

CS-0149: Introduction to Statistics Using Baseball

126 Adele Simmons Hall; Tuesday & Thursday 2:00-3:20PM

Instructor: Ethan Meyers **Office:** 133 ASH
Email: emeyers@hampshire.edu **Phone:** 559-5500

Office Hours: Office hours are Monday from 2:30-3:30pm, Tuesday from 12:30-1:30pm and by appointment. Please feel free to email me to set up an appointment.

About the Course

Course Objectives: The goal of this course is to provide students with an understanding of statistical methods and reasoning that are useful for analyzing baseball and other types of data. Additionally students will learn how to use the R programming language to analyze data.

Distribution: Mind, Brain, and Information

Course Website: <https://moodle.hampshire.edu/course/view.php?id=5532>

Course books and other material: The following books are required for this course and can be purchased from the Hampshire College Bookstore:

Albert, J. (2003). *Teaching Statistics Using Baseball*. Washington, D.C.: The Mathematical Association of America.

Sawchik, T. (2015). *Big Data Baseball*. New York, N.Y. Flatiron Books.

Additional reading material will be posted to the class website.

Recommended books:

Marchi, M., Albert, J. (2014). *Analyzing Baseball Data with R*. Boca Raton, F.L.: CRC Press.
Lewis, M. (2003). *Moneyball*. New York, N.Y.: W. W. Norton & Company.

Course Format: The format of the course will be lectures, discussions, and in-class activities. There will be some overlap between the material covered in class and the material covered in the book. Students must complete the assigned readings as well as attend class in order to master the material covered in this course. Adjustments to the required reading may be made frequently, which also makes it important to attend class to hear announcements concerning the reading assignments.

Expectations for work completed: In this course students are expected to spend at least six to eight hours a week of preparation and work outside of class time. This time includes reading sections of the textbook, completed assigned problems, and analyzing data using R.

Course topics: The course will cover descriptive statistics, inferential statistics, and more advanced topics that are applicable to analyzing baseball data and data from other fields.

Course Assignments: The course assignments will consist of readings, weekly worksheets where you will use R to analyze data, and a final project where you will do a more in depth analysis investigating one specific question. You will also be required to do a presentation on one baseball player or on another topic related to the class, and a class presentation on your final project. The questions presented on the worksheet problems may be discussed with other members of the class, but the work turned in must be your own since it is important that you understand all the material yourself. Obviously copying from other students is not allowed. Everyone who you discussed the worksheet with must be listed on the top of the worksheet.

Access to R: Will we use the R programming language to analyze data in this class. The easiest way to do these assignments will be to use the online version of RStudio which can be found at: <https://asterius.hampshire.edu/>

General Course Policies

Due Dates: In order to understand the material covered in class it is **critical** that you keep up with the assignments. Thus **all assignments must be turned in on time** as indicated on the class website. **Failure to turn in work in on time will result in receiving a no evaluation for this class.** In particular a three strikes policy will be implemented in this class: **if you fail to turn in three assignments on time you will not receive an evaluation for the class.** Extensions may be granted in rare circumstances if you contact me well in advance to get my permission but otherwise will not be granted.

Attendance: Class attendance is a requirement of this course. I also expect you to be to class on time. Missing class will have an impact on your evaluation.

Laptops: If you own a laptop, you may bring it to class in order to take notes and complete in class exercises, although it might be easier to simply use the computers that are already in the classroom. However, obviously you should not use your laptop for email, web browsing or any other activity that is not related to the class.

Email: Email from this class will be sent to your Hampshire account. Make sure to check this email account at least twice a day. You can always have your Hampshire email forwarded to another account if that is easier for you.

Requirements for Evaluation: Students are required to attend class, do the assigned readings, and complete all the assigned worksheets. **Any student who does not turn in the assignments on time will NOT receive an evaluation for the course.**

Incompletes: An “incomplete” will be granted at the discretion of the instructor only under unusual and incapacitating circumstances. If you miss more than two classes, or feel that you are in danger of not completing the course for any reason, please come see me as soon as possible. Any student seeking an incomplete must (1) request the “incomplete” in writing *prior to the last week of class*, (2) provide appropriate written documentation of the illness or circumstances, and (3) make specific arrangements with the instructor to complete the required coursework. Students will have four weeks to complete any missed exams/assignments to replace the incomplete with an evaluation. Failure to complete the required work within this time frame will result in a “no evaluation”.

Special Needs: I encourage students with disabilities or special needs to discuss with me whatever arrangements may be needed during the first week or two of class. Please visit Aaron Ferguson, Director, Office of Accessibility Resources and Services, at x5498 or accessibility@hampshire.edu, or stop by his office in Lemelson. Getting the proper documentation in ensures you can receive appropriate accommodations.

Academic dishonesty: All students and faculty, whether at Hampshire or another institution are bound by the ethics of academic integrity. Academic dishonesty includes: cheating on assignments; plagiarizing (misrepresenting as one’s own anything done by another, including quoting from readings without acknowledging sources); submitting the same or substantially similar papers for more than one course without consent of all instructors concerned; depriving another of necessary course materials; sabotaging another’s work; and fabricating data, information, or citations.

All cases of suspected plagiarism or academic dishonesty will be referred to the Dean of Advising who will review documentation and meet with student and faculty member. Individual faculty, in consultation with the Dean of Advising, will decide the most appropriate consequence in the context of the class. This can range from revising and resubmitting an assignment to failing the course. Beyond the consequence in the course, CASA considers first offenses as opportunities for education and official warning. Multiple or egregious offenses will have more serious consequences. Suspected instances of other breaches of the ethics of academic integrity, such as the falsification of data, will be treated with the same seriousness as plagiarism and will follow the same process. For further details about Hampshire’s policies regarding the Ethics of Scholarship please see Non Satis Non Scire: <https://handbook.hampshire.edu/node/87>

Course Outline

(Schedule will likely change as the semester proceeds so, please check the class website)

- Jan. 26 **Introduction and overview of the class**
TSUB: Chapter 1
- Jan. 31 **An introduction to baseball and R**
TSUB: Chapter 2
Big Data Baseball Chapter 1
- Feb. 2 **Summary statistics and plots for a single batch of data**
Big Data Baseball Chapter 2
Worksheet 1 - Due at 11:59pm on Sunday February 5th
- Feb. 7 **Exploring categorical and quantitative data**
TSUB: Chapter 2
- Feb. 9 **Quantifying variability**
Big Data Baseball Chapter 3
Worksheet 2 - Due midnight on Sunday February 12th
- Feb. 14 **Quantifying variability continued**
TSUB: Chapter 3
Worksheet 3 - Due midnight on Sunday February 19th
- Feb. 16 *Advising day – no classes*
- Feb. 21 **Relationship between measurements: correlation**
Big Data Baseball Chapter 4
- Feb. 23 **Relationship between measurements: regression 1**
TSUB: Chapter 4
Worksheet 4 – Due midnight Sunday February 26th
- Feb. 28 **Relationship between measurements: regression 2**
Big Data Baseball Chapter 5
- Mar. 2 **Multiple regression**
Worksheet 5 – Due midnight Sunday March 5th
- Mar. 6 **Data manipulation**
TSUB: Chapter 5
- Mar. 8 **Descriptive Statistics review**
- Mar. 14, 16 *Spring break – no classes*

- Mar. 21 **Introduction to Probability using table top games (all-star baseball)**
TSUB: Chapter 5
- Mar. 23 **Probability using table top games continued (Strat-o-Matic)**
Big Data Baseball Chapter 6
Worksheet 6 – Due midnight Sunday March 26th
- Mar. 28 **Tree diagrams and the binomial distribution**
TSUB: Chapter 5
- Mar. 30 **Binomial and Normal distributions**
Spring training games begin for most teams
Big Data Baseball Chapter 7
Worksheet 7 – Due midnight Sunday April 2nd
- Apr. 4 **Introduction to statistical inference**
TSUB: Chapter 6
- Apr. 6 **Hypothesis tests for a single proportion**
Big Data Baseball Chapter 8
Worksheet 8 – Due midnight Sunday April 9th
- Apr. 11 **Hypothesis tests for two proportions**
TSUB: Chapter 8
- Apr. 13 **Hypothesis tests for two proportions and two means**
TSUB: Chapter 9
Worksheet 9 – Due midnight Sunday April 16th
- Apr. 18 **Randomization tests for comparing two or more means**
Big Data Baseball Chapter 10
- Apr. 20 **Parametric tests for comparing two or more means**
Worksheet 10 – Due midnight Sunday April 23rd
- Apr. 25 **Confidence intervals**
Big Data Baseball Chapter 11
- Apr. 27 **Review**
Worksheet 11 – Optional
- May 2 **Final project presentations**
- May 4 **Final project presentations, review and conclusions**
Big Data Baseball Epilogue