 Statistics and Sports MS 251

Fall 2019

**Instructor**: Michael Lopez

**Office**: Harder 221

**Email:** mlopez1@skidmore.edu

**Twitter:** @StatsbyLopez

**Office Hours:** Friday 10-12 or by appointment

**Class Meetings:** WF 8:40-10:00

**Class Website:** <https://github.com/statsbylopez/StatsSports/>

**Text (optional):** *Introductory Statistics with Randomization and Simulation* by Diez et. al

**Link to download:** http://www.openintro.org/stat/textbook.php

**Link to purchase (Amazon):** http://www.amazon.com/dp/1500576697

**Course Goals**: Students will:

-Gain an understanding of how statistical tools have been applied in the sports world

-Learn advanced methods for analysis, visualization, and implementation

-Gain familiarity with nonparametric and parametric statistical tools

-Contrast in-sample versus out-of-sample statistics with respect to sports

**Course Skills:**

-Computing skills using *R* & *R-studio*

*-*Reproducible analysis using *RMarkdown*

-Multivariate linear regression, logistic regression, model diagnostics

**Textbook**: Readings will be posted on Github. There is no textbook. See <https://github.com/statsbylopez/StatsSports/blob/master/Readings.md>

**Computing:** We’ll use R & RMarkdown throughout the semester. You have two options:

1. Use your laptop, download the R statistical environment (downloadable from <http://www.r-project.org/>) and the RStudio interface (downloadable from [www.rstudio.org](http://www.rstudio.org))
2. Use the Skidmore server. Access on or off campus can be round at: <http://r.skidmore.edu>

**Writing:** Your ability to communicate results, which may be technical in nature, to your audience, which is likely to be non-technical, is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

**Grading:**

**Homework & Participation [25%]** Homework is the most effective way to reinforce concepts learned in class. There will be weekly homework assignments. Most often, questions will relate to material in the reading that will be covered in class. Doing statistics is the best way to learn statistics. You are welcome to collaborate with other students, but you must turn in your own work and write up all assignments in your own words.

RMarkdown will be used as to ensure reproducible work and of a standard format. Using RMarkdown, *only* HTML or PDF copies will be accepted. These must be printed and handed in during class.

On the top of every homework, write the names of *everyone* you collaborated with in doing problems. A basic principle of scholarship is that once gives credit to all who contributed to the findings. Copying and pasting sentences, paragraphs, or blocks of *R* code from another student is not acceptable and will receive no credit. All students, staff and faculty are bound by the Skidmore College Honor Code.

Late homework is not accepted. I will drop the lowest homework score.

Homeworks are graded out of 5 points:

1-3 out of 5 points: Most questions attempted, minimal effort

4 of 5 points: All questions attempted, complete effort, graded questions incorrect

4.5 of 5 points: All questions attempted, complete effort, graded questions partially correct

5 of 5 points: All questions attempted, graded questions perfect

**Project Presentation and Technical report [25%]** The major milestone in this class will be conducting (with a group) a statistical investigation on a question of interest to you. For each, you may collect primary data by hand or you can use data available on the Internet. You may also use data that we have used in class, as long as it is for a different project.

You will prepare a project proposal describing your study and obtain approval from me before you begin the investigation. During the last week of class, you (and your group) will give a 10-minute oral presentation of your study. We will spend time in class looking at what data is available on the web and about writing a project proposal. You will write up your report using RMarkdown and submit it before finals week.

**Exams [25% each]** There will be two take-home exams. These will involve write-ups of a statistical analysis, done using RMarkdown.

**Disability accommodations**: Any student with special needs requiring accommodations should give me his/her memo of accommodations in a timely manner. It is the student’s responsibility to follow up with me regarding all accommodations that require my participation. The student is advised to ensure full use of testing accommodations by coming to talk to me at least three days before any test.

**Attendance**: Your attendance in class is crucial, as is your punctuality. We are all going to learn this material together, so we need to have everyone present and working. Accommodations for an unavoidable absence can be made in advance via email; one necessary absence during the semester is not unusual; having more than two is uncommon.

**Collaboration:** Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete homework assignments, and prepare for exams. However, every word that you write must be your own. Copying and pasting sentences, paragraphs, or blocks of *R* code from another student is not acceptable and will receive no credit. No interaction with anyone but the instructor is allowed on any exams or quizzes. All students, staff and faculty are bound by the Skidmore College Honor Code.

**Additional policies:**

Cell phones: If your phone rings, I get to answer it! Just kidding (sort of)

Computers: Each Friday will feature a lab. If you have a laptop you can bring, please bring it to that class. Wednesday will consist of a more formal discussion and notes.

**Calendar (subject to change)**

|  |  |  |
| --- | --- | --- |
| Date | Topics | Assignments |
| 4-Sep | What is `statistics in sports’? |  |
| 6-Sep | Lab: Intro to R/RStudio/RMarkdown | HW 0 |
| 11-Sep | Player valuation in baseball |  |
| 13-Sep | Lab: Correlation, R-squared, scatter plots | HW 1 |
| 18-Sep | Multiple linear regression modeling |  |
| 20-Sep | Lab: Baseball analysis via Lehman database | HW 2 |
| 25-Sep | Linear regression & prediction |  |
| 27-Sep | Lab: Prediction for MLB batter/pitcher stats | HW 3 |
| 2-Oct | NFL - Logistic regression, kickers, expected points |  |
| 4-Oct | Lab: NFL kickers | HW 4 |
| 9-Oct | No class |  |
| 11-Oct | Football - game theory: 4th downs, field goals | HW 5 |
| 16-Oct | Basketball – possessions, shot difficulty |  |
| 18-Oct | Lab: NBA shot maps & hot-hand (?) | HW 6 |
| 23-Oct | Hockey stats: Stein’s estimator, shrinkage/shooting % | Exam 1 take home |
|  | No class – study day |  |
| 30-Oct | Hockey overview |  |
| 1-Nov | Lab: hockey expected goals | HW 7 |
| 6-Nov | Soccer & expected goals |  |
| 8-Nov | Lab: NWSL games | HW 8 |
| 13-Nov | Adjusted plus-minus & regression specification |  |
| 15-Nov | Ridge and other regression models (split lab) | HW 9 |
| 20-Nov | Lab: Ridge regression and adjusted plus/minus |  |
|  | Thanksgiving |  |
| 27-Nov | Team ratings and BTM |  |
| 29-Nov | Lab: Team ratings | HW 10 |
| 4-Dec | *Take home exam 2: basketball and hockey* |  |
| 6-Dec | Project work |  |
| 11-Dec | Presentation | Final Paper: 12/17 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |