TEXAS TECH UNIVERSITY
Department *of* Physics & Astronomy

ASTR 3300/5300-003 Astrostatistics
Syllabus Spring 2025

SCHEDULE

2:00 - 3:20 pm TR at Sci 204

INSTRUCTOR

Dr. Nihan Pol | Sci 017 | npol@ttu.edu
Office hours: 2:00 pm - 5:00 pm Wednesday or by appointment via email.

DESCRIPTION

The purpose of this course is to equip you with the tools necessary to understand and extract meaning from real world datasets. Real datasets are usually messy; they are not uniformly sampled in time, have gaps in them, the calibration uncertainties are not always fully specified, and they might be susceptible to unknown noise processes. However, several techniques already exist to deal with such data and we will cover these in this course. As such, this course will be focused on getting practical experience with implementing these methods in Python, and we will not dwell too much on the philosophy of inference. While we will draw a lot of examples from astronomy in this course, the methods are much more general and applicable to data from other fields, such as particle and condensed matter physics, engineering, and finance.

LEARNING OUTCOMES AND TOPICS

- The role of probability in inference.
- Frequentist and Bayesian inference.
- Bayesian parameter estimation and model selection using MCMC.
- Exploratory data analysis and visualization.
- Regression analysis.
- Time-series analysis.
- Machine learning and neural networks.

COURSE PREREQUISITES

While there are no firm pre-requisites for taking this course, please understand that this is an upper-level course and the work expectations during the course will be at that level. We list a few example pre-requisites courses that will give you a sense of what knowledge you

are expected to possess, though courses with equivalent coverage will also be considered. In addition to these pre-requisites, it will be desirable, though not required, for you to have some experience with programming in Python and basic Git and GitHub knowledge. These *pre-requisites may be waived on a case-by-case basis*; please reach out to Dr. Nihan Pol if you would like to discuss the waivers.
Reference pre-requisites: ASTR 2401, PHYS 2302, PHYS 3301, and PHYS 4325 or MATH 3350, or MATH 3354.

RESOURCES

Recommended:

Textbook: "Statistics, Data Mining, & Machine Learning in Astronomy" – Z. Ivezic, A.J. Connelly, J. T. VanderPlas, & A. Gray., Updated Edition (course textbook)

- ISBN: 9780691198309.
- You can purchase this directly from the Follett bookstore or any other retailer.

Other useful texts:

- "Bayesian Logical Data Analysis for the Physical Sciences" – P.C. Gregory
- "Modern Statistical Methods for Astronomy" – E.D. Feigelson & G. J. Babu
- "Bayesian Data Analysis" – A. Gelman, J. Carlin, H. Stern, D. Bunson, A. Vehtari, D. Rubin
  - *Free: http://www.stat.columbia.edu/~gelman/book*
- "Practical Statistics for Astronomers"— J. V. Wall & C. R. Jenkins
- "Python Data Science Handbook"— J. T. VanderPlas
  - *Free: https://github.com/jakevdp/PythonDataScienceHandbook*
- "Information theory, inference, and learning algorithms"— D. MacKay
  - *Free:* https://www.inference.org.uk/mackay/itila/book.html
- "Data analysis recipes: Fitting a model to data"— D. Hogg, J. Bovy, D. Lang
  - Arxiv: https://arxiv.org/abs/1008.4686
- "Data analysis recipes: Probability calculus for inference"— D. Hogg
  - Arxiv: https://arxiv.org/abs/1205.4446

WEB TOOLS

Blackboard. The course web page is (or will soon be) on the Blackboard system. Course announcements will be posted on this site. Blackboard will also be used for electronic communications, and to post other relevant course material (such as lecture notes, grades, homework assignments etc.). Please check Blackboard and your Texas Tech email at least once a day to ensure you are not missing important course information.

ASSESSMENT AND GRADING

The final grade is based on class participation (30%), homework (40%), and a final project (30%). The details for these are given in their respective sections below.
There is no extra credit or other means of altering your score.

My grading philosophy is weighted in favor of you showing that you understand the problem(s). Be as explicit as you can in all your work, and show all steps in the calculation to receive full credit. If you know your final answer is wrong or incomplete, say so! This tells me that you understand the material.

You are strongly encouraged to confer with your classmates on homework assignments, but I expect the work you submit to be your own. I will easily be able to use git tools to check whether code has been copied. Cases of cheating will lead to an automatic zero points on the assignment.

