BSTA 551 Biostatistics in Practice (Fall 2025)

Course Director: Mary Putt, PhD, ScD (mputt@upenn.edu)

Location: 202 Anatomy Chemistry

Tuesday/Thursday: 10:15-11:45

Office hours (In person): 11:45-12:30 Tuesday or by appointment

Description: BSTA 511 focuses on rigorous scientific thinking in the practice of biostatistics, both in the application of existing methods to biomedical problems and development of new methodologies. Part 1 of the course covers topics fundamental to an understanding of the role of biostatistics in medical research including study design, ethics, reproducibility, exploratory data analysis and critiquing the literature. With these basic concepts in hand, Part 2 moves to a discussion of more advanced topics under study by the Biostatistics faculty, for example, research frontiers in statistical genomics, predictive modeling, high-dimensional and functional data analysis, big data, health policy and causal inference.

Objectives:

- Provide a foundation for key concepts used in the practice of biostatistics
- Introduce motivation and development of new methodologies in biostatistics
- Practice critical thinking of applications and methodologies
- Practice communication skills
- Practice thoughtful and constructive criticism of others' work

Evaluation:

Class Participation: (30%)

Consistent class participation is expected for each student. Participation points will be evenly distributed across all lectures and will be determined in advance of the class by the instructor for that day. Examples of activities might include attending class, completing a quiz, submitting an assignment, or participating in a CANVAS discussion board. Attendance may be collected. Students are expected to attend all lectures.

Project 1: (35% of the grade)

Students will be assigned to groups. Each group will consider a case study highlighting either a novel application of biostatistics or some controversy in the use of biostatistics for a particular application. Case studies will be provided by the instructor or students may choose a study of their own interest. The initial case study provided by the instructor will include one or more published papers. Students will read the papers to determine the

scientific question of interest and the study design, assess the success of the study/study publication for answering the question of interest and think about next steps. This process will often entail further reading of the peer-reviewed literature. Critiques might incorporate topics covered by instructors in the first month of the semester e.g. was the study design appropriate, did the figures/tables effectively and validly convey the study results, are there better ways of presenting the data, are there concerns about reproducibility or ethics. Optionally, students may describe a novel statistical method used in the analysis or design and describe the advantages/disadvantages of the method. Students will be evaluated on the quality of a classroom presentation, quality of slides and abstract describing their project and constructive peer-review of a fellow students' presentation.

Groups: Groups for the first project are pre-assigned based on random pairings through CANVAS. You may change your groups as long as everyone agrees to do so.

Project 2: (35% of the grade) Students will form their own groups or work on their own. This project involves a literature review of a novel statistical method of interest to the student and/or a data analysis comparing the novel and more traditional methods. Students are encouraged to seek guidance on possible topics from any of the instructors or indeed the Biostatistics faculty. Students will be evaluated on the basis of a classroom presentation and a short writeup of their findings. The presentation/writeup should include a description of the scientific question of interest, the motivation for the new method (why didn't the original method do the job effectively?), and a summary of advantages/disadvantages of the new method. Students will be evaluated on the quality of their classroom presentation, a brief writeup of their findings and a constructive peer-review of a fellow student's presentation.

Late Policy: Assignments/quizzes are generally due 1 week after the lecture. Individual instructors may have different expectations, for example some quizzes may be due before the lecture. There are no points given for a late submission. The class participation points will be calculated by dropping the lowest grade.

Student Accomodation for Disabilities: A student in need of accommodations must contact the Student Disabilities Services (SDS) through the Weingarten Learning Center. After a formal evaluation through SDS, reasonable accommodations will be determined.

Plagiarism – defined by the University's Code of Academic Conduct as the use of existing ideas, data, or language without proper attribution to the original source – is prohibited in all forms. No large language model (LLM)-driven will be accepted as a credited author in your work. All cited author attributions included in your answers must demonstrate

accountability for the work, and LLM tools cannot take such responsibility. As a result, students are not allowed to copy (in part or in whole) or cite (in part or in whole) any result from a query posed to a LLM application Use without prior approval from the course director or dissertation advisor is not allowed.

For BSTA 511, this policy applies to any written work using Artificial Intelligence**. If the instructor or guest lecturer has concerns about whether a LLM or other AI tool was used for a written assignment, the student will be asked to meet with the instructor and provide an oral discussion of their findings for the assignment.** The grade for the assignment will be the minimum of the grade for the written and oral discussions.

Month	Week	Date	Topic	Instructor
August				
	1	26 (T)	Syllabus: Introduction (Big Data Conference First Project Description)	Mary Putt, ScD
		28 (Th)	Responsible Use of Data (ENAR Webinar)	Virtual (Sarah Ratcliffe, PhD) Mary Putt ScD Preceptor University of Virginia
September				
	2	2(T)	Reproducible Research (Markdown, SAP's, Github)	Jesse Hsu, PhD.
		4(Th)	Study Design I (Clin Trials)	Rui Feng, PhD.
	3	9(T)	Study Design I (Observational Studies)	Mary Putt, ScD
		11(Th)	Case Studies	Mary Putt, ScD
	4	15(M)	Conference: Big Data/Causal Inference in Biomedical Research Penn Big Data Conference	Optional Registration ($50 per student includes food) *Attend*: Attend at least 1 session from 10:00-11:45 AM
		16(T)	Exploratory Data Analysis	Mary Putt, ScD
		18 (Th)	Big Data Conference Discussion	Hongzhe Lee, PhD
	5	23(T)	Project 1 Preparation	Mary Putt, ScD
		25(Th)	Survey Sampling	Dawei Xie, PhD
	6	30(T)	No Class (Makeup is Big Data Conference)	
October				
		2(Th)	Project 1 Presentations	Students/Mary Putt
	7	7(T)	Project 1 Presentations	Students/Mary Putt
		9(Th)	Fall Break	
	8	14(T)	Clinical Trials in Industry	Devan Mehrotra, PhD
		16(Th)	Smart Trials	Kristen Linn, PhD
	9	21 (T)	Data Integration	Jiayin Zheng, PhD
		23(Th)	Xie Lab: Research Topics	Sharon Xie, PhD
	10	28 (Th)	Causal Inference	Alisa Stephens, PhD

		30(Th)	Mitra Lab: Research Topics	Nandita Mitra, PhD
November				
	Week	**Date**	**Topic**	**Instructor**
	11	4(T)	Topics in Survival Analysis	Doug Schaubel, PhD
		6(Th)	AI-driven Smart Spatial Omics	Mingayo Li, PhD
	12	11(T)	Large Language Models	Qi Long, PhD
		13(Th)	Imaging	Taki Shinohara, PhD
	13	18(T)	Functional and/or Bayesian Data Analysis	Jeff Morris, PhD
		20 (Th)	Statistical Measures of Agreement	Jarcy Zee, PhD
	14	25(T)	Project 2 preparation	Students/MPutt
		27(Th)	Thanksgiving	
December				
	15	2 (T)	Student Project 2	Students/MPutt
		4 (Th)	Student Project 2	Students/MPutt