

STAT 470 : Introduction to Sports Analytics: Survey of Modeling, Methodology, and Communication

Name:	E-mail:	Phone:
Josh Havstad	jhavstad@calpoly.edu	562-881-0435

General Information

Short Course Title: Sports Analytics

Course Description: Applying advanced statistical techniques, with an emphasis on using methodology commonly found in the sports industry, to answer questions that might arise as a sports statistician. Data collection and data wrangling. Spatial analysis, regression models with an emphasis on predictive modeling, general additive models, classification trees, and clustering methods. Model assessment and selection. How to communicate ideas to various levels of an organization (players, coaches, general managers, etc.). Heavy use of statistical software (R, Tableau, SQL). Creating reproducible documents (GitHub).

Prerequisites: STAT 302, STAT 324 or 334, and STAT 331.

Course Requirements

Course Requisites

Type	Course	Justification
Prerequisite	STAT 302	Students will need to understand simple linear regression, multiple regression, prediction, and basics of statistical modeling to be successful in this advanced course.
Prerequisite	STAT 324 or 334	Students will need experience with advanced regression concepts such as logistic and nonlinear regression, as well as knowledge of variable selection, to have a foundation in some of the advanced concepts of sports analytics.
Prerequisite	STAT 331	Students will need to be familiar with data wrangling and computing in R to answer the

		problems that will arise in this course.
--	--	--

Course Delivery and Resources

What is the primary modality in which the course is intended to be taught:

Face-to-Face, Traditional (FT)

Indicate other modalities in which the course is intended to be taught:

Please answer the following questions about direct instruction and out-of-class work for EACH of the modalities selected above (or its equivalent for online). For each modality listed above please include the hours estimate on a weekly basis. For example FT: x hours, FO: y hours, etc.

Hours of face-to-face or synchronous instruction (may include instruction through web-conferencing software such as Zoom, Skype, Microsoft Teams, or its equivalent):

4

Briefly describe planned methods of direct instruction face-to-face or synchronous (e.g., lecture, discussion, small group problem-solving, videos, demonstrations, etc.):

Lecture, in-class discussion, problem-solving, guest-speakers, and activities

Hours of direct instruction online (asynchronous):

0

Briefly describe planned methods of direct instruction online:

None

Hours of out-of-class work or its equivalent:

8

Briefly describe planned methods for engaging students in out-of-class work or its equivalent for online (e.g., assigned reading, homework problems, non-faculty mediated discussion board posts, individual/group projects, papers, service-learning, etc.):

Textbook readings, homework, group projects/presentations, final project/report

Enrollment capacity by modality:

35

Indicate the names of faculty members who will initially teach the course, and if one (or more) of the online modalities (FO, RO, LO, HY, & FL) are being proposed, please briefly provide their prior online experience and/or training:

Beth Chance, Immanuel Williams

Program Learning Objectives

Select a program to display program learning objectives supported by this course.

BS Statistics (QUARTER CURRICULUM)

PLO1: Have a good working knowledge of the most commonly used statistical methods, including statistical modeling and omnipresent role of variability, efficient design of studies and construction of effective sampling plans, exploratory data analysis, and formal inference process.

PLO2: Have background in probability, statistical theory, and mathematics, including especially calculus, linear algebra and symbolic and abstract thinking.

PLO3: Be able to synthesize and apply knowledge of common inferential methods, understanding the limitations of procedures and appropriate conclusions.

PLO4: Communicate effectively (written and oral) with skills in collaboration (within and between disciplines) and teamwork, and in organizing and managing projects.

PLO5: Have a good mastery of several standard statistical software packages and facility with data management strategies.

PLO6: Have a focused concentration in an area of application outside the discipline of statistics.

Course Learning Objectives and Assessment Methods

	Course Learning Objective	Modality Assessment Method(s)	Program Learning Objective(s)
1	Apply data cleaning and data wrangling techniques to transform sports data into a usable format suitable for	Homework assignments, exams, projects	PLO5

	analysis and visualization.		
2	Explain how sports analytics is utilized across various areas of a sports organization and its direct impact on decision-making processes.	Homework assignments, exams, and projects	PLO3, PLO4, PLO6
3	Utilize statistical software and programming tools such as R, SQL, and Tableau to create informative visualizations of sports data.	Homework assignments, exams, and projects	PLO1, PLO5, PLO6
4	Identify and apply appropriate statistical modeling and machine learning techniques to address questions that may arise in the sports industry	Homework assignments, exams, and projects	PLO1, PLO3, PLO5, PLO6
5	Translate technical concepts and statistical results into both verbal and written recommendations or conclusions that could be presented to diverse audiences within a sports organization	Homework assignments, exams, and projects	PLO3, PLO4, PLO6

Expanded Course Content

List textbooks, materials, and/or other resources for the course.

Textbooks, materials, and/or other resources
<u>Mathletics</u> by Wayne L. Winston <u>Analytic Methods in Sports</u> by Thomas A. Severini <u>Scorecasting</u> by L. Jon Wertheim <u>Hands on Machine Learning with R</u> by Bradley Boehmke and Brandon Greenwell <u>Basketball on Paper</u> by Dean Oliver <u>Sports Analytics</u> by Benjamin Alamar

Football Analytics with Python and R by Eric Eager and Richard Erickson

Mastering Sports Analytics with R by Lauren Gallardo

Analyzing Baseball Data with R by Jim Albert

Provide a detailed outline of the content for this course:

Week	Readings	Topics of Discussion	Labs, Activities, Assignments
1		Overview of sports analytics, how it impacts and is used in various sectors of the industry, current trends, and popular terminology. Introduction to data scraping, data wrangling, data ethics, exploratory data analysis techniques with players' health data.	Homework, Quiz
2		Introduction to SQL and relational databases Case study: Using Lahman's baseball database and predicting team success with Pythagorean expectation	Homework, Activity
3		Introduction to Tableau for visualizing data. Creating a profile, table and graph creation, publishing a dashboard.	Quiz
4		Importance of communication, learning how to write a report, practicing and critiquing communication/presentation skills, watch sports analytics conference presentations	Homework
5		Logistic and multivariate regression for NFL field goal success. Introduction to in-game win probability models with football data.	Activity
6		Overview of sports betting, odds and probability, and evaluation of betting odds with Brier Scores.	Project
7		Ranking players and teams with popular sports metrics. Bradley-Terry, ELO, plus-minus.	Homework, Quiz

8		Intro to classification techniques (KNN, decision trees, support vector machines) with basketball data. Emphasis on interpretations and limitations.	Homework
9		Introduction to spatial data analysis and visualization. Case study: Analyzing shot location and spatial data in basketball.	Homework, Quiz
10		Overview of optimization and linear programming methods in sports. Case study: team building and lineup management.	Homework
11		None	Final Project Report, Final Exam

Final Assessment

What will be the method for final assessment for this course?

In-person comprehensive final exam, final project report

Will the final exam occur during the designated time period?

Yes