QBIO 310 Statistical Thinking for Quantitative Biology

Units: 4
Spring 2025 Semester
Lecture: Mondays and Wednesdays, 2:00 – 3:20pm
Location: KAP 144
Discussion: Tuesdays 2:00 – 2:50pm (RRI 421),
or Tuesdays 3:00 – 3:50pm (RRI 421)

Instructor: Peter Calabrese
Office: RRI 404B
Office Hours:
Contact Info: petercal@usc.edu, 213-740-2434
For office hours, I will be in my office and I will also be on Zoom (https://usc.zoom.us/j/4898518195). It is up to you if you want to meet in person or online

Teaching Assistant:
Office Hours:
Location:
Contact Info:

Course Description

This is an upper-division course designed to introduce computational biologists to statistical theory for data analysis. Students will also learn basic programming skills in the statistical programming language R. The course is more mathematically demanding and more focused on general theory than BISC 305. At the same time, it is gentler and more targeted at biological data than courses that cover similar material in the math department, such as MATH 407 and 408. We will spend approximately 2/3 of the semester exploring simple linear regression, taking time to learn some statistical theory, to view linear regression from non-parametric/semi-parametric, likelihood-based, and Bayesian perspectives, and to implement methods in R. The remainder of the semester will be a tour of some important techniques useful for describing, visualizing, and modeling different types of data, including from studies with multiple independent variables or dichotomous outcomes.

Learning Objectives

By course's end, our aim is that you will be able to:

- Discuss the philosophy involved in typical statistical estimation and inference, in which models are posited as data-generating processes with unknown parameters.
- Read and understand mathematical descriptions of simple statistical models.
- Explain the assumptions involved in justifying various views of the least-squares line, including a minimal "exploratory data analysis" view and views arising from semiparametric, parametric, and Bayesian models.
- Understand probabilistic and statistical concepts including expectation, variance, covariance, correlation, the law of large numbers, the central limit theorem, bias, consistency, efficiency, confidence intervals, *p* values, power, bootstrapping, permutation tests, likelihood, prior distributions, and posterior distributions.
- Design legible and informative data displays.
- Learn new methods for data analysis, such as linear regression, ANOVA, generalized linear models including logistic regression, principal component analysis, and linear mixed models, identifying principled reasons for choosing analysis methods.
- Explore the properties of statistical procedures using simulation and probability calculations.
- Use R to analyze and plot data, as well as write code to implement basic versions of procedures like bootstrapping and permutation testing.

Prerequisite(s):

There are no specific requirements to enroll. The main requirement is that you have an interest in learning about using mathematics and computation to support scientific claims with data. Beyond that, comfort with algebra is very helpful. Some familiarity with the ways in which statistical analyses are used in research is helpful—if the words "mean," "median," "mode," "scatterplot," "standard deviation," "t-test," "confidence interval," are at least vaguely familiar, you are covered on this dimension. However, the way in which we approach these questions is quite different from in, for example, AP statistics. We will use some basic calculus, and we will be programming. Courses in these areas will likely help you feel comfortable initially but are not required.

If you have taken MATH 407 and 408 or equivalent courses, then the material in this class would be repetitive for you, and you are urged to take a different course.

Co-Requisite(s): none

Concurrent Enrollment: none

Recommended Preparation: none

Course Notes

In this course, we will take the time to learn one statistical method deeply first, and then we will add breadth at the end. This involves some mathematics and computer programming. Some of you may not have had math classes for a while and may have little experience programming. That will make the course a bit harder, but it is still possible to succeed with hard work and a good attitude. Lecture slides will be posted.

Technological Proficiency and Hardware/Software Required

We will use R, a programming language designed for statistical computing. R is available free online from the R Project website, https://www.r-project.org/. We will also use RStudio, an interactive development environment designed for use with R. RStudio is also free. (Download RStudio Desktop from https://www.rstudio.com/products/rstudio.) RStudio requires an active R installation.

Required Readings and Supplementary Materials

Statistical Thinking from Scratch: A Primer for Scientists, by M.D. Edge, Oxford University Press, 2019.

Assignments

There will be HW assignments most weeks (due Wednesdays before midnight). You will submit these assignments on Brightspace.

There will also be a term paper. In this term paper you will select a dataset, and use what you have learned during the semester to analyze this dataset. The term paper will be due near the end of the semester.

Examinations

There will be an in-person midterm exam and a take-home final exam.

