Sports Analytics: Course Description and Syllabus

Time and room

The class meets on Wednesdays from 9:00 am–12:15 pm in Warren 207
Dates: First class January 28, 2015; last class March 4, 2015

Instructor

Professor Mark Broadie
Office: 415 Uris Hall, fourth floor
E-mail: mnb2@columbia.edu
Phone: 212-854-4103
Office hours: Wednesdays, 12:30 pm–2:00 pm (or by appointment)

Teaching assistant

TBA

Course description

Sports analytics refers to the use of data and quantitative methods to measure performance and make decisions to gain advantage in the competitive sports arena. This course builds on the Business Analytics core course and is designed to help students to develop and apply analytical skills that are useful in business, using sports as the application area. These skills include critical thinking, mathematical modeling, statistical analysis, predictive analytics, game theory, optimization and simulation. These skills will be applied to sports in this course, but are equally useful in many areas of business. Typical questions addressed in sports analytics include: How much is a player on a team worth? How to rank players or teams? How to predict future performance of players or teams? How likely are extreme performances, i.e., streaks? Which decision is more likely to lead to a win (e.g., attempt a stolen base or not in baseball, punt or go for it on fourth down in football, dump and chase or not in hockey, pull the goalie or not in hockey)?

Class sessions will involve a mixture of lecture, discussion, and hands-on analysis with computers in class. Each session will typically address a question from a sport using an important analytical idea (e.g., mean reversion) together with a mathematical technique (e.g., regression). Because of the “laboratory” nature of part of the sessions, students should bring their laptops to each class.

Prerequisites and connections with core courses

The prerequisites for the course are the Managerial Statistics and Business Analytics core courses. Good Excel skills are required and we will use the Business Analytics Excel add-in for analyzing
data. Detailed knowledge of sports is not required, but familiarity with the rules of baseball, football, basketball, hockey and golf will be assumed. There are many websites and popular books where you can learn about these sports, if needed.


**Recommended books**

- *Every Shot Counts*, 2014, Mark Broadie, Gotham Books
- *Trading Bases*, 2013, Joe Peta, Dutton
- *Scorecasting*, 2011, Moskowitz and Wertheim, Crown Archetype
- *Analyzing Baseball Data with R*, 2013, Chapman & Hall/CRC

**Course work and grading**

There will be a homework assignment for every class. These may be done individually or in a group of up to three people (which can change from assignment to assignment). If you would like help finding a team, please e-mail the TA. Assignments will be a mix of data analysis, thought questions, and reading. In computing your homework assignment grade, the lowest score will be dropped.

The final project can be on a sports topic of your choosing and can be done in groups of three or less. A good level to shoot for would be an article that could be published on [fivethirtyeight.com/sports](http://fivethirtyeight.com/sports) or [sportsanalyticsblog.com](http://sportsanalyticsblog.com).

The project could also be on a topic from the *Mathletics* book not covered in class or from the academic literature. You could investigate an interesting result from one sport and apply it to another sport. The write-up should be approximately five pages, not including figures, charts, or supporting material in an appendix. Project presentations will be given on the last class session.

Class participation (present and on time, prepared and participating) is important for learning the material and for us learning from each other—and is more fun for all of us.

Grades will be based on the following weights:

- Final project and presentation: 40%
- Class participation (including attendance): 10%
- Homework (lowest score dropped): 50%

**Software**

All assignments are expected to be done in Excel (and possibly together with the Business Analytics add-in). There is no requirement to use other software for the assignments (though we might use Visual Basic or R in class). However, if you are familiar with another statistical package or programming language, you are welcome to use it for the project (e.g., you might want to use R if your project involves analysis of pitch/fx data).
Course outline (subject to change)

Session 1. Rating sports teams and players

- Course overview
- Which basketball team is the best? Which team has the best offense?
- How to account for the strength of the schedule or the strength of the field?
- Applications of logistic and fixed effects regression models
- Reading: Mathletics, chapters 40 and 43

Session 2. Baseball and football decision-making

- Baseball: analysis of bunting and base-stealing strategies
- Football: analysis of run versus pass, punt or go-for-it
- Win probability added, player value added
- Applications of state space / Markov chain and game theoretic methods
- Reading: Mathletics, chapters 6, 20–24

Session 3. Fantasy sports

- Draft strategies, daily leagues, accounting for variability
- Performance prediction: exponential smoothing, mean reversion
- Application of integer and nonlinear optimization
- Reading: Mathletics, chapters 40 and 43

Session 4. Baseball

- Guest speaker, Professor Jerry Kim: Are umpires biased? An analysis using PITCHf/x data
- Analysis of streaks
- Applications of simulation
- Reading: Mathletics, chapter 11
- Optional reading: “Seeing stars: Matthew effects and status bias in major league baseball umpiring,” Management Science, J. Kim and B. King.
Session 5. Golf

- Measuring golf performance: What is the key to Tiger’s success?
- Optimal strategies in golf: trading off risk and reward
- Applications of simulation
- Optional reading: *Every Shot Counts*, chapters 5 and 8

Session 6. Project presentations