Physics 441/Data Sci 421

Statistical Methods for Physicists and Astronomers - Fall 2023

Adam Miller (amiller@northwestern.edu)

TA: Saarah Hall (saarah@u.northwestern.edu)

When: Mon & Wed 1:30 — 2:50 PM

Where: Tech L361

Office Hours: Mon 3:00—3:40 PM, Tech F123
Thurs 9:30—10:30 AM, room 8069, 1800 Sherman Ave
also by appointment

For Saarah Hall:
Thurs 3:30—4:30 PM, Dearborn Room 14

[Office hours are a time to meet with Professor Miller and/or TA Hall to ask questions about the course content including lectures, problem sets, and the final project. Other discussion topics are also welcome]

Textbook: Statistical Data Analysis, Glen Cowan; ISBN- 10: 0198501560, ISBN-13: 9780198501565

[All the material required to complete the course work will be covered during lectures, the textbook is not strictly required]

Teaching Method: Lectures, hands-on activities, discussion, problem sets, and reading assignments.

Final: Project, **due December 6, 11:59 pm CST**.

Course website: https://canvas.northwestern.edu/courses/198692

Course feedback: can be provided at any point in time via an anonymous google form – https://forms.gle/SRABYTCLDkhtWadB8

Software to download/install (instructions are available via CANVAS):
python and jupyter (+required packages in python)

Note – assignments in this class will be written in jupyter notebooks (i.e., python). Assignments may be submitted in other formats, and programming may be done in other languages (though some of the problems at the end of the quarter will be easier to solve in python). It is expected that students are familiar with concepts such as: reading and writing data from/to comma separated value text files, producing scatter and line plots of the data, generating random numbers (using existing functions within well-documented software packages, e.g., numpy), writing short analysis scripts, for loops, if… then… and else… commands, reading documentation to apply (simple) functions from well maintained modules (e.g., scipy). Students with limited programming experience will be expected to learn the relevant basics outside of class. The CIERA REACHpy module provides a decent introduction: https://github.com/CIERA-Northwestern/REACHpy

Learning Objectives: At the end of this course, students should be able to:

1. Understand what a *probability density function* is, and what role it plays.
2. Have an understanding of *uncertainties*, and how uncertainty estimation influences inference.
3. Understand what a *test statistic* is and what properties it should have.
4. Use the principles of *maximum likelihood*.
5. Know how to make a *point estimate* and interpret a *confidence interval.*
6. Know how to perform linear and non-linear *regression* (i.e., fitting models to data).
7. Know how to generate *synthetic data sets* including by means of MCMC.
8. Understand the subtleties of *hypothesis tests*.
9. Appreciate the distinctions between *frequentist* and *Bayesian statistics*.
10. Apply *Bayes' Rule* and understand the principles of *Bayesian Inference.*
11. Understand the basics of *nonparametric methods.*
12. Comfortably use *Python* to analyze data and produce basic scientific computing results.
13. Produce a written document, with supporting figures and tables, summarizing the numerical analysis of a data set.

Learning materials: example data sets from physics and astronomy (and other fields) will be used to demonstrate the statistical principles covered in the class

POLICIES:

Problem sets will be assigned. They are due one week after they are assigned, unless explicitly stated otherwise.

Late problem sets will be accepted, up to one week after the initial due date. Assignments turned in after this date will have a final score that is reduced by 10% of the total number of points for every week they are late.

Late submissions of the final project proposal will not be allowed as the proposal must be reviewed with sufficient time to provide feedback on the final project.
Late submissions of the final project will be accepted through Dec 8, 2023. Assignments turned in after the due date (11:59 pm, Dec 6, 2023 - please note 12:02 am on Dec 7, 2023 will be considered late) will have a final score that is reduced by 10% of the total number of points for every day they are late.

Please note that the late assignment policy is to allow for unexpected circumstances, this should not be treated as the "true" due date since each problem set builds upon previous assignments, and significant time is needed to complete the final project. For this reason it is advisable to stay on top of the homework due dates.

Attendance policy: None, but strongly encouraged.

Grading policy: The course will be graded on the problem sets (65%), and the final project (35%). Assigned readings are optional and will not be graded.

Final project: for the final project, students will analyze a data set relevant in their field. Students will also have to submit a proposal for the final project 3 weeks before it is due. The proposal will detail the data set and methods to be used to analyze the data as part of the final project. The analysis should include inference (i.e., there should be a model fit to data or something similar) producing a final result. Students may select whatever project interests them the most. The final submission will be a jupyter notebook detailing the end-to-end data analysis. The notebook must also include text detailing the data set and analysis that was conducted.

======
Class Overview: The course is designed for graduate students in physics, astronomy, Earth science, and engineering. The goal is for students to learn the basic methods and techniques for the statistical analysis of experimental data. Emphasis will be placed on the wide variety of available methods. Through the problem sets, students will gain wide experience with statistical methods. Note –

this is not a course on mathematics or machine learning; the focus will be on understanding and modeling data.
======

IMPORTANT DATES:
Wednesday, Sep 20 – first day of class
Monday, Sep 25 – last day to add/change classes
Friday, Oct 27 – last day to drop a class
Friday, Nov 10 – last day to change to P/N
Wednesday, Nov 29 – last day of class
Wednesday, Dec 6 – final assignment due

ACADEMIC INTEGRITY:
Students in this course are required to comply with the policies found in the booklet, "Academic Integrity at Northwestern University: A Basic Guide". All papers submitted for credit in this course must be submitted electronically unless otherwise instructed by the professor. Your written work may be tested for plagiarized content. For details regarding academic integrity at Northwestern or to download the guide, visit: https://www.northwestern.edu/provost/policies-procedures/academic-integrity/index.html

ACCESSIBILITY STATEMENT:
Northwestern University is committed to providing the most accessible learning environment as possible for students with disabilities. Should you anticipate or experience disability-related barriers in the academic setting, please contact AccessibleNU to move forward with the university's established accommodation process (e: accessiblenu@northwestern.edu; p: 847-467-5530). If you already have established accommodations with AccessibleNU, please let me know as soon as possible, preferably within the first two weeks of the term, so we can work together to implement your disability accommodations. Disability information, including academic accommodations, is confidential under the Family Educational Rights and Privacy Act.

