**University of Pennsylvania**
**Division of Biostatistics**
**Subject Guide**

# BSTA 670: Programming and Computation for Biomedical Data Science

**Credit points:** **1.0**
**Semester:** **Spring 2022**
**Time:** **T/Th 8:30-9:50am EST**
**Location:** **Zoom (through Jan 21), 418 Blockley**
**Meeting ID:** 942 9556 0765
**Passcode:** 091530
**Zoom link:**
https://upenn.zoom.us/j/94295560765?pwd=dUxyK1lnRlFEaVdxcUJHUERPZTRSQT09&from=addon

**Course Instructor:** Kristin A. Linn
Assistant Professor of Biostatistics
Email: klinn@pennmedicine.upenn.edu
Office: 220 Blockley Hall
**Office hours: Tuesdays 4:30-5:30pm or by appointment**
**Location:** Zoom (through Jan 21), 220 Blockley
**Meeting ID:** 963 1731 6664
**Passcode:** 768833
**Zoom link:**
https://upenn.zoom.us/j/96317316664?pwd=a1UrSGxBTWZQZkQrcVdWYUhNNkNRUT09&from=addon

**TA** Danni Tu
Email: Danni.Tu@pennmedicine.upenn.edu
**Office hours: Wednesdays 3:30-4:30pm**
**Location:** Zoom (through Jan 21), TBD (840 Blockley requested)
**Meeting ID:** 964 4415 7535
**Passcode:** 388925
**Zoom link:**
https://upenn.zoom.us/j/96444157535?pwd=UUVvTU02TEUyVXhGdnhROHpyeVp3dz09&from=addon

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

**Subject 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. Optimization
6. Numerical Integration

**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: March 8 and 10 (Spring Break), March 29 (ENAR), and April 12 (Penn clinical trials conference).

**Assessment:** All assignment materials will be submitted on Canvas. Grades will be based on the following components:

Homework: 50% (5 @ 10% each)
Data analysis midterm and in-class presentation: 20%
Final project: 30%

**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.

**Midterm:** The midterm grade will comprise an analysis of a public data set in a Python notebook, an in-class presentation of the work, and blinded peer reviews of other students' presentations. More details about the midterm project will be provided in February. Class attendance is required March 1 and 3.

Code/notebook due: Feb 28, 2022 by 5:00pm EST
Student presentations: March 1 and 3, 2022, 8:30-9:50am EST
Peer reviews due: March 17, 2022 by 5:00pm EST

**Final Project:** Students will replicate and extend the results of a recently published Monte Carlo stimulation experiment. The final project will include an R package containing simulation code and a report written in .Rmd that fully reproduces the simulation experiment. Additional details about the final project requirements will be given later in the semester.

All project materials due: May 9, 2022 by 5:00pm EST

**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.