

Spring, 2025

## Stat 470 – Sports Analytics: Modeling and Methodology

**Instructor:** Beth Chance

**TA:** Andrew Martinez

**Class Time:** Tuesdays/Thursday 4-6pm 38-122;

**Office Hours:** M 4-5pm, T 10-11am, W 4-5pm, R 10-11am, F 2-3 or by appointment, including Calendly, or any time my door is open

**Office:** Faculty Office Building East 25-235

**Email:** bchance@calpoly.edu (a very good way to reach me)

**Course Listserv:** stat-470-01-2254@calpoly.edu

**Course Webpages:** Canvas (<http://my.calpoly.edu>)

**Co-requisite:** Stat 302/313; Recommended 324/334 and Stat 331

**Course Objectives:** This course introduces students to the fundamental principles and methods of sports analytics. Analytic topics may include web scraping, data wrangling, metric estimation and evaluation, regression/predictive modelling, optimization, ranking, spatial statistics, multilevel modelling, Bayesian inference, clustering methods. Use of R/RStudio/Quarto. Discussion of topics pertinent to most professional sports and business operations, including how to communicate ideas to players, coaches, and general managers.

**Learning Outcomes:** By the end of this course students will be able to

1. Obtain sports data and understand its limitations
2. Implement and interpret methods used in existing sports analytics research
3. Critically analyze the merits and limitations of existing sports analytics research
4. Develop their own research question in sports analytics and conduct a data analysis

### **Texts/Materials:**

Required: *Introduction to Sports Analytics using R* by R Elmore and A Urbaczewski

Assigned readings

Bring a laptop with you to class each day

**Groups:** You will be assigned to sports group. Each group will be expected to do a mini-presentation, probably Tuesday, giving an overview of how sports analytics is used in the sports. You should bring in information from the text as well as external sources. You should also describe an interesting case study in your sport, as well as any key data repositories for that sport. This will be the same group you complete the final project with. Step one is choose a team name!

### **Grading:**

Assignments 30%

Presentations 20%

Participation/Discussion 10%

Exam 15%

Final project/presentation 25%

I anticipate that we will have cutoffs like 70/72/78/80/82/88/90/92/ for C-/C/C+/B-/B/B+ etc.

*Spring, 2025*

**Participation.** Participation in this class is an important part of your learning experience. While attendance is not required for every lecture, it is strongly encouraged. Participation will be assessed through in-class discussion, including journal articles, office hour visits, and attendance/questions for guest speakers. You will be expected to give one group presentation Weeks 2-8.

**Final project.** One of the most exciting parts of this class is the opportunity to explore a sports analytics question of your choosing. To reflect real-world scenarios where team members are often assigned, project groups will be **formed by me** with **6 members**, based on a brief survey of student preferences. This ensures a diverse range of skills and perspectives in each team. To help you stay on track and receive feedback, there will be two intermediate deadlines before the final project write-up is due: a **project proposal** and a **project progress report**. These checkpoints will allow me to guide you in refining your project idea and analysis. During Week 10, you will make a YouTube short of your project. During the final exam time period, all groups will present their findings to the class.

The final project constitutes 25% of your overall grade, distributed as follows:

- **Project Proposal:** 5%
- **Project Progress Report:** 5%
- **Final Write-Up:** 10%
- **Presentation:** 5%

Detailed expectations and grading criteria for the final project will be shared later in the course.

Spring, 2025

**Tentative schedule**

<b>Week</b>	<b>Dates</b>	<b>Topics</b>	<b>Speaker</b>
1	4/1	Ch. 1, 2: Intro to Course and R, Careers in Sports Analytics	Ethan Moore
	4/3	Ch. 3: R, Web scraping, APIs, Shiny apps	Mia Hodges, Nate Ngo
2	4/8	Ch. 4: Baseball, Multiple regression, GAMs	Baseball I
	4/10	Pythagorean expectation, Regression to the mean	Ben Baumer
3	4/15	Logistic regression	Baseball II
	4/17	SQL, Tableau, Dashboards	
4	4/22	Ch. 5: Basketball	Basketball
	4/24	Win probability models	
5	4/29	Ch. 6: Football, Wins above replacement	Football
	5/1	Ranking, Bradley Terry Models	Jared James
6	5/6	Ch. 7: Hockey, Multilevel models	Hockey, Sam Ventura
	5/8	Classification techniques	<b>Midterm</b>
7	5/13	Ch. 8: Soccer, K-means clustering	Soccer, Julia
	5/15	Tennis	
8	5/20	Ch. 9: Golf, Markov Chains	Misc, Brad?
	5/22	Volleyball	
9	5/27	Ch. 10 Sports Wagering	Rus Davtian
	5/29	Optimization, Linear programming	Andrew Martinez
10	6/3	Work Day	
	6/5	Work Day	
Final	Final Project Presentations		

### **Academic Integrity**

The University Code of Academic Integrity is central to the ideals that undergird this course. Students are expected to be independently familiar with the Code and to recognize that their work in the course is to be their own original work that truthfully represents the time and effort applied. Violations of the Code are most serious and will be handled in a manner that fully represents the extent of the Code and that befits the seriousness of its violation. You and your student peers must have a strong commitment to personal and professional integrity that informs your behavior both before and after graduation, discouraging you from creating a false appearance of achievement by presenting the work of others as your own, or bending or breaking the rules of any situation. This includes the unauthorized use and sharing of information/course resources on the internet.

### **COVID-19 Compliance, Classroom, and Campus Safety**

Cal Poly is committed to protecting the health and safety of the campus community. Taking preventative steps, as well as monitoring your health and staying home if you are feeling unwell, will help protect the entire Cal Poly community. By participating in this course, you agree to abide by all campus safety protocols. Please note that safety protocols may change throughout the quarter. You may wish to bring in your own device for computer work.

### **Support**

If you face emotional or economic challenges this quarter, you are not alone, and Cal Poly can help during this time of crisis. For example:

- Cal Poly's Basic Needs initiative: [basicneeds.calpoly.edu](http://basicneeds.calpoly.edu)
- Student Care Resources: <https://coronavirus.calpoly.edu/student-care-resources>
- Cal Poly Coronavirus website: <http://coronavirus.calpoly.edu/>
- Disability Resource Center: [drc@calpoly.edu](mailto:drc@calpoly.edu)
- Cal Poly's Counseling Services (805-756-2511)

If I can help you in any way to access the resources above, or if you have any questions about student care resources, please let me know, including textbook access.

### **Responsibilities**

Through classroom discussion and online communication, I welcome individuals of all ages, backgrounds, beliefs, races, ethnicities, social classes, genders, gender identities, gender expressions, national origins, documentation statuses, religious affiliations, sexual orientations, abilities – and other visible and nonvisible differences. All members of this class are expected to contribute to a respectful and inclusive environment for every other member of the class. This does not mean we cannot disagree or have different ideas. It does mean we try to consider perspectives other than our own, though they may differ from our own beliefs/experiences. If you experience disrespect or discrimination in this class, please report your experiences to me or the Statistics Department chair ([aschaffn@calpoly.edu](mailto:aschaffn@calpoly.edu)).