Grading Breakdown

Assessment Tool (assignments)	% of Grade
Homework	30
Midterm Exam	30
Term Paper	10
Final Exam	30
TOTAL	100

Grading Timeline

Assignments and exams will be graded within one to two weeks of submission. Grades will be entered on Brightspace.

Additional Policies

You can work together on the HW assignments (every student must submit their own assignment). A 50% grade deduction will be imposed for late homework, and no homework later than one week will be accepted.

The professor reserves the right to make changes to the syllabus; these changes will be announced as early as possible so that students can adjust their schedules.

Course Schedule (Subject to change)

Week 1
Jan 13: Intro, course policies. Data quality and biases.
Reading: *STFS* Prelude and chapter 1.

Unit 1: Mathematical and computational tools for statistics

Jan 15: The statistical programming language R, part 1: R markdown and RStudio, interacting with R, exploratory data analysis.
Reading: *STFS* chapter 2.

Week 2
Jan 20: Martin Luther King, Jr. Day, no class
Jan 22: R, part 2 - Functions and Loops, data input/output
Reading: *STFS* Appendix B.

Week 3
Jan 27: Probability 1 (Foundations, Axioms, independence, conditional probability, Bayes' Theorem)
Reading: *STFS* chapter 4 (through the end of section 4.2 / Box 4-2, pp 38-48)
Jan 29: Probability 2. (Discrete and continuous random variables, pdfs and cdfs, distribution families)
Reading: *STFS* chapter 4 (sections 4.4-4.8, pp 48-58)

Week 4
Feb 3: Probability 3. (Expectation, Variance, and the law of large numbers)
Reading: *STFS* chapter 5 (through the end of section 5.2, pp 60-68)
Feb 5: Probability 4. (Correlation and covariance; The central limit theorem)
Reading: *STFS* chapter 5 (sections 5.3 and 5.5)

Week 5
Feb 10: Probability 5. (Poisson)
Feb 12: Probability 6. (Conditional distributions; Least squares)
Reading: *STFS* chapter 3 and chapter 5 5 (sections 5.4, 5.6-5.7)

Unit 2: Basic statistical theory

Week 6
Feb 17: President's Day, no class
Feb 19: Properties of Estimators: Bias, Variance, Mean Squared Error, and Consistency.
Reading: *STFS* interlude; chapter 6 through the end of section 6.4.

Week 7
Feb 24: Properties of Estimators: Efficiency and Robustness. Decision Theory.
Reading : *STFS* chapter 6, sections 6.5-6.10.
Feb 26: Standard errors and confidence intervals
Reading: *STFS* chapter 7 through the end of section 7.2.

Week 8
March 3: p values and hypothesis tests
Reading: *STFS* chapter 7, sections 7.3-7.4
March 5: Power and effect size, multiple testing. Criticisms of NHST
Reading: *STFS* chapter 7, sections 7.6-7.9 (skip optional section 7.5)

Week 9
March 10: Statistics in the News and Midterm review
March 12: **Midterm Exam in-class**

March 17 – 21: Spring Break, no class

Unit 3: Approaches to estimation and inference

Week 10
March 24: Plug-in estimators, the method of moments, and the bootstrap.
Reading: *STFS* chapter 8 through the end of section 8.2.
March 26: Permutation tests.
Reading: *STFS* chapter 8, sections 8.3-8.5.

Week 11
March 31: Maximum-likelihood estimation.
Reading: *STFS* chapter 9, through section 9.2; skip optional section 9.2.2.
April 2: Wald test, score test, and likelihood-ratio test.
Reading: *STFS* chapter 9, sections 9.3-9.5.

Unit 4: Models for data analysis

Week 12
April 7: Assessing linear regression assumptions, multiple linear regression
Reading: *STFS* Postlude, through the end of section Post.2.1
April 9: Special cases and relatives of linear regression 1 (t-test, ANOVA)
Reading: Course notes Brightspace

Week 13
April 14: Special cases and relatives of linear regression 2 (ANOVA, Chi-squared test)
Reading: Course notes posted on Brightspace
April 16: Generalized linear models
Reading: *STFS* sections Post.2.2-end

Week 14
April 21: The Bayesian Alternative: Priors and posteriors.
Reading: *STFS* chapter 10.
April 23: Bayesian continued

Week 15
April 28: Neural Networks
April 30: Review

Term paper due Friday, May 2
Take-home final exam due Monday, May 12

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) - (213) 740-5086 | Title IX – (213) 821-8298
equity.usc.edu, titleix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776
dsp.usc.edu
Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Campus Support and Intervention - (213) 821-4710
campussupport.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.