University of Pennsylvania
Division of Biostatistics
Subject Guide

BSTA 6700: Programming and Computation for Biomedical Data Science

Credit points: **1.0**
Semester: **Spring 2026**
Time: **M/W 1:45-3:15PM**
Location: **BRB 252**

Course Instructor:
Kristin A. Linn
Associate Professor of Biostatistics
Email: klinn@pennmedicine.upenn.edu
Office: Richards, D002 (basement level, D-tower)
Office hours: Mondays, 3:30-4:30pm
Location: Richards, D-tower basement conference room

TAs

Kosta Tsingas
Email: Konstantinos.Tsingas@pennmedicine.upenn.edu
Office hours: Tuesdays, 11:30am-12:30pm
Location: Zoom (link on Canvas)

Meghan Gerety
Email: Meghan.Gerety@pennmedicine.upenn.edu
Office hours: Thursdays, 12-1pm
Location: TBD

Pre-requisites: BSTA 620, 621, and 651; or permission of instructor.

Course Aims: The course will cover programming and computational fundamentals in Python and R. It will concentrate on computational tools that are useful for statistical research and computationally intensive analyses. The goal is for students to develop a knowledge base and skill set that includes a wide range of modern computational tools needed for statistical research and data science. Topics may include, but are not limited to:

1. Reproducible research and programming
2. Algorithms
3. Simulation
4. Computer storage and arithmetic
5. Numerical Integration
6. Optimization

Course Materials: All course materials will be available on Canvas. Canvas is assessable from the Penn library: https://canvas.upenn.edu

Software: A combination of R and Python will be used.

Textbook: None required.

Breaks: There will be **no class** on:
January 19 (MLK Jr. Day)
February 25 (DBEI Research Day)
March 9 and 11 (Spring Break)
March 16 and 18 (ENAR conference)
April 20 (UPenn's Annual Clinical Trials Conference)

Quizzes: Several quizzes will be given at the beginning or end of lecture throughout the semester, and quiz dates will not be announced in advance. I will provide 1 page of blank paper for each student which will be turned in with your name and work. No materials, laptops, phones, etc. may be consulted during the quizzes. Please bring a pencil or pen to every lecture. If a quiz is missed for an excused absence (e.g., PhD interview), your final quiz average will be determined by averaging your grades on the other quizzes. A quiz missed because of an unexcused absence will be given a 0 and will be included in the final quiz average.

Midterm Exam: The midterm exam will be held on **March 4, 2026** during lecture hours. No materials, laptops, phones, etc. may be consulted during the midterm. You will be responsible for bringing a pencil or pen.

Assessment: All assignment materials will be submitted on Canvas. Grades will be based on the following components:
In class quizzes: 10%
Problem sets: 40% (10% each)
Midterm exam: 20%
Final project: 30%

Students are encouraged to discuss strategies for solving problem sets, but all submitted code should reflect each student's unique implementation. **Evidence of shared code will be penalized.**

Late Policy: Late assignments will receive a maximum of half credit. An assignment submitted 1 minute after the deadline will be considered late. Assignments more than 3 days late will not be graded and will receive no credit. **If you have a pre-existing commitment or special**

circumstance (e.g., conference travel, family emergency) please let me know as far in advance as possible so that we can make alternative arrangements for submitting your work.

Final Project: Details about the final project requirements will be given later in the semester.

Use of Generative AI Tools I encourage you to use foundation models such as ChatGPT, GitHub Copilot, etc., in combination with critical thinking skills to further your educational development. If you use these models to obtain quick solutions, you may be missing out on learning opportunities and potentially stifling your own creativity. Keep in mind large language models may produce incorrect statements and fake citations, and code generation models may produce incorrect outputs. **If you use materials produced by foundation models in a problem set, you must acknowledge which models were used and for which parts of the problem.** It is also important to "show your work" to get full or partial credit, i.e., **document what prompts you used** to obtain your outputs.

Useful resources:

Git documentation and book by Chacon and Straub: https://git-scm.com/book/en/v2

Python documentation: https://docs.python.org/3/

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT press.

Wickham, H (2015). *Advanced R*. CRC Press.

Matloff, N (2011). *The Art of R Programming*. No Starch Press.

Monahan, J (2011). *Numerical Methods of Statistics* (second edition). Cambridge University Press.

Givens, G.H., & Hoeting, J.A. (2013) *Computational Statistics*. Second edition. Wiley.

Cheney, W, & Kincaid D. (2008) *Numerical Mathematics and Computing*. Sixth edition. Thomson.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Academic Integrity: Academic integrity is the intellectual currency of our community. Our department expects honorable behavior from everyone. While study groups and strong interactions between classmates are encouraged (including working on homework together), your work on the quizzes, midterm, and final project must be your own. Students are expected to be familiar with and comply with Penn's Code of Academic Integrity, which is available online. Any case of suspected cheating on any of the graded assessments will be directed to the Center for Community Standards & Accountability (CSA). Possible consequences are zero on the assignment, quiz or exam; F in the course; note on your transcript; suspension; or expulsion. If you have any doubts or questions about what constitutes academic misconduct, please do not hesitate to contact me.