RELIGIOUS OBSERVANCE STATEMENT:
Northwestern is committed to fostering an academic community respectful and welcoming of persons from all backgrounds. To that end, the policy on academic accommodations for religious holidays stipulates that students will not be penalized for class absences to observe religious holidays. If you will observe a religious holiday during a class meeting, scheduled exam, or assignment deadline, please let

me know as soon as possible, preferably within the first two week of class. If exams or assignment deadlines on the syllabus fall on religious holidays you observe, please reach out so that we can discuss that coursework.

STATEMENT OF INCLUSIVITY:
This course strives to be an inclusive learning community, respecting those of differing backgrounds and beliefs. As a community, we aim to be respectful to all students in this class, regardless of race, ethnicity, socio-economic status, religion, gender identity or sexual orientation.

UNAUTHORIZED RECORDINGS:
Unauthorized student recording of classroom or other academic activities (including advising sessions or office hours) is prohibited. Unauthorized recording is unethical and may also be a violation of University policy and state law. Students requesting the use of assistive technology as an accommodation should contact AccessibleNU. Unauthorized use of classroom recordings – including distributing or posting them – is also prohibited. Under the University's Copyright Policy, faculty own the copyright to instructional materials – including those resources created specifically for the purposes of instruction, such as syllabi, lectures and lecture notes, and presentations. Students cannot copy, reproduce, display, or distribute these materials. Students who engage in unauthorized recording, unauthorized use of a recording, or unauthorized distribution of instructional materials will be referred to the appropriate University office for follow-up.

SUPPORT FOR WELLNESS AND MENTAL HEALTH:
Northwestern University is committed to supporting the wellness of our students. Student Affairs has multiple resources to support student wellness and mental health. If you are feeling distressed or overwhelmed, please reach out for help. Students can access confidential resources through the Counseling and Psychological Services (CAPS), Religious and Spiritual Life (RSL) and the Center for Awareness, Response and Education (CARE). Additional information on all of the resources mentioned above can be found here:

https://www.northwestern.edu/counseling/

https://www.northwestern.edu/religious-life/

https://www.northwestern.edu/care/

https://www.northwestern.edu/studentaffairs/timelycare.html

USE OF GENERATIVE AI:
Closed: You are prohibited from using Generative Artificial Intelligence (GAI) to produce any materials or content related to this course. Any use of GAI will be viewed as a potential academic integrity violation.

CLASS SCHEDULE:

Lecture 1: *Wednesday* Sep 20 (Lecture) Introduction to uncertainty. Broad overview of the types of problems that will be encountered in this course

Sep 21 - No office hours Adam Miller

Lecture 2: Sep 25 (Lecture) Probability
Homework (due **Oct 2**): probability
No office hours for Professor Miller

Lecture 3: Sep 27 (remote Lecture) Random Variables and Probability Distribution Functions

Lecture 4: Oct 2 (Lecture) Expectation Values and Propagation of Uncertainty
Homework (due **Oct 9**): probability distribution functions

Lecture 5: Oct 4 (Lecture) Common Probability Distributions (binomial, multinomial, Poisson, Gaussian, exponential, etc)

Lecture 6: Oct 9 (Lecture) Parameter Estimation (Hypothesis Tests; if there is time)
Homework (due **Oct 16**):

Reading Day: Oct 11 (reading assignment)
Oct 12 - No office hours Adam Miller AND Saarah Hall

Lecture 7: Oct 16 (Lecture) Hypothesis Tests and Maximum Likelihood
Homework (due **Oct 23**):

Lecture 8: Oct 18 (Lecture) Maximum Likelihood II

Lecture 9: Oct 23 (Lecture) The Method of Least Squares
Homework (due **Oct 30**):

Lecture 10: Oct 25 (Lecture) Confidence Intervals

Oct 27 - NU drop deadline

Lecture 11: Oct 30 (Lecture + hands-on activity) Random Numbers, Monte Carlo method, bootstrapping, calculating pi
Homework (due **Nov 06**): curve fitting

Lecture 12: Nov 1 (Lecture) Outliers and non-linear models

Lecture 13: Nov 6 (Lecture) Introduction to Bayesian statistics
Homework (due **Nov 13**): **Final project proposal** (late submissions not allowed)

Lecture 14: Nov 8 (Lecture) The Markov Chain Monte Carlo method

Lecture 15: Nov 13 (hands-on activity) Metropolis-Hastings Algorithm
Homework (due **Nov 20**): Bayesian statistics

Lecture 16: Nov 15 (Lecture) Bayesian marginalization, dealing with uncertainties that aren't "simple"

Lecture 17: Nov 20 (Lecture) Gaussian Process Regression
Homework (due **Nov 27**): Bayesian model fitting

Lecture 18: Nov 22 (Lecture + hands-on activity) GPs II

Lecture 19: Nov 27 (Lecture + hands-on activity) Model Selection + Bayesian Evidence

Lecture 20: Nov 29 (Lecture + hands-on activity) Summary of Bayes and Frequentist

Final project due Dec 6, 11:59 pm CDT.