

Statistics and Sports

Spring 2016



Instructor: Michael Lopez

Office: Harder 217

Email: mlopez1@skidmore.edu

Twitter: @StatsbyLopez

Phone: 580-5297

Office Hours: Wednesday 10-12 or by appointment

Class Meetings: TR 12:40-2:00 (Harder 101)

Text (optional): *Introductory Statistics with Randomization and Simulation* by Diez et. al

Link to download: <http://www.openintro.org/stat/textbook.php>

Link to purchase (Amazon): <http://www.amazon.com/dp/1500576697>

Course Goals: Students will:

- Gain an understanding of how statistical tools have been applied in the sports world
- Learn advanced methods for analysis, visualization, and implementation
- Gain familiarity with nonparametric and parametric statistical tools
- Contrast in-sample versus out-of-sample statistics with respect to sports

Course Skills:

- Computing skills and basic webscraping using *R* & *R-studio*
- Reproducible analysis using *RMarkdown*
- Multivariate linear regression, logistic regression, model diagnostics

Textbook: Readings will be posted on blackboard. There is no textbook

Computing: The use of the R statistical environment (downloadable from <http://www.r-project.org/>) with the RStudio interface (downloadable from www.rstudio.org) is thoroughly integrated into the course. Access on or off campus can be found at: <http://r.skidmore.edu>

Writing: Your ability to communicate results, which may be technical in nature, to your audience, which is likely to be non-technical, is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

Grading:

Homework & Participation [15%] Homework is the most effective way to reinforce concepts learned in class. There will be weekly homework assignments. Most often, questions will relate to material in the reading that will be covered in class. Doing statistics is the best way to learn statistics. You are welcome to collaborate with other students, but you must turn in your own work and write up all assignments in your own words.

Wherever possible, RMarkdown will be used as to ensure reproducible work and of a standard format.

On the top of every homework, write the names of *everyone* you collaborated with in doing problems. A basic principle of scholarship is that once gives credit to all who contributed to the findings. Copying and pasting sentences, paragraphs, or blocks of *R* code from another student is not acceptable and will receive no credit. All students, staff and faculty are bound by the Skidmore College Honor Code.

Late homework is not accepted. I will drop the lowest homework score.

Homeworks are graded out of 5 points:

- 1-3 out of 5 points: Most questions attempted, minimal effort
- 4 of 5 points: All questions attempted, complete effort, graded questions incorrect
- 4.5 of 5 points: All questions attempted, complete effort, graded questions partially correct
- 5 of 5 points: All questions attempted, graded questions perfect

Additionally, 5% of your participation grade will be based on your readings before the Tuesday classes.

Project Presentation and Technical report [25%] The major milestone in this class will be conducting (with a group) a statistical investigation on a question of interest to you. For each, you may collect primary data by hand or you can use data available on the Internet. You may also use data that we have used in class, as long as it is for a different project.

You will prepare a project proposal describing your study and obtain approval from me before you begin the investigation. During the last week of class, you (and your group) will give a 10-minute oral presentation of your study. We will spend time in class looking at what data is available on the web and about writing a project proposal. You will write up your report using RMarkdown and submit it before finals week.

Exams [20% each] There will be three take-home exams (the third one will come during exam week). These will involve write-ups of a statistical analysis, done using RMarkdown.

Disability accommodations: Any student with special needs requiring accommodations should give me his/her memo of accommodations in a timely manner. It is the student's responsibility to follow up with me regarding all accommodations that require my participation. The student is advised to ensure full use of testing accommodations by coming to talk to me at least three days before any test.

Attendance: Your attendance in class is crucial, as is your punctuality. We are all going to learn this material together, so we need to have everyone present and working. Accommodations for an unavoidable absence can be made in advance via email; one necessary absence during the semester is not unusual; having more than two is uncommon.

Reading: Readings will be posted online. Please see the class website, statsbylopez.com, for details

Collaboration: Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete homework assignments, and prepare for exams. However, every word that you write must be your own. Copying and pasting sentences, paragraphs, or blocks of *R* code from another student is not acceptable and will receive no credit. No interaction with anyone but the instructor is allowed on any exams or quizzes. All students, staff and faculty are bound by the Skidmore College Honor Code.

Guest speaker etiquette:

We are planning at least 4 guest speakers. Some comments:

- Come prepared - do reading
- DO:
 - Be respectful
 - Ask good questions
 - Share reasonably-held opinions & analytics facts
- DO NOT:
 - Be disrespectful
 - Be unprepared
 - Not be able to support your opinions
 - Get in to issues of fandom
 - Bug them before/after class

Additional policies:

Cell phones: If your phone rings, I get to answer it! Respect the rest of the class and turn your phone on silent, airplane mode, or off

Computers: Each Thursday will feature a lab. If you have a laptop you can bring, please bring it to that class. Tuesday will consist of a more formal discussion and notes.

Calendar (subject to change)

Date	Topics	Assignments
26-Jan	What is 'statistics in sports'?	
28-Jan	Lab: Intro to R/RStudio/RMarkdown	HW 0
2-Feb	Player valuation in baseball	
4-Feb	Lab: Correlation, R-squared, scatter plots	HW 1
9-Feb	Multiple linear regression modeling	
11-Feb	Lab: Baseball analysis, Lehman database	HW 2
16-Feb	NFL - Logistic regression, kickers, expected points*	
18-Feb	Lab: NFL kickers	HW 3
23-Feb	Football - game theory: 4 th downs, field goals	
25-Feb	Lab: Simulations in R.	HW 4
1-Mar	<i>Take home exam 1: football and baseball^</i>	
3-Mar	Lab: March Madness, gambling, game theory	
8-Mar	Basketball – possessions, shot difficulty**	
10-Mar	Lab: Shot charts, hot hand theory	HW 5
22-Mar	Hockey – shooting stats, dump-ins,	
24-Mar	Lab: OT likelihood, prisoners dilemma	HW 6
29-Mar	Stein's estimator, shrinkage, mean reversion	
31-Mar	Lab: Shootouts & shooting percentages	HW 7
5-Apr	Referee analytics	
7-Apr	Lab: NHL/NFL refs (make up calls, late game bias)	HW 8
12-Apr	Power rankings, Elo, Bradley Terry ***	
14-Apr	Lab: BTM in R, NBA power rankings	HW 9
19-Apr	<i>Take home exam 2: basketball and hockey^</i>	
21-Apr	Soccer analytics	
26-Apr	Lab – expected goals****	
28-Apr	Project work	HW 10
3-May	Project presentations	
7-May		<i>Technical Report</i>
??-May		<i>Take home exam 3</i>

Guest speakers (tentative)

*Ben Morris, FiveThirtyEight

**Seth Partnow, Nylon Calculus

***Sharon Katz, ESPN Stats & Info

****Sarah Rudd, StatDNA

^Take home exams are posted on Monday, due Thursday at noon.