

# ASTRON 3705/PHYS 3704

## Statistics and Data Science

### Course Information

Meeting Time: Monday and Wednesday, 12:15 PM – 1:30 PM, 303 Old Engineering Hall (unless otherwise specified)

Credits: 3

Prerequisites: None

### Instructor Information

Lecturer: Prof. Jeffrey Newman

Office: 310 Allen Hall

Office Hours: TBD

Email: [jnewman@pitt.edu](mailto:jnewman@pitt.edu)

Phone: (412) 592-3853

### Textbook

*Practical Statistics for Astronomers*, by Wall & Jenkins, **Second Edition**

Despite the title, the portions of the textbook that we will be using are not dependent upon prior astronomical knowledge. It is available through the bookstore or online through the University Library System at [https://pitt.primo.exlibrisgroup.com/permalink/01PITT\\_INST/g3767I/alma9979140963406236](https://pitt.primo.exlibrisgroup.com/permalink/01PITT_INST/g3767I/alma9979140963406236)

From the PittCat record, you can click on either Ebook Central link in the white box, under “View Online”.

### Course Description

Current research in physics and astrophysics relies heavily on advanced techniques for data analysis and the ability to deal with large datasets. The principal aim of the course will be to provide practical experience with statistical methods, including but not limited to applications to astronomical and physics data. It will cover topics such as data analysis techniques, interfacing with data files, statistics and machine learning methods, and experiment design. Some prior knowledge of

astronomy will be helpful for some additional material that may be covered in optional lectures but will not be necessary for the class. However, students **should** have some prior experience with computer programming (language does not matter). This course is designed to be useful for both physics and astrophysics students with either theoretical or observational/experimental interests.

## Course Objectives

The principal goal of this course is for students to gain the skills needed to perform research in astrophysics or physics that makes use of measurements from data (in any fashion, including interpretation). For example, astrophysical theorists make frequent use of statistical “survey” datasets, and particle phenomenologists will compare predictions to the results of experiments; this course is intended to be helpful for *anyone* pursuing research. The emphasis will be on developing practical skills in applied statistics and machine learning, data processing/interfaces, and programming, primarily through hands-on experience with all of these.

By the end of this course, you should be able to:

- Apply a variety of statistical tests to measurements, and identify the correct test for the problem being faced
- Apply Monte Carlo and resampling techniques to predict distributions of errors, estimate significances, etc.
- Perform linear and nonlinear curve fitting
- Apply maximum likelihood techniques and utilize robust statistics to make measurements
- Read in data files from existing datasets, select objects/items of interest, and apply statistical methods to those data
- Apply a few machine learning techniques for both classification and regression (continuous variable prediction)
- Perform data analysis, I/O, and plotting in the Python programming language

If time allows, we will also cover basics of Fisher matrix, Markov Chain Monte Carlo, and/or broader sets of machine learning techniques. We may supplement these activities with a couple of astronomy-specific lectures, which will be optional for students from other research areas.

The capstone of this course is for students to undertake original research projects applying statistical or machine learning methods to a problem (which may be related to/build on but should not be identical to their current research activities). In general, I will try to adjust course content to reflect student interests.

## Grading

Grading will be based on:

30%	Homework
20%	In-class activities
40%	Final project
10%	Astro Coffee Participation OR paper summary

### Homework:

Every week or two there will be 1-2 small homework problems/tasks.

Students are allowed (and encouraged!) to collaborate on homework assignments in developing basic algorithms, but must present their own work (programs/plots, results, etc.).

Please provide all relevant code or notes describing how you did a calculation, to help me understand the cause of any errors. I do not expect a detailed writeup of Python-based problems (e.g., it is not necessary to spend time on nicely formatted equations) unless I specify otherwise. If I say I want a plot, I want a plot; if I want a particular number, I care more about the process of getting that number in Python than a detailed derivation of the underlying methods. Source code or notes allows me to assess that process and help me to provide more useful feedback.

Conceptual problems are a different case, and I do expect you to explain your reasoning fully there.

I will accept at most 2 late homeworks from any letter-grade-option student, except in extreme (e.g., medical) circumstances or via prior arrangement more than two days before the deadline. The first late homework will have no penalty, the second one will have a grade penalized by 10%. Late homeworks must be turned in within 24 hours of the nominal due date unless otherwise arranged in advance.

