Graduate Group in Epidemiology and Biostatistics

BSTA 6100: Biostatistical Methods for Epidemiology

Fall 2025

1 Course Overview and Objectives

This one-semester course is designed to provide a strong foundation in biostatistical methods for epidemiologic research, intended for students entering a PhD program in epidemiology. Covered topics include introductory probability theory, estimands, large-sample theory, hypothesis testing, confidence intervals, linear regression, and generalized linear models, all with a throughline of likelihood-based inference. The course will consist of interactive lectures and labs designed to develop hands-on analytic skills. Upon completion of this course, students will be able to

1. Explain and implement key statistical concepts that underlie data analysis for epidemiologic research;
2. Understand principles of statistical estimation and inference, including asymptotic properties, hypothesis testing, and interval estimation;
3. Implement data analyses in R, perform appropriate diagnostic checks, and report and interpret results in a reproduceable format;
4. Translate acquired biostatistical knowledge to novel research contexts.

2 General Course Information

Instructor:	Nicholas J. Seewald, Ph.D. seewaldn@pennmedicine.upenn.edu 624 Blockley Hall
Teaching Assistant:	Hong Xiong hong.xiong@pennmedicine.upenn.edu
Class Times:	Tuesday, 10:15 AM - 11:45 AM Thursday, 10:15 AM - 11:45 AM
Class Location:	505 Blockley Hall
Credits:	1.0 course unit
Prerequisites:	Single variable calculus, Prior coursework in statistics at the undergraduate level

Advisory Prerequisites:	Linear algebra at the undergraduate level
Corequisites:	EPID 7010: Introduction to Epidemiologic Research
Required Materials:	Access to a computer capable of running R, RStudio, and Quarto (running, e.g., Windows 10/11, macOS 12+), Course materials such as lecture slides will be provided via Canvas.
Instructor Office Hours:	TBD and by appointment, 624 Blockley Hall
TA Office Hours:	TBD.

3 Course Structure

Lectures

Lectures will focus on introducing concepts with worked, interactive examples. Students will be expected to actively participate in the interactive components of lecture.

Labs

Labs will involve a brief (20-25 minute) interactive tutorial on using R to accomplish relevant data management and analytic tasks, followed by a structured mini-project to be accomplished in pairs or small groups. Lab material will be linked to that week's lecture topic.

4 Assessments and Grading

The grade for the course will be based on the following components:

- **Labs and Problem Sets (60%).** There will be approximately 7 lab assignments and 5 problem sets, each weighted equally. The lowest *two* scores (including zeros) will be automatically dropped. Labs will focus on developing students' ability to analyze data and synthesize results into scientific reports. Problem sets will focus on developing students' theoretical understanding of course material, including mathematical facility.
- **Exam 1 (15%): October 16, 2025.** The first exam will cover material from the first half of the course and consist of questions assessing knowledge of theoretical and applied aspects of biostatistics.
- **Exam 2 (15%): Date TBD.** The second exam will cover material from the entire semester and consist of questions assessing knowledge of both theoretical and applied aspects of the biostatistical methods discussed in the course. *The exam is a cumulative exam and will cover material from prior to exam 1.*

Both problem sets and labs are due one (1) week after they are assigned, before the start of class that day, unless otherwise noted. Labs must be submitted electronically on Canvas as .html files rendered using Quarto. Problem sets may be submitted electronically or on paper.

Both exams will be in-person and handwritten. Use of computers is not allowed, though students will be permitted to use one (1) 8.5"x11" double-sided cheat sheet that is handwritten and must be turned in with the exam.

Letter Grades

Letter grades for the course will be curved to a distribution based on weighted final scores.

5 Course Policies

5.1 Academic Integrity

Intellectual development requires honesty, responsibility, and doing your own work. Taking ideas or words from others – plagiarism – is dishonest and will result in a failing grade on the assignment and possibly other disciplinary actions, including reporting to the University of Pennsylvania's Center for Community Standards and Accountability. If you have a question about what constitutes academic misconduct, please contact the course director(s) and/or epidemiology PhD program leadership. You should also consult *Academic Integrity at the University of Pennsylvania: A Guide for Students*, which can be found here: https://catalog.upenn.edu/pennbook/code-of-academic-integrity/.

It is the opinion of the instructor that the use of generative "artificial intelligence" and large language models (LLMs) is, in the vast majority of cases, an impediment to learning. Using these models deprives you, as a learner, of the opportunity to grapple with difficult problems and reason your way through them. Material generated by these models may also be inaccurate, incomplete, or otherwise incorrect. As such, **use of generative artificial intelligence tools (e.g., ChatGPT, Claude, Gemini, etc.) in this course is not allowed.** While neither the TA nor the instructor will actively police the use of these tools, the course is designed such that reliance on them will almost certainly result in a lower grade than had you learned the material without them.

5.2 Addressing Issues Related to Disabilities

If you think you need an accommodation for a disability (including and not limited to mental health conditions), let the instructor know as soon as possible. Some aspects of the course can be modified to facilitate your inclusion and progress. We will work with the Weingarten Center to determine appropriate academic accommodations.

