

## **Sports and Gaming Analytics**

### **Proposal and Syllabus**

Professor Abraham Wyner

[ajw@upenn.edu](mailto:ajw@upenn.edu)

303 ARB

**Monday, April 29<sup>th</sup>- Friday, May 3<sup>rd</sup>. Philadelphia, Campus.**  
**½ credit course**

### **Overview**

The “Moneyball revolution” in sports and inspired great interest in the transformative potential of statistics. This “1/2 credit course will introduce students to the growing field of sports analytics while creating for students an opportunity to practice and improve their analytical skills on real problems that are accessible and fun for anyone with an interest in sports and a foundational knowledge of statistics.

While the context will be sports related and the expectation of students is that they are interested and knowledgeable about most major sports, the skills and techniques gained will be widely applicable and generalizable with applications in diverse areas. The course is very applied and very data driven. Students will conduct hands on work with real data using JMP software, R or Python. Along the way, students will learn new techniques for analyzing data and gain practical and useful skills that will be broadly applicable across many areas.

**Pre-requisites:** Statistics 6130 or 6210. Class size will be limited.

**Format:** The course will meet for three hours of lecture, discussion and hands on work every morning for 5 days. Students will work in groups on various in-class data analysis tasks and present results. There will be office hours and group work every afternoon. Guests from a subset of the local Philadelphia area sports teams (Flyers, Phillies, Eagles, Union, Sixers) will talk with the students about analytics in their organization.

There will be two graded assignments. The first assignment will be a simple prediction contest using Kaggle. The second assignment can be a replication and extension of a published analysis or an entirely original project that addresses a specific sports related managerial decision on or off the field. Projects, in both parts, will be collaborative involving teams of 3-6 students.

### **Grading:**

- 25% Class participation
- 25% Prediction Assignment
- 50% Final Presentation.

## **Class Content and Course Schedule:**

1. **Monday, April 29<sup>th</sup>. Regression Bootcamp.** A refresher course in both univariate and multivariate linear and non-linear regression. Virtual Class (asynchronous, because of Passover holiday).
2. **Tuesday, April 30<sup>th</sup>. Regression Bootcamp Continued:** The material will be presented in a series of applications to simple problems in sports and gaming.

**Applications:** Predicting Batting Averages, study the NFL Combine, and build models for predicting Putt success in golf and field goal success in football.

- Nationality Bias in Olympic Diving. Non-parametric statistical methods for estimating p-values. Controlling for confounders and causal inference.
- Moneyball Modeling (discovering the relationship between payroll and winning).
- Pythagorean Theorems (modeling the relationship between scoring and winning).

Afternoon: Prediction Contest (Kaggle).

### **3. Wednesday, May 1<sup>st</sup>. Statistical Methods and Applications:**

- i) Multiple Regression Modeling: Building a power rank from historical data on game outcomes. Measuring Park-Factors in baseball.
- ii) Logistic Regression, ELO Model and Bradley-Terry models: Building power ranks and modeling team performance using win-loss records. Measuring the home-field advantage. Studying different measures of success: How do we judge the accuracy of a prediction problem.
- iii) Machine Learning: Tree models. Random Forests, Boosting and Neural Networks applied to prediction model in sports. Win Probability Models, Expected Points Modeling. JMP software will be used which allows speedy introduction of new methods through its visual platform.

Afternoon: Group Project work and office hours. Guest Speaker.

### **4. Thursday, May 2<sup>nd</sup>. Probability and Gaming:**

- i) Expected value and Risk. The St. Petersburg Paradox. The Martingale betting strategy.
- ii) Lotteries and the reverse lottery.
- iii) Sports betting and implied odds. Looking for an edge using football bets.

- iv) The stock market through the lens of gambling. Log Optimal portfolios and Kelly betting.
- v) Birthday problems and application to overfitting and back testing.
- vi) Monty Hall, selection bias and the power of hidden information.
- vii) The Hot Hand. Is it real?

Afternoon: Group work on Projects, office hour. Guest Speaker.

## **5. Friday, May 3<sup>rd</sup>. The insights that Changed the Game: Real World Sports Analytics.**

- i) Baseball: Player Evaluation. In-Game strategy (To Steal/Bunt or not).
- ii) Football: Fourth Down Decision making.
- iii) Golf: Stokes Gained.
- iv) Basketball: The 4-factors. Plus-Minus. Adjusted Plus-Minus and Regularized Adjusted Plus Minus.
- v) Soccer: Expected Goals (xG) vs. Goals.

**Afternoon: Presentations of Student Work.** Students will present their projects to the class. Presentations will be 10-15 minutes in length.

### Potential Reading List:

1. **Moneyball**, Michael Lewis.
2. **Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won** [L. Jon Wertheim](#), [Tobias Moskowitz](#)
3. **Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football** 2012 by [Wayne L. Winston](#) (Author).
4. Emerson, John W., et al. “**Assessing Judging Bias: An Example From the 2000 Olympic Games.**” *The American Statistician*, vol. 63, no. 2, 2009, pp. 124–131. *JSTOR*, [www.jstor.org/stable/25652240](http://www.jstor.org/stable/25652240).
5. TVERSKY, A. and GILOVICH T. (1989b), “**The Hot Hand: Statistical Reality or Cognitive Illusion?**” *Chance*, 2, 31–34.