# AstroStats

## ASTR/PHYS 390/590
## Spring, 2023

## 1 Course Information

- **Day and Time**: Thursday from 4:30 - 7:10 pm
- **Location**: 220 Planetary Hall
- **Instructor**: Bob Weigel
- **Email**: rweigel@gmu.edu
- **Office Hour**: Thursday 3:30-4:30 pm
- **Course URL**: http://rweigel.github.io/astrostats
- **Credits**: 3 (Lecture)
- **Prerequisites**:
  - 390: Grade of B or higher in PHYS 251 or permission of instructor;
  - 590: Enrollment in a COS or VSE graduate program.
- **Catalog Description**: Bayesian and frequentist statistical and data analysis methods applied to data and problems in astrophysics and the space sciences.

HWs: HW1

## 2 Motivation

- Many entry-level jobs for BS, MS, and PhD typically involve exploratory data analysis and require a basic understanding of statistics.
- Many of our graduate students are working on data-intensive research problems that require an understanding of modern statistical methods.
- Many BS, MS, and PhD students have not had a traditional statistics course as an undergraduate.

## 3 Methods of Instruction

- In-class time will be approximately 50% lecture/demonstration and 50% in-class discussion.
- I usually have you do a reading and basic follow-up problems before I discuss a topic in class. I find content that I present "sticks" if students have experiemented with the concepts before I discuss them.
- You will be asked to participate in class discussions during class and on Discord. Talking and writing about statistics will help you understand statistics.
- Homework assignments will be 50% hand-written and 50% programs. I will ask students to present their solutions during class and for other students to comment.
- I give detailed solutions to homework problems. However, I will not be able to provide detailed

feedback or identify many errors in your code. I will discuss general types of issues that I observed on the homework problems during class.

- As a result, I may ask you to meet with me to explain your reasoning. I will comment on your how you solved the problem and how you presented your results. Like ordinary writing, they can always be improved. I want you to leave this class knowing how a working scientist organizes analyses and presents results.

## 4 Programming

- Any programming language may be used for homework and the final project. I will write my solutions in Python.
- I assume that you are proficient in Python (or another language) at the level of a B grade in PHYS 251. The topics that you should be familiar with are covered in my notes for PHYS 251.
- I have provided an extensive set of notes on Python techniques that you will use and should be familiar with.

## 5 Textbook

Practical Statistics for Astronomers, 2nd Edition (2012), by J.V. Wall and C.R. Jenkins | Amazon | Author's web page |

This textbook covers modern statistical methods used in astronomy. It is not a comprehensive textbook, however. The textbook is most useful as a reference to understand the approaches and types of problems considered in the astronomical literature. For this reason, I will supplement this textbook with copies of background material. I'll typically provide two references that explain the background material.

Two general references on statistics that I highly recommend are

1. Principles of Statistics, M.G. Bulmer – A short Dover book that covers fundamental statistical topics in and parsimonious manner. Available used on Amazon for  $5.00.
2. Probability and Statistics for Engineering and the Sciences (8th ed), J.L. Devore – A commonly used textbook for upper-division engineering and science majors. | Amazon |

In the References section of this syllabus, many additional references and resources are listed.

## 6 Evaluation

- **Homework**: 40% - Approximately one per week; extra problems will be assigned for students registered in 590. Most homework assignments will include at least one problem that requires the use of astronomy or space science data. Assignments are due before class starts.
- **Midterm**: 30%
- **Final project**: 30%
- **Final course**: Grades for undergraduates are determined from numerical course grades using 90%–100% A, 80%–90% B, 70%–80% C, 60%–70% D, < 60% F. For graduate students, 70% and lower is an F (the graduate grade scale does not have a D).

## 7 Topics

The topics listed below are covered in Chapters 2-7 of the textbook. Supplementary references and notes will be provided as-needed.

