16**	HW3 Due Wed 2/15
7	2/21, 2/23	13. Confidence Intervals 14. Bootstrap confidence intervals	
8	2/28, 3/2	15. QQ Plots, Tests to compare distributions 16. Maximum Likelihood Estimation	HW4 Due Sun 3/5
		(3/5 - 3/12 Spring Break)	
9	3/14, 3/16	17. Frequentist and Bayesian inference using MLE 18. Continuous Bayes, Beta distribution and posterior distributions	
10	3/21, 3/23	19. Bayesian credible intervals 20. Frequentist Hypothesis Testing	
11	3/28, 3/30	21. p values, Type I, II, S, M Errors 22. A/B Testing, Power Calculation	HW5 Due Sun 4/3
12	4/4, 4/6	23. ANOVA, Hypothesis testing in R 24. Linear Regression	
13	4/11, 4/13	25. Exam 2 Review **26. Exam 2, 4/13**	HW6 Due Wed 4/12
14	4/18, 4/20	27. Linear regression, gradient descent, bias-variance tradeoff 28. Subset selection	
15	4/25, 4/27	29. Generalized linear model 30. Final Exam Review	HW7 Due Fri 4/28

Academic Dishonesty

Academic honesty is expected of all students in all examinations, papers, laboratory work, academic transactions and records. Note that copying homework solutions from online sources is considered academic dishonesty. Possible sanctions include, but are not limited to, appropriate grade penalties, course failure (indicated on the transcript as a grade of E), course failure due to academic dishonesty (indicated on the transcript as a grade of XE), loss of registration privileges, disqualification and dismissal. For more information, see http://provost.asu.edu/academicintegrity.

Students with Disabilities

Disability Accommodations: Qualified students with disabilities who will require disability accommodations in this class are encouraged to make their requests to me at the beginning of the semester either during office hours or by appointment. Note: Prior to receiving disability accommodations, verification of eligibility from the Disability Resource Center (DRC) is required. Disability information is confidential.

Establishing Eligibility for Disability Accommodations

Students who feel they will need disability accommodations in this class but have not registered with the Student Accessibility and Inclusive Learning Services (SAILS), formerly the Disability Resource Center (DRC), should contact SAILS immediately. Their office is located on the first floor of the Matthews Center Building. SAILS staff can also be reached at: 480-965-1234 (V), 480-965-9000 (TTY). For additional information, visit: www.asu.edu/studentaffairs/ed/drc. Their hours are 8:00 AM to 5:00 PM, Monday through Friday.

Policy on Threatening Behavior

All incidents and allegations of violent or threatening conduct by an ASU student (whether on-or off campus) must be reported to the ASU Police Department (ASU PD) and the Office of the Dean of Students. If either office determines that the behavior poses or has posed a serious threat to personal safety or to the welfare of the campus, the student will not be permitted to return to campus or reside in any ASU residence hall until an appropriate threat assessment has been completed and, if necessary, conditions for return are imposed. ASU PD, the Office of the Dean of Students, and other appropriate offices will coordinate the assessment in light of the relevant circumstances.

Classroom behavior: Make sure you arrive on time for class

Excessive tardiness will be subject to sanctions. Under no circumstances should you allow your cell phone to ring during class. Any disruptive behavior, which includes ringing cell phones, listening to your mp3/iPod player, text messaging, constant talking, eating food noisily, reading a newspaper will not be tolerated. The use of laptops (unless for lecture note taking), cell phones, MP3, IPOD, etc are strictly prohibited during class. Students who engage in disruptive classroom behavior may be subject to various sanctions. The procedures for initiating a disruptive behavior withdrawal can be found at https://clas.asu.edu/resources/disruptive-behavior.

Absences related to religious observances/practices: If you will be absent from class due to a religious observance or practice, it is your responsibility to inform the instructor during the first week of class. Your instructor will work with you on alternative and reasonable arrangements for any time missed.

Absences related to university sanctioned events and activities: If you will be absent from class due to participation in a university sanctioned event/activity, it is your responsibility to inform the instructor during the first week of class. Your instructor will work with you on alternative and reasonable arrangements for any time missed.

Inclusion

The School of Mathematical and Statistical Sciences encourages faculty to address and refer to students by their preferred name and gender pronoun. If your preferred name is different than what appears on the class roster, or you would like to be addressed using a specific pronoun, please let your instructor know.

Policy on Sexual Discrimination

Arizona State University is committed to providing an environment free of discrimination, harassment, or retaliation for the entire university community, including all students, faculty members, staff employees, and guests. ASU expressly prohibits discrimination, harassment, and retaliation by employees, students, contractors, or agents of the university based on any protected status: race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, and genetic information.

Title IX

Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at https://sexualviolenceprevention.asu.edu/faqs.

As a mandated reporter, I am obligated to report any information I become aware of regarding alleged acts of sexual discrimination, including sexual violence and dating violence. ASU Counseling Services, https://eoss.asu.edu/counseling, is available if you wish to discuss any concerns confidentially and privately.