### In-class activities:

We will periodically split up into small groups for students to work to solve problems. Class time will be a mix of lecturing with activities (e.g., solution of some problem or a programming activity). This will often yield some work product which will contribute to your in-class activities grade (as completion/noncompletion, not graded in detail). If you do need more time, submission after class will be acceptable. **You will be asked to send me all of your Jupyter notebooks at the end of the semester for this part of your grade.**

**Astro Coffee:** I expect Astrophysics graduate students to regularly attend astro coffee (the journal club held twice a week). This does not mean that you have to be present for every single meeting, but you should be there frequently. I expect that over the course of the semester each astrophysics student will present at least two papers. I will set up a google form that you can fill out quickly each time you present so that I have a record. Grading will be based on completion: zero papers = F, 1 paper = C, 2 papers = A .

As an alternative to this, non-astronomy students can produce a 2-3 page summary of a paper in your field that applies some statistical or machine learning method, written for nonexperts (see examples of the desired level or detail at [astrobites.org](http://astrobites.org) ).

### **Final Project:**

The capstone of the course will be an original project – development of a new data reduction algorithm, a new analysis of a dataset, etc. I will provide a list of some project ideas early in the semester which students may sign up for, or am happy to discuss your own project ideas (e.g., ways to incorporate new statistical methods into your current research). Projects should **not** simply consist of the research you are already doing, however. Many projects should be suitable for working in groups of 2 or 3.

At the end of the semester, students will give a 10-20 minute presentation of their project to the class. They will also provide a brief writeup (this should be provided in a **manuscript format with a minimum font size of 12 points**; the length should ideally be 2-3 pages, with 5 pages for the text plus figures **maximum**) to allow both oral and writing skills to be evaluated; presentation of results is a key skill for all scientists.

## **Academic Integrity**

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity. Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

To learn more about Academic Integrity, visit the Academic Integrity Guide for an overview of the topic. For hands-on practice, complete the Understanding and Avoiding Plagiarism tutorial.

## Disability Services

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 140 William Pitt Union, (412) 648-7890, [drsrecep@pitt.edu](mailto:drsrecep@pitt.edu), (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

## Diversity and Inclusion

The University of Pittsburgh does not tolerate any form of discrimination, harassment, or retaliation based on disability, race, color, religion, national origin, ancestry, genetic information, marital status, familial status, sex, age, sexual orientation, veteran status or gender identity or other factors as stated in the University's Title IX policy. The University is committed to taking prompt action to end a hostile environment that interferes with the University's mission. For more information about policies, procedures, and practices, see: <https://www.diversity.pitt.edu/civil-rights-title-ix-compliance/policies-procedures-and-practices> .

I ask that everyone in the class strive to help ensure that other members of this class can learn in a supportive and respectful environment, and to follow the [departmental code of conduct](#). If there are instances of the aforementioned issues, please contact the Title IX Coordinator, by calling 412-648-7860, or e-mailing [titleixcoordinator@pitt.edu](mailto:titleixcoordinator@pitt.edu). Reports can also be filed online: <https://www.diversity.pitt.edu/make-report/report-form>. You may also choose to report this to a faculty/staff member; keep in mind that as mandatory reporters they are required to communicate this to the University's Office of Diversity and Inclusion. If you wish to maintain complete confidentiality, you may also contact the University Counseling Center (412-648-7930).

## Take Care of Yourself

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking time to relax. Despite what you might hear, using your time to take care of yourself will actually help you achieve your academic goals more than spending too much time studying. All of us benefit from support and guidance during times of struggle. There are many helpful resources available at Pitt. An important part of the college experience is learning

how to ask for help. Take the time to learn about all that's available and take advantage of it. Ask for support sooner rather than later – this always helps. If you or anyone you know experiences any academic stress, difficult life events, or difficult feelings like anxiety or depression, we strongly encourage you to seek support. Consider reaching out to a friend, faculty or family member you trust for assistance connecting to the support that can help.

The University Counseling Center is here for you: call 412-648-7930 and visit their website.

If you or someone you know is feeling suicidal, call someone immediately, day or night: University Counseling Center (UCC): 412 648-7930  
University Counseling Center Mental Health Crisis Response: 412-648-7930 x1  
Resolve Crisis Network: 888-796-8226 (888-7-YOU-CAN)

If the situation is life threatening, call the Police: On-campus: Pitt Police: 412-268-2121 Off-campus: 911

## **Copyright Notice**

All course materials should be protected by copyright. United States copyright law, 17 USC section 101, et seq., in addition to University policy and procedures, prohibit unauthorized duplication or retransmission of course materials.

See [Library of Congress Copyright Office](#) and the [University Copyright Policy](#).

**Note:** The schedule and procedures in this course are subject to change. Any changes will be posted on the ASTRON 3705/PHYS 3704 Canvas site and announced in class in the case of major changes.