## 7.1 Python

1. NumPy
2. Random number generators
3. Data Visualization

## 7.2 Basic Concepts in Probability

1. Mathematical, Frequentist, and Bayesian probabilities
2. Conditional probability
3. Bayes' theorem
4. Counting - permutations and combinations
5. Probability distributions
6. Overview of statistical fallacies

## 7.3 Elements of Statistics

1. Expectation, mean, variance, and bias
2. Random variables, distributions, quantiles, mean, variance
3. Joint distributions, covariance, correlation, independence
4. The law of large numbers
5. Central limit theorem
6. Bootstrapping

## 7.4 Hypothesis Tests

1. Frequentist
2. Bayesian
3. Bootstrap and Jackknife

## 7.5 Parameter Estimation - Basic

1. Least-squares
2. Maximum likelihood

## 7.6 Parameter Estimation - Advanced

1. Linear models
2. Nonlinear models

The EMCEE Markov chain Monte Carlo Python library

# 8 References

The following references are not needed for this course, but you may find them useful for alternative explanations of topics in the textbook.

## 8.1 General

1. Principles of Statistics, M.G. Bulmer - A short Dover book that covers fundamental statistical topics in and parsimonious manner. Available on Amazon for $3.00.
2. Many of the topics covered in this course are covered in the MIT OCW course Introduction to Probability and Statistics.
3. Dealing with Uncertainty - A Guide to Error Analysis 2nd Edition, M. Drosg - Covers basics of uncertainty and error analysis at Freshman physics lab level. Very good but very expensive (> $100 on Amazon).
4. Statistical Methods in Experimental Physics, F. James - A standard reference that you should have a copy of.
5. Statistics, D. Freedman, R. Pisani, and R. Purves - A basic introduction to statistics with many examples and extended discussion of topics.
6. Probability for the Enthusiastic Beginner, D. Morin - A very basic introduction to probability.
7. Probability and Statistics for Engineering and the Sciences (8th ed), J.L. Devore - A commonly used textbook for upper-division engineering and science majors.
8. Probability and Statistics, M.H. DeGroot and M.J. Schervish (4th ed) – Similar in scope and coverage of Devore.

## 8.2 Bayes

### 8.2.1 General

3. Bayes' Rule - A Tutorial Introduction to Bayesian Analysis (2013), by J.V. Stone. This book is a good starting point.
4. Doing Bayesian Data Analysis by Kruschke - The best overview that I am aware of. The examples are in R, but translation to other languages is straightforward. Most of the value of this book is in the descriptions of the concepts.
5. Bayesian Statistics – an Introduction by P.M. Lee is a classic introductory textbook.
6. How to become a Bayesian in eight easy steps: An annotated reading list, by Etz et al., 2018.
7. Data Analysis - A Bayesian Tutorial (2006), by D.S. Sivia and J. Skilling.
8. Neural Networks for Pattern Recognition (2006), by C.M. Bishop - Advanced undergraduates or first-year PhD students; the first few chapters have a good introduction to Bayes' rule.
9. Teaching Statistics in the Physics Curriculum: Unifying and Clarifying Role of Subjective Probability (1999), by G. D'Agostini
10. The Elements of Statistical Learning (2009; 2nd Edition), by T. Hastie, R. Tibshirani, and J. Friedman – This is a classic book on Machine Learning/Statistical Learning.

11. Validation of Software for Bayesian Models Using Posterior Quantiles by Cook, Gelman, and Rubin, 2006 (https://www.jstor.org/stable/27594203).
12. Section 3.4 of the lecture notes on Hierarchal Models by Junker contains in Section 3.4 a detailed introduction to a common example covered in many books – finding the posterior when the standard deviation of the population is known and a prior distribution is available.

### 8.2.2 Bayesian vs. Frequentist

1. The primary reference is Jaynes 1976.
2. Frequentism and Bayesianism: A Python-driven Primer (2015), by J. VanderPlas (arxiv.org/pdf/1411.5018.pdf).
3. Lecture notes on the comparison of frequentist and Bayesian Inference, by J. Orloff and J. Bloom (MIT18_05S14_Reading20.pdf)
4. Statistical Inference Showdown: The Frequentists vs. The Bayesians (2017), by K. Dubovikov.
5. The Earth is Round covers the mis-use of significance tests in the psychological literature.

### 8.2.3 Credible Intervals

In the following references, the terms "Bayesian Intervals", "Bayesian Confidence Intervals", and "Credible Intervals" are used to mean the same thing.

