# STATS 50: Mathematics of Sports

**Stanford University**
**Quarter**: Spring 2018
**Room**: Hewlett TC 103
**Time**: Mon, Wed, Fri
2:30 PM - 3:20 PM


**Instructor**: Maxime Cauchois
maxcauch@stanford.edu
Office: Sequoia 220 Fishbowl
Hours: Mondays 4-5pm

**TA**: Damian Pavlyshyn
damianp@stanford.edu
Office: Sequoia 207 Bowker
Hours: Thursdays 2-4pm

**TA**: Marius Tirlea
mtirlea@stanford.edu
Office: Sequoia 220 Fishbowl
Hours: Wednesdays 12-1pm


## Abstract


The spring 2019 edition of Stats 50 (cross-listed as MCS 100) will be bent towards data exploration. The overlying question of this course is how to leverage all the data available in sports to provide professional sports players with meaningful and useful insights on their own performance. We will most notably focus on the evaluation of individual and collective performance (covering regression to the mean or regularized adjusted plus-minus), adaptive in-game strategies (who should shoot the ball?), but also fairness in sports (how to make sure that all participants have equal chances?). A few guests speakers will shed light on more specific problems encountered in professional sport.

We will review linear regression early in the course and use the statistical programming language R to implement the ideas discussed. We recommend that you have taken an introductory statistics course prior to this one, but this is not mandatory. Exposure to linear algebra and basic probability theory is however strongly advised.


## Goals of the course

By the end of the course, the hope is that you have learned how to:

- understand and interpret advanced statistics reported in the media;
- evaluate individual players in terms of how their performance leads to team success;
- interpret players statistics in small sample sizes and against varying opponent quality;
- use data to inform in-game strategic decisions in a variety of sports.

## Tentative schedule

Week 1: Introduction, survey of data and stats in sports
Week 2: Linear regression and Markov Chains
Week 3: Player evaluation: wOBA and Bayesian framework
Week 4: Regression to the mean
Week 5: Bradley-Terry model
Week 6: Regularized adjusted plus-minus
Week 7: In-game strategy
Week 8: Miscellaneous topics

Week 9: Final project presentations
Week 10: Final project presentations, continued

## R

For every statistical technique learned in this class, we will also learn how to implement in on real data in R, the free statistical language. Friday is R day, meaning that students are expected to bring laptops to class on Friday and follow along by running the same code in R sessions on their own laptops. Visit `www.r-project.org` to download R for free before the first R day (Friday, April 5).

## Textbook

There is no required text for the course. Some useful references are:

- The Book: Playing the Percentages in Baseball, by Tango, Lichtman and Dolphin
- Analytic Methods in Sports, by Severini
- Mathematics in Sport, by Townend
- Mathletics, by Winston
- Optimal Strategies in Sports, by Ladany and Machol

## Course Material

Some class notes, the syllabus, and announcements will be available on Canvas (see `canvas.stanford.edu`).

## Grading

The grading will be divided into three parts:

- *Homework (40%).* Four homework assignments will be due on April 17, May 1, May 8 and May 15 (all Wednesdays, subject to change). Each assignment will be announced at least one week in advance. Each homework submission should be typeset on a computer, and then submitted on Gradescope: think of it as a report you would have to write for the front-office of a professional sport team! It also should be confined to one page, front and back. To clarify, this means that two pages is the maximum length allowed for a homework submission. Students may work in pairs on homework assignments.
- *Project (40%).* The final project is a research paper on the subject of your choosing. Ideally, your project should include the following steps: pose a research question with respect to sports performance or analysis, gather some relevant data, and analyze it using the different methods and techniques seen in class, or even some statistical tools out of scope. In the end, your project should try to give a concrete answer to the problem at stake. The maximum length for the research paper is eight pages (i.e. four front and back), and the deadline is Wednesday, May 29. You will be expected to give a 10-minute presentation on your work in the final weeks of the course. Students are encouraged to work in pairs or in three-person teams on the final project.

- *Participation (20%).* Class participation makes up a very significant portion of the grade, so it is in your best interest to speak up during class discussions. Half of the class participation grade is based on attendance on Fridays (R days) and on days when guest speakers visit the class. The other half is based on actually talking during class discussions. We want to hear from you, and do not hesitate to give us feedback on the course!