Advanced Topics in Biostatistics: AI Tools for Data Science and Statistics

140.850, 3rd Term, 2025–2026

2026-02-18

Instructor

Erik Westlund, PhD
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
ewestlund@jhu.edu

Course Description

As AI tools are rapidly adopted across research and industry, there is a growing need for statisticians and data scientists to understand what these tools are capable of and how to use them responsibly. This course provides practical approaches for integrating large language models and agent-based tools into statistical workflows. Students learn how to structure AI-assisted processes for analysis, simulation, and pipeline development, along with core skills in context management and agent orchestration. The course emphasizes AI safety, privacy, and responsible handling of sensitive data. Students will discuss these topics and do hands-on exercises to test the strengths and weaknesses of AI tools in practice.

Course Details

Dates	February 19, February 26, March 5, March 12, 2026
Time	9:00–10:20 AM ET
Location	Zoom (link provided on CoursePlus/Email)
Format	Lecture, live demos, discussion, hands-on exercises
Grading	Pass/Fail

Course Learning Objectives

By the end of this course, students will be able to:

- Describe how large language models work at a conceptual level, including their capabilities and limitations for statistical work.
- Evaluate the ethical implications of AI tool use in research, including privacy, bias, reproducibility, and academic integrity.
- Use code to work with data, not the LLM itself, to protect sensitive and regulated data (including PHI) in privacy-sensitive contexts.
- Navigate the landscape of AI tools — chat interfaces, IDE integrations, CLI agents, and supporting tools — and select appropriate tools for different tasks.
- Use AI assistants to write, debug, and audit code for data cleaning, visualization, and statistical analysis.

- Build analysis workflows on synthetic data and deploy validated code to secure data environments using Git.
- Critically assess AI-generated output and identify errors, hallucinations, and inappropriate statistical choices.

Schedule

Session 1 (February 19): Foundations: History, Today, and Ethics

We will share our experiences using AI and study the foundations of AI tools. We will discuss ethics and concerns people have with using the tools.

Session 2 (February 26): The Toolbox

We will introduce and demonstrate various tools available to us, including major models (Claude, GPT, Gemini), open source models, chat tools, IDEs (VS Code, Positron, Cursor), CLI tools (Claude Code, Codex, OpenCode), and supporting utilities (tmux, Git).

- **Due before class:** Email course project topic to ewestlund@jhu.edu

Session 3 (March 5): AI-Assisted Statistical Workflows

We will discuss how to use these tools to summarize datasets, generate synthetic/simulated data, and build models. We will pay special attention to how to work with PHI using these tools.

- **Due:** Make progress on course project; prepare for check-in

Session 4 (March 12): Synthesis and Looking Forward

We will spend most of this session sharing our experiences with our projects, but will leave it open to address lingering questions.

- **Due:** Course project materials and 2–3 page reflection; in-class share

Course Project

Each student will attempt an ambitious task using AI tools — something you know little about, or that feels beyond your current capability. The goal is not to produce publication-quality work. It is to explore the limits and possibilities of these tools in a low-stakes way and to report honestly on what happened. See the course project page for full details.

Date	Milestone
February 19 (Session 1)	Assignment introduced; begin brainstorming
February 26 (Session 2)	Topic proposal due via email
March 5 (Session 3)	Progress check-in during class
March 12 (Session 4)	Project and reflection due; in-class share

Tools and Subscriptions

Students should subscribe to at least one AI tool for the duration of the course. Options include:

Tool	Cost
OpenAI ChatGPT Plus	$20/month
Anthropic Claude Pro	$20/month
Google Gemini Advanced	$20/month (one month free trial)

Tool	Cost
Cursor Pro	$20/month
GitHub Copilot Pro	$10/month

Variety across the class is encouraged — we will compare how different tools handle the same problems. If financing is a concern, please reach out to the instructor. Google Gemini Advanced offers a free trial that can cover the course period.

Students should also ensure that Git is installed and a GitHub account is registered before the first class. Mac and Linux typically have Git pre-installed. Otherwise, follow directions at git-scm.com.

Course Materials

There is no required textbook. Course readings and further reading are listed on the course website.

Methods of Assessment

This is a pass/fail course. Assessment is based on:

- **Participation (50%):** Attend all sessions and engage actively in discussions and hands-on exercises.
- **Course project (50%):** Attempt your project and present your experience to the class in Session 4.

Generative AI Policy

Using AI tools is the subject of this course. Their use is permitted, encouraged, and expected. It is nevertheless the student's responsibility to understand the output of these tools and ensure their correctness. Students are strongly encouraged to approach these tools as learning aids and not crutches.

Academic Ethics

Students enrolled in the Bloomberg School of Public Health of The Johns Hopkins University assume an obligation to conduct themselves in a manner appropriate to the University's mission as an institution of higher education. Students should be familiar with the policies and procedures specified under Policy and Procedure Manual Student-01 (Academic Ethics) and the Student Conduct Code (Student-06), available at my.publichealth.jhu.edu.

Student Health and Well-being

If you are struggling with anxiety, stress, depression, or other mental health related concerns, please consider connecting with resources:

- Student support: bit.ly/bsphstudentsupport
- Mental Health Services: wellbeing.jhu.edu/MentalHealthServices
- Behavioral Health Crisis Support Team (24/7): 410-516-9355

Disability Accommodations

Student Disability Services (SDS) provides accessible and inclusive educational experiences for students with disabilities. To request accommodations:

1. Complete the SDS online application via AIM
2. Submit documentation using the provided link after application submission
3. Schedule a meeting with Audrey Ndaba

More information: Student Disability Services