Statistical Techniques in Astronomy
ASTR 234, Fall 2023

Instructors: Andy Skemer and J. Xavier Prochaska
Format: In person—we will use a whiteboard at times.
Office Hours: By appointment and scheduled co-working sessions when possible.
Textbooks: Wall and Jenkins, *Practical Statistics for Astronomers, 2nd edition* (library online)
Sivia, *Data Analysis: A Bayesian Tutorial, 2nd edition* (library online)
Ivezic et al., *Statistics, Data Mining and Machine Learning in Astronomy* (library online)
Evenson, *Data Assimilation: The Ensemble Kalman Filter.* Free online--https://link.springer.com/book/10.1007/978-3-642-03711-5
Final Exam: None

Overview: *Statistical Techniques in Astronomy* will introduce students to basic data analysis techniques with an emphasis on astronomical applications. Students will learn to (1) identify the appropriate statistical technique to apply to an astronomical problem, (2) execute that statistical technique with real astronomical data, and (3) communicate their findings in writing and oral presentations. By the end of the course, students will have developed a portfolio of analytic and computational techniques that they can apply to their own research.

Over the past several years, we have slowly shifted the class to reflect the boom of machine learning and artificial intelligence in Astronomy research. The first ~5 weeks of the course will be classical and Bayesian statistics (taught by Andy). The second ~5 weeks will be machine learning and artificial intelligence (taught by X).

Course Format: The course will be taught in a seminar style. We will meet twice a week, covering one topic (or chapter) each week. For the Monday class, students will have already done the assigned reading, and will be prepared to go over the assigned problems together. For the Wednesday meeting, students will take turns presenting on assigned topics, which might include, for example, an application of that week's work to an astronomical problem, or a comparison of different techniques.

Reading: The readings are short but very dense. We highly recommend that you read the material carefully, so that you understand almost all of it before starting the problem set. In a way, the reading is the most critical aspect of the class and you *must* stay current because the material is cumulative.

Problem Sets: Problem sets should be mostly completed in advance of the Monday seminar. We will go over the problems on Monday (you will volunteer to present problems), and your written solutions are due on Wednesday at the beginning of class (Slack to Andy and X). You may work in groups, but all work that you turn in, including the results of computer simulations, must be your own.

Presentations: We will take turns giving presentations on various topics on Wednesdays (3-4 per week). These will be on a volunteer basis, so you can sign up for topics that interest you, and you don't have to do it on weeks when you know you'll be especially busy. By the end of the quarter, everyone should sign up for approximately the same number of presentations.

Programming: The problem sets and presentations will frequently involve programming. Please program in whatever language you like.

Grading:
This class is graded S/US. In order to receive an S, you are expected to (1) attend all class sessions unless you have a valid excuse, (2) turn in all homework and (3) complete your assigned presentations. If you are not able to turn in homework or complete your presentations by the end of the quarter, you may request an incomplete. The incomplete can be resolved by turning in all of your homework and completing all of your presentations. For this class, incompletes must be resolved before the start of the Fall 2023 quarter, or else the grade will lapse to an Unsatisfactory.

Schedule:
Week 1: Probability and Monte Carlo (Wall & Jenkins Chapter 2)
Week 2: Statistics and Expectations (Wall & Jenkins Chapter 3)
Week 3: Hypothesis Testing (Wall & Jenkins Chapter 5)
Week 4: Bayes (Sivia)
Week 5: Data Modeling and Parameter Estimation (Wall & Jenkins Chapter 6 and 7)
Week 6: Correlation and Regression (Wall & Jenkins Chapter 4)
Week 7: Gaussian Processes (Rasmussen & Williams; Chapters 1 & 2)
Week 8: Principal Component Analysis & Variants (Francis & Wills (1999))
Week 9: Convolution Neural Networks and Auto-Encoders
Week 10: Natural Language Processing and Chat-GPT

(Subject to change)

Students with Disabilities
UC Santa Cruz is committed to creating an academic environment that supports its diverse student body. If you are a student with a disability who requires accommodations to achieve equal access in this course, please submit your Accommodation Authorization Letter from the Disability Resource Center (DRC) to Andy and X.

Sexual Harassment / Title IX
Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the Campus Advocacy Resources & Education (CARE) Office by calling (831) 502-

2273. In addition, Counseling & Psychological Services (CAPS) can provide confidential, counseling support, (831) 459-2628. You can also report gender discrimination directly to the University's Title IX Office, (831) 459-2462. Reports to law enforcement can be made to UCPD, (831) 459-2231 ext. 1. For emergencies call 911.

Please note that Andy and X are mandatory reporters, which means that we are required to report Title IX violations to the Title IX office, even if the person who tells one of us wishes the information to be confidential.