- The primary reference is Jaynes 1976. This is not an easy read as a first introduction. The following references provide a more introductory explanation: Kruschke and Liddell 2018; VanderPlas 2014. A blog post by VanderPlas goes over one of the examples in Jaynes. Read the comments to see the disagreements about the Frequentist and Bayesian approach.
- Levy 2012 is also a good introduction.
- Credible intervals are described by the authors of the easystats.
- testscience.org has a brief example that compares confidence intervals with credible intervals.
- Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification by Liu, Ye, and Ling, 2012 (https://doi.org/10.1029/2011WR011289)
- `bayes_mvs` in SciPy uses https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1277&context=facpub

### 8.2.4 MCMC

- https://twiecki.io/blog/2015/11/10/mcmc-sampling/
- Data analysis recipes: Using Markov Chain Monte Carlo (2017), Hogg and Foreman-Mackey (available from arxiv.org and iopscience.iop.org); Provides an overview of the motivation for MCMC sampling, discussions of its use and abuse, and detailed recipies for implementation.
- Ravenzwaaij, Cassey, and Brown, 2018

### 8.2.5 Misc

- https://stats.stackexchange.com/questions/421221/bayesian-inference-feeding-posterior-back-in-as-prior

- https://projecteuclid.org/journals/statistical-science/volume-32/issue-1/How-Principled-and-Practical-Are-Penalised-Complexity-Priors/10.1214/16-STS603.full
- http://www2.denizyuret.com/ref/aitkin/posterior-bayes-factors.pdf
- https://towardsdatascience.com/what-is-rejection-sampling-1f6aff92330d
- https://stats.stackexchange.com/questions/2356/are-there-any-examples-where-bayesian-credible-intervals-are-obviously-inferior
- https://link.springer.com/article/10.3758/s13423-016-1221-4

### 8.3 Astronomy

1. Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy (2017), S. Sharma.
2. Bayesian Methods for the Physical Sciences: Learning from Examples in Astronomy and Physics (2015), S. Andreon and B. Weaver.
3. Bayesian Models for Astrophysical Data: Using R, JAGS, Python, and Stan, J.M. Hilbe, R.S. de Souza, and E.E.O. Ishida.
4. Modern Statistical Methods for Astronomy: With R Applications (2012), E.D. Feigelson and G.J. Babu.
5. Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data, Željko Ivezic, Andrew J. Connolly, Jacob T VanderPlas, and Alexander Gray.
6. Computational Bayesian Statistics: An Introduction (2019), M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller.

### 8.4 Software

1. Think Stats - A free book that demonstrates standard statistical calculations in Python.
2. EMCEE Overview EMCEE software documentation

## 9 General Policies

### 9.1 Academic Integrity

Any instance of cheating or plagiarism is a violation of the Honor Code Pledge and will result in a score of zero on the exam or paper and referral to the Honor Committee. The website for the Office of Academic Integrity is https://oai.gmu.edu/.

### 9.2 Disability

If you have a disability and need academic accommodations, please contact Disability Services. Their website is https://ds.gmu.edu/. All academic accommodations must be arranged through Disability Services.

### 9.3 Diversity and Inclusion

We seek to create a learning environment that fosters respect for people across identities. We welcome and value individuals and their differences, including gender expression and identity, race, economic status, sex, sexuality, ethnicity, national origin, first language, religion, age, and ability. We encourage all members of the learning environment to engage with the material personally but to also be open to exploring and learning from experiences different than their own. Mason's nondiscrimination policy is at https://universitypolicy.gmu.edu/policies/non-discrimination-policy/.

## 9.4 Communication

If you have a question whose answer may be of interest to other students, please post it to Discord and make it visible to other students. You can set the post so that you are anonymous.

If you need to send communicate with me about something private, please send it to `rweigel@gmu.edu` from your MasonLive email address.

## 9.5 University Resources

- Learning Services https://learningservices.gmu.edu/
- Student Support and Advocacy Center https://ssac.gmu.edu/
- Counseling and Psychological Services https://caps.gmu.edu/