Several studies suggest that graduate students are at greater risk for mental health issues than those in the general population. This course is meant to be challenging, but it is not my intent to contribute to experiences of stress, depression, or anxiety. Please reach out to me if you find yourself feeling low. I also encourage you to reach out to the different mental health services at Penn, including CAPS and GAPSA. I will make every effort to maintain the confidentiality of the information you share with me, but please be aware that in cases of disclosing experiences of sexual assault, self-injury or intent to injure others, I may be required by the University to report any concerns.

5.3 Collaborative Work

You are welcome to collaborate on labs and problem sets with fellow students; however, all work turned in must be your own. Any fellow students with whom you collaborated must be identified on the submitted assignment.

5.4 Late Work

Submitting work late creates difficulties for both the student and the instructor. For the instructor, consistent grading across student assignments becomes difficult if assignments are not turned in and graded together. For students, learning to keep to deadlines or to ask for extensions in advance is an important aspect of professional development. Because the lowest lab and problem set scores are dropped, **no late work will accepted** *unless* the student opts in advance to use one or more "late days" for the assignment.

Students are allowed two (2) late days to be allocated across the semester's assignments as they wish. A late day is one (1) 24-hour grace period (including weekends) for late submission of an assignment. A student may choose to use both late days for a single assignment or may allocate single late days to two different assignments. Due dates change according to the below table:

Original due date	Due date with 1 late day	Due date with 2 late days
Tuesday	Wednesday	Thursday
Thursday	Friday	Saturday

Late days are not applied automatically. Use of one or more late days **must** be declared to the instructor (and received in writing via email) in advance of the assignment's due date and time. Receipt of such a declaration after the due date and time will result in the late day not being applied and the student's work will not be accepted late. There is neither a penalty nor a reward for unused late days.

5.5 Regrade Requests

Errors in grading are infrequent, but do happen. If you notice an error in grading on a problem set or lab, you have **one week** from the date it was returned to you to request a regrade. Requests must be made in writing and include the reason you believe your answer was correct. No regrade requests will be honored after this one week period. Regrades for objective (e.g., multiple choice) questions will not be considered.

You may ask for a regrade if

- Your answer matches the answer in the solutions, but the grader did not realize it. Your explanation should make it clear why you believe your answer is the same.
- Your answer is different from the solutions, but your answer is also correct. Your explanation should make it clear that you have read through the solutions and should indicate why you think that your answer is equally good.

5.6 Coding Policy

Unannotated computer code or output will not by itself constitute an acceptable answer to any question on any assignment, unless otherwise noted or unless the submission is entirely self-

explanatory (which is extremely rare).

Lab assignments will be submitted as raw (uncompiled) .qmd files which must compile on the grader's computer, as well as . No attempts at troubleshooting will be made; it is the student's responsibility to ensure their work is reproducible from uncompiled code.

5.7 Good Course Citizenship

Students are expected to be good citizens of the course. This means coming to class prepared to engage with material, asking questions, and completing assignments to the best of your ability. Engage respectfully with the instructor and fellow students both in and out of class sessions, understanding that people learn in ways that may be different from you.

Because data is collected by and about humans, it necessarily encodes aspects of our proclivities and biases. As a result, this course may touch upon difficult topics related to race, gender, inequality, class, and oppression. We each come into this class with different perspectives that can be shared to enhance our understanding of these issues. Please enter these conversations with respect, curiosity, and cultural humility. You should be open to alternative perspectives and be willing to revise beliefs that are based on misinformation. Generally, your ideas and experiences can always be shared during these conversations but please refrain from dismissing the experiences of others. Personal attacks of any kind will not be tolerated. Plan to treat your fellow students and the instructor with respect. This includes things like arriving at class on time, coming in quietly if you are late (or leaving quietly if you need to leave early), and focusing on the task at hand, and using your fellow students' preferred names and pronouns.

5.8 Class Attendance & Engagement

The best way to succeed in this course is to attend class. Active engagement in the course is expected. This can be demonstrated in a number of ways, including consistent class attendance, participation in discussion forums, attending office hours, etc.

5.9 Email and Communications

All email communications **must** come from a upenn.edu email address. Please allow 2 business days for a response (i.e., not counting weekends). If you have not received a response within two business days, please send a gentle reminder. Nick typically replies to email between 8am and 6pm. If your email is lengthy or requires a lengthy response, consider attending office hours to discuss.

6 Office Hours

Students are invited to attend office hours to get help with understanding concepts from lecture, homework, and to prepare for the exam. The goal of office hours, whether the instructor's or the TA's, is to provide additional support in students' development as independent problem solvers. To this end, the instructor and TA will advise students that seek help on assignments, but will not directly provide answers. Similarly, the instructor and TA will advise students with code structure and help students develop strong debugging skills, but will not directly debug code.

Students must have first made a good-faith attempt at assignments before asking for help in office hours.

7 Scientific Rigor and Reproducibility

This course is designed to teach students how to select and apply appropriate statistical methods to address questions in epidemiology in a transparent and reproducible way through clear presentation of results, conclusions, and statistical programming.

8 Key Dates

The following is a tentative list of key dates for the course.

- September 18: Dr. Seewald traveling; asynchronous lecture.
- October 9: Fall break; no class.
- October 16: Exam 1
- November 27: Thanksgiving; no class.
- EXAM 2 TBD