All assignments (but not their grades) will be available to the rest of the class in the GitHub repository. My hope is that you will learn collaboratively from each other's approaches. Science is a collaborative endeavor: This statement is a fair warning that your work will be freely available to all other students in the class. Sharing knowledge and expertise in this way is intended in the spirit of scientific collaboration among your classmates. But to reiterate, I will easily be able to use git tools to check whether code has been copied

The thresholds for letter grades are 85% for A, 75% for B, 65% for C, and 60% for D.

Late submission of assignments or final project
Barring special arrangements made in advance of the due date, late submissions of portions of the final project will not be accepted for credit. Barring prior arrangements, late submissions for lecture notebooks and homework assignments will be subject to the following deductions: 1 day late –25%, 2 days late –50%, 3 days late –75%, 4+ days late will not be accepted for credit.

LECTURES and CLASS PARTICIPATION

The first part of lectures will be a lecture format, with me explaining the material. The second part of the lecture will be an individual and collaborative problem-solving session using Jupyter notebooks and Python running on your personal laptop computers (or any equivalent machine). If you do not have a personal machine, please get in touch with me as soon as possible.

Participation credit will be assigned by submitting your completed copy of the lecture Jupyter notebook with the required tasks highlighted therein. The deadline for submitting these completed lecture Jupyter notebooks will be 11:59 am CT each Monday. Credit will be given for making a reasonable attempt at completing all the tasks in the Jupyter notebook.

HOMEWORK

Homework assignments will be available to access on Fridays, and will be due by 11:59 pm CT the Friday of the following week. We will be using GitHub to submit homework assignments. A demonstration for how to do this will be given in class during the first week of classes. There will be 9 homework assignments.

To get the highest grade on the homework, all solutions must contain suitable explanation or commentary to describe the methods that were used to find the solution. Your code will be judged based on (i) how well it is commented, (ii) how well it is structured, (iii) how well it is made compact and optimized, (iv) its speed, and (v) its efficacy in delivering the correct answer.

A lack of explanation or a severe deficit in one of the criteria above will lead to a point deduction.

Final Project

You will complete a capstone project on a topic of your choosing at the end of this course. This should employ analysis techniques learned during this course and can be applied to data from your field of study. Note that this is not a semester-long project, but rather a project to be completed in the last few weeks of the semester.

You will present your work in the form of a journal-style article, i.e. including an abstract, introduction, methods, figures, references, and an appendix section. Your primary goal is to take something complicated and find a simpler way to explain it (in your own words). Don't just regurgitate information from the textbook or other resources.

The length of the article should be 4—5 pages. It should include an appendix that consists of Jupyter notebooks that can reproduce the results from the main article.

The grading rubric for the final project is as follows:
- By March 21st, please email me your topic for the final project (10%).
- By March 28th, please upload a 1-page abstract for your project to Blackboard (10%).
- During class time towards the end of the semester (dates: TBD), make a conference-style 10-minute presentation to the class about your project (20%).
- By May 9th, submit journal-style article with Jupyter notebooks to Blackboard. This will be assessed by the quality of the background/motivation and explanations, appropriateness of figures/references, and quality of writing and analysis (60%).

AI USE

Large language models like ChatGPT are readily available. However, they are not a replacement for your own thinking, reasoning, and comprehension. Use of ChatGPT, BARD, Claude, or other AI chatbots to provide complete solutions for homework or the final project is prohibited.

However, you may treat it like an expert colleague, and in fact, I recommend GitHub's Copilot for coding assistance. You may use it to help resolve misunderstandings, provide other examples or explanations, or other things that supplement the learning process. But remember, it can be wrong! If you use AI for something, then cite it appropriately.

ILLNESS-BASED ABSENCE POLICY

If at any time during this semester you feel ill, in the interest of your own health and safety as well as the health and safety of your instructors and classmates, you are encouraged not to attend face-to-face class meetings or events. When you return and once you have provided proper documentation, we will make arrangements for missed work on an individual basis. All medical documentation must be date and time-stamped. In case of an illness that will require absence from class for more than one week, the student should notify their Academic Dean.

DISTRIBUTION OF COURSE MATERIALS

All course materials students receive or to which students have online access are protected by copyright laws. Students may use course materials and make copies for their own use as needed, but unauthorised distribution and/or uploading of materials without the instructor's express permission is strictly prohibited. Students who engage in the unauthorised distribution of copyrighted materials may be held in violation of the University's Code of Conduct and/or liable under Federal and State laws.

ADA STATEMENT

Any student who, because of a disability, may require special arrangements in order to meet the course requirements should contact the instructor as soon as possible to make any necessary arrangements. Students should present appropriate verification from Student Disability Services during the instructor's office hours. Please note: instructors are not allowed to provide classroom accommodations to a student until appropriate verification from Student Disability Services has been provided. For additional information, please contact Student Disability Services in Weeks Hall or call 806-742-2405.

