STAT 143: Course Information

Instructors

Laurie Shaw, Research Scientist & Lecturer, Department of Statistics, laurieshaw@g.harvard.edu
Christy Huo, Teaching Fellow, Department of Statistics, rhuo@g.harvard.edu

Class time & Location

Wednesday & Friday, 3-4.15pm (EST)
Zoom (connection details on canvas site)

Sections & Office hours (tentative)

Sections: Monday 9:00 am to 10:00 am and 9:00 pm to 10:00 pm (EST).
Office hours (Christy Huo): Tuesdays, 10:00am to 12:00pm (EST)
Office hours (Laurie Shaw): Thursdays, 1:00pm to 3:00pm (EST) (Zoom link)

Course Goals

The amount of data collected during games and tournaments has increased enormously over the last few decades, as has the demand from professional teams, national federations and the media for new techniques for analysing this data. This course will cover a wide variety of quantitative topics in sports analytics, including measuring team strength and predicting outcomes, situational analysis, player evaluation and working with tracking data. The course will be hands-on, including live data analysis in class and an emphasis on project work (group and individual). Play-by-play data for all sports covered during the course will be provided, as will Python code for implementing some of the methods that we discuss in class.

Prerequisites

This course applies statistical methods taught in STAT 110, STAT 111 and STAT 139. We will move through topics fairly quickly, so it is strongly advised that you complete these courses before taking STAT 143. The STAT 139 requirement may be waived if instead you have completed any of the following courses: CS 109A, CS 181, ECON 1123 or ECON 1126; however, you should have completed at least one of these courses plus STAT 110 and STAT 111 before taking STAT 143.

It is important that you have previous experience with data analysis in Python, R or Matlab. Play-by-play data and python code will be provided throughout this course to show you how to apply the

methods that we discuss in class and to complete the coursework and you will be expected to inspect, run and adapt this code for coursework assignments.

Python

Python is the main programming language used in this course to demonstrate the application of methods discussed during class. In order to run the examples, you will need to install python on your computer. If you have not previously used python, I recommend that you install the Anaconda distribution, which includes all the main libraries that we will use. Anaconda is freely available for download here.

If you have not previously used python: Google provides a free online python class which introduces most of the basic concepts. I suggest that you complete Lectures 1.1-1.3 and the associated problems during the first week of classes. To get you fully up to speed, in the first few weeks of STAT 143 the Sections will be used to teach a short 'crash-course' in data analysis in python (focussing on an introduction to the two most important libraries: NumPy and Pandas).

Coursework and problem sets

There will be two problem sets and two short projects issued during the course. You will have one week to complete the problem sets and two weeks to complete the short projects. There will also be a final project, which you will have two weeks to complete (including time in class).

Evaluation

There are no exams in this course. Instead, the course will be assessed based on coursework, problem sets and a final project. The total grade will be weighted as follows:

- Problem sets: 10%
- Short projects: 40%
- Final project report: 40%
- Final project presentation: 10%

Webpage

Everything of interest will be posted on the web: lecture notes, handouts, problem sets and their solutions, the course calendar, seminar transparencies, useful links, etc. The URL is https://canvas.harvard.edu/courses/85302

Textbooks & Background reading

Sports Analytics is a relatively new field and there is no single textbook that covers all the material in this course. However, many of the topics are covered in *Handbook of Statistical Methods and Analyses in Sports* (edited by Albert, Glickman, Swartz and Koning). The sections on baseball are also covered in *Teaching Statistics Using Baseball* (by Jim Albert).

Course Outline (subject to change)

Lecture #	Lecture Title	Sport	Date (day)
1	Introduction to Sports Analytics		1/27 (W)
Module 1: Measuring Team Strength and Predicting Outcomes			
2	Normal models of score differentials in the NFL (I)	Football	1/29 (F)
3	Normal models of score differentials in the NFL (II)	Football	2/03 (W)
4	Logistic models and the Elo system	Football	2/10 (W)
5	Poisson models for low-scoring sports	Hockey	2/12 (F)
6	Simulating matches and tournaments	Hockey/Soccer	2/17 (W)
Module 2: Situational analysis			
7	Valuing states (I. Markov Chains)	baseball	2/19 (F)
8	Valuing states (II. Applications to baseball)	baseball	2/24 (W)
9	Working with PITCHf/x data: Clustering pitch types	baseball	2/26 (F)
10	When should you go for it on fourth down? (I)	Football	3/3 (W)
11	When should you go for it on fourth down? (II)	Football	3/5 (F)
12	Guest Lecture	Football	3/10 (W)
13	xG: Measuring chance quality in soccer (I)	Soccer	3/12 (F)
14	xG: Measuring chance quality in soccer (II)	Soccer	3/17 (W)
Module 3: Player evaluation			
15	The +/- score (I. Basic and adjusted +/-)	Basketball	3/19 (F)
16	The +/- score (II. Regularized adjusted +/-)	Basketball	3/24 (W)
17	Streaks, momentum and the hot hand (I)	Basketball	3/26 (F)
18	Streaks, momentum and the hot hand (II)	Basketball	4/2 (F)
Module 4: Tracking data			
19	Introduction to working with tracking data	Soccer	4/7 (W)
20	Guest Lecture: William Spearman (Liverpool FC)	Soccer	4/9 (F)
Final Project			
21-23	<PROJECT WORK IN CLASS>		4/14,16,21
24	Project presentations		4/23

Academic Integrity Policy

Discussion and the exchange of ideas are essential to doing academic work. For assignments in this course, you are encouraged to consult with your classmates as you work on problem sets. However, after discussions with peers (or course instructional staff such as tutors, TF/TAs, course assistants), make sure that you can work through the problem yourself and ensure that any answers you submit for evaluation are the result of your own efforts. In addition, you must cite any sources, such as books, articles, websites, lectures, documents, etc., that have helped you with your work using appropriate citation practices. You are not permitted to use any documents that appear directly to answer the assignment questions. A good rule of thumb: if a fellow student asks if you would like to discuss a homework problem, we encourage you to say "yes"; if a fellow student asks to see your answer to a homework problem or python code in its entirety, the answer is "no."