Syllabus

INTEGBIO 201/120 COMPBIO 210 – Spring 2022
Introduction to Quantitative Methods In Biology (4 units)

Lectures: 1:00-2:00 Etcheverry 3108

Co-instructor: Prof. Rasmus Nielsen
Email: rasmus_nielsen@berkeley.edu

Co-instructor: Prof. Peter Sudmant
Email: psudmant@berkeley.edu

Co-instructor: Prof. John Huelsenbeck
Email: johnh@berkeley.edu

Co-instructor: Prof. Denis Titov
Email: titov@berkeley.edu

Course description: This course provides a fast-paced introduction to a variety of quantitative methods used in biology and their mathematical underpinnings. While no topic will be covered in depth, the course will provide an overview of several different topics commonly encountered in modern biological research including a review of basic concepts of calculus, differential equations and systems of differential equations, a review of basic concepts in linear algebra, an introduction to probability theory, Markov chains, maximum likelihood and Bayesian estimation, measures of statistical confidence, hypothesis testing and model choice, permutation and simulation, and several topics in statistics and machine learning including regression analyses, clustering, and principal component analyses. The course includes a lab section focusing on building student skills in modern computational methods for biological data analysis using python and R.

Prerequisites: Graduate standing **or** Biology 1A, Biology 1B, a course in statistics such as Data 8, Stat 2 or Stat 20, and two semesters of college level math including calculus such as Math 10A and Math 10B. Undergraduate students engaged in honors research, or other supervised research, are preferred. Graduate students who have not previously taken any statistics courses may benefit from taking Data 8, Stat 2 or Stat 20, or similar courses, before enrolling. Previous knowledge of R is not necessary.

Course Format: The course consists of 3 hours of lectures (MWF 1:00 – 2:00 pm) and three hours of computer exercises.

Course readings : There is no required text. The course will be based on lecture notes developed by the instructors. Lecture notes for each week will be posted on bCourses.

Requirements and Grading: Submission of weekly lab report to GSI. Attendance at lectures and participation in classroom discussions and computer exercises are required of all students. There will be four in-class exams on 2/9, 3/7, 4/6, and 4/29. The computer labs will account for 30%, the four exams will each account for 10%, and a project due May 13 will count for 30% of the final grade. Graduate students have an option to let the labs count for 30% of the grade and the project count for 70% of the grade. Graduate students are also encouraged to use the project to apply methods to their own datasets.

Lecture Schedule

Dates	Topic	Instructor
1/19	Introduction	RN
1/21, 1/24, 1/26, 1/28, 1/31, 2/2, 2/4, 2/7	**Lecture 1.** Python Intro **Lecture 2.** Differential Equations Intro **Lecture 3.** Numerical solutions of ODEs using SciPy **Lecture 4.** Graphical methods of analyzing ODEs **Lecture 5.** Analytical solutions of ODEs using SymPy **Lecture 6.** Predator-Prey dynamics ODE models **Lecture 7.** COVID19 dynamics ODE models **Lecture 8.** Cell Signaling dynamics ODE models **Lab 1**: R & Python Basics **Lab 2**: Discrete Time Modeling **Lab 3**: Analytical Solutions to ODEs and Systems of Equations	DT
2/9	**Exam I**	DT
2/11, 2/14, 2/16, 2/18, 2/23, 2/25, 2/28, 3/2, 3/4	Introductory probability theory I (axioms of probability, conditional probability, Bayes' formula, discrete random variables, Binomial, Poisson). Introductory probability theory II (expectation, variance, continuous random variables, exponential distribution, normal distribution and its properties). Markov chains (discrete time Markov chains, transition probabilities, classification of states, stationary distribution, examples from evolutionary biology). **Lab 4**: Random number generation, drawing random variables from probability distributions, expected values, variance, cumulative probabilities, quantiles. **Lab 5**: Graphing in R **Lab 6**: Understanding Markov chains	JH

3/7	**Exam 2**	JH
3/9, 3/11, 3/14, 3/16, 3/18, 3/28, 3/30, 4/1, 4/4	Estimation, bias, method of moments estimation, maximum likelihood estimation, Bayesian estimation. Confidence and credible intervals, hypothesis testing, model choice. Regression analysis. Permutation, simulation, and bootstrap **Lab 7**: Simple statistics in R. More on data-frames and on accessing and extracting data in R. **Lab 8**: R exercise on constructing likelihood functions, optimizing the functions, and performing likelihood ratio tests. **Lab 9**: Linear regression in R using lm. Visualizations of data and residuals.	RN
4/6	**Exam 3**	RN
4/8, 4/11, 4/13, 4/15, 4/18, 4/20, 4/22, 4/25, 4/27	Measures of distance and similarity, hierarchical & k-means clustering, gaussian mixture modelling & expectation maximization, introduction to linear algebra, dimensionality reduction and PCA, hidden markov models **Lab 10**: K-means. Hierarchical Clustering. Visualization of clustering using dendrograms. **Lab 11**: PCA **Lab 12**: HMM implementation	PS
4/29	**Exam 4**	PS