ACADEMIC INTEGRITY STATEMENT

Academic integrity is taking responsibility for one's own class and/or course work, being individually accountable, and demonstrating intellectual honesty and ethical behaviour. Academic integrity is a personal choice to abide by the standards of intellectual honesty and responsibility. Because education is a shared effort to achieve learning through the exchange of ideas, students, faculty, and staff have the collective responsibility to build mutual trust and respect. Ethical behaviour and independent thought are essential for the highest level of academic achievement, which then must be measured. Academic achievement includes scholarship, teaching, and learning, all of which are shared endeavours. Grades are a device used to quantify the successful accumulation of knowledge through learning. Adhering to the standards of academic integrity ensures grades are earned honestly. Academic integrity is the foundation upon which students, faculty, and staff build their educational and professional careers. [Texas Tech University ("University") Quality Enhancement Plan, Academic Integrity Task Force, 2010].

RELIGIOUS HOLY DAY STATEMENT

"Religious holy day" means a holy day observed by a religion whose places of worship are exempt from property taxation under Texas Tax Code §11.20. A student who intends to observe a religious holy day should make that intention known in writing to the instructor prior to the absence. A student who is absent from classes for the observance of a religious holy day shall be allowed to take an examination or complete an assignment scheduled for that day within a reasonable time after the absence. A student who is excused under section 2 may not be penalized for the absence; however, the instructor may respond appropriately if the student fails to complete the assignment satisfactorily.

STATEMENT OF ACCOMMODATION FOR PREGNANT STUDENTS

Any pregnant student will be provided with supportive measures as would be provided to a student with a temporary medical condition including:

1. ability to maintain a safe distance from hazardous substances, areas, or activities;
2. excused absences
3. ability to make up missed assignments or assessments
4. additional time for assignment completion; and access to instructional materials and recordings of classes for which the student has an excused absence.

Any student who is pregnant or parenting a child up to age 18 may contact Texas Tech's designated Pregnancy and Parenting Liaison to discuss support available through the University. The Liaison can be reached by emailing titleix@ttu.edu or calling 742-7233. Please go to this link for more details and important information.

CIVILITY IN THE CLASSROOM STATEMENT

Texas Tech University is a community of faculty, students, and staff that enjoys an expectation of cooperation, professionalism, and civility during the conduct of all forms of university business, including the conduct of student-student and student-faculty interactions in and out of the classroom. Further, the classroom is a setting in which an exchange of ideas and creative thinking should be encouraged and where intellectual growth and development are fostered. Students who disrupt this classroom mission by rude, sarcastic, threatening, abusive or obscene language and/or behaviour will be subject to appropriate sanctions according to university policy. Likewise, faculty members are expected to maintain the highest standards of professionalism in all interactions with all constituents of the university (www.depts.ttu.edu/ethics/matadorchallenge/ethicalprinciples.php).

SCHEDULE (subject to change without notice)

Section	Topics	Reading	Notes
Software setup and quick introduction to git and GitHub	• Installing Python, miniconda, and environment with necessary software for this course. • Quick introduction to git and GitHub. • Submitting homeworks using GitHub.	N/A	Week 1
Probability and Statistical Distributions	• Probability theory • Random variables • Probability and frequency • Central limit theorem	Ivezic, Ch. 1 & 3	Week 2—3

Section	Topics	Reading	Notes
	• Generating random draws from arbitrary distributions		
Frequentist Inference	• Point estimation • Least squares estimation • Maximum likelihood estimation • Bootstrapping and jack-knifing • Comparison of distributions	Ivezic, Ch. 4	Week 4—5
Bayesian Inference	• Priors • Parameter uncertainty quantification • Model selection • Conditional distributions • Marginalization	Ivezic, Ch. 5	Week 6—7
Data Exploration and visualization	• Non-parametric and parametric density • estimation • Dimensionality reduction • Principal Component Analysis • Visualizing data	Ivezic, Ch. 6, 7	Week 8—9
Regression, parameter estimation, and model selection	• Formulating a model • Likelihoods • Markov chain Monte Carlo • Practical parameter estimation and model selection • Cross-validation techniques	Ivezic, Ch. 8	Week 10—13
Time-series analysis	• Deterministic and stochastic processes • Auto-correlation and cross-correlation • Structure function • Random Gaussian processes • Fourier power spectrum, Lomb-Scargle periodogram • Bayesian spectral estimation	Ivezic, Ch. 10	Week 14
Deep Learning	• Neural networks • Adding hidden layers • Fully connected, recurrent, and convolutional networks • Practical deep learning • Examples	Ivezic, Ch. 9	Week 15