

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

140.850, 4th Term, 2025–2026

2026-02-18

## Instructor

Erik Westlund, PhD  
Department of Biostatistics  
Johns Hopkins Bloomberg School of Public Health  
[ewestlund@jhu.edu](mailto: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

<b>Dates</b>	February 19, February 26, March 5, March 12, 2026
<b>Time</b>	9:00–10:20 AM ET
<b>Location</b>	Zoom (link provided on CoursePlus)
<b>Format</b>	Lecture, live demos, discussion, hands-on exercises
<b>Grading</b>	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: Ethics and Foundations — February 19

*What are we dealing with, and what are the stakes?*

Block	Min	Topic
Discussion	15	How are you using AI now? What feels OK, what feels uncomfortable?
Lecture	15	AI ethics and responsible use: privacy and PHI, bias, reproducibility, academic integrity, environmental cost — a framework for reasoning, not a list of rules
Case study	10	How this syllabus was built with AI — is that OK? Why or why not?
Lecture	15	What is an LLM? Tokens, context windows, temperature, stochasticity. Why the same prompt gives different answers. Model landscape: GPT, Claude, Gemini, DeepSeek
Hands-on	20	Everyone gets the same statistical task; use your own tool; compare results across tools — what worked, what hallucinated?
Wrap-up	5	Introduce moonshot assignment; tools and subscriptions overview

**Homework:** Brainstorm your moonshot task. Try a second AI tool you haven't used before for a small task and note the differences.

### Session 2: The Toolbox — February 26

*What's out there, and how do you use it?*

Block	Min	Topic
Discussion	10	Debrief: what happened when you tried a new tool? Surprises?
Lecture + demo	20	The AI tools: models and model selection, chat windows, autocomplete tools, IDE agents (VS Code, Cursor, RStudio, Positron), CLI agents (Claude Code, etc.), Model Context Protocol
Lecture + demo	15	The supporting tools: command line basics, tmux for persistent sessions, Git and GitHub, API keys vs. subscriptions, cost management

Block	Min	Topic
Hands-on	30	Get set up with an IDE or CLI agent. Work through a guided exercise: use an integrated tool (not a chat window) to complete a coding task.
Wrap-up	5	Moonshot check-in

**Homework:** Make progress on your moonshot. Ensure Git is set up and working.

### Session 3: AI-Assisted Statistical Workflows — March 5

*Doing real work: from prompting to pipelines*

Block	Min	Topic
Lecture + demo	15	Context and prompting: structuring prompts for statistical work, project context files, system prompts, good prompt vs. bad prompt
Lecture + demo	15	The code-not-data principle in practice. Synthetic data as a development strategy. Git as a bridge between AI-assisted and secure data environments
Lecture	10	AI for statistical thinking: data cleaning, EDA, visualization, modeling, DAGs and causal assumptions, simulation, model validation — where AI is a thought partner and where it confidently misleads
Hands-on	30	Build an analysis pipeline on synthetic data using AI tools: cleaning, summary statistics, visualization, a simple model. Audit the output — catch the mistakes.
Wrap-up	10	Moonshot progress reports. Prepare a 3-minute informal share for next week.

**Homework:** Finish your moonshot attempt. Prepare to share what you tried, what worked, and what didn't (3 minutes, informal, no slides required).

### Session 4: Synthesis and Looking Forward — March 12

*What did we learn, and where do we go from here?*

Block	Min	Topic
Lecture	10	Reproducibility and documentation: organizing AI-assisted projects, documenting workflows, project setup and progress tracking
Moonshot share	40	Lightning rounds: each student shares their moonshot (~3 min) — what they attempted, what worked, what failed, what surprised them. Class discussion after each.
Discussion	15	Looking forward: picking models for different tasks, staying current as tools change, where these tools are headed, finding your personal line
Wrap-up	15	Course retrospective: what was most useful? What do you wish we covered? Feedback for the pilot.

## Moonshot Project

Each student will attempt a “moonshot” 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.

- **Session 1:** Brainstorm your task
- **Session 3:** Work session and progress check-in
- **Session 4:** Informal 3-minute presentation to the class

## 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)
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](https://git-scm.com).

## Course Materials

There is no required textbook. Readings and resources will be shared on CoursePlus.

## Readings

- Gloaguen, T., Mündler, N., Müller, M., Raychev, V., & Vechev, M. (2025). [Evaluating AGENTS.md: Are Repository-Level Context Files Helpful for Coding Agents?](#)

## 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.
- **Moonshot project (50%):** Attempt your moonshot task 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](https://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](https://bit.ly/bsphstudentsupport)
- Mental Health Services: [wellbeing.jhu.edu/MentalHealthServices](https://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](#)