HARVARD T.H. CHAN | SCHOOL OF PUBLIC HEALTH

BST 227: Introduction to Statistical Genetics
Fall 2 2024
Monday and Wednesday 3:45 - 5:15 pm Kresge 200

Instructor Information

Faculty:
Sharon Lutz
Associate Professor, Department of Population Medicine, HPHCI/Harvard Medical School
Associate Professor, Department of Biostatistics, Harvard T.H. Chan School of Public Health
Email: smlutz@hsph.harvard.edu
Zoom for the office hour (link on canvas)
Office hours: TBD

Teaching Fellows:
Nick Su
Email: ssu@hsph.harvard.edu
Zoom for the office hour (link on canvas)
Office hours: Friday 1-2

Yuchen Zhao
Email: yuchenzhao@hsph.harvard.edu
Zoom for the office hour (link on canvas)
Office hours: Thursday 1-2

Laboratory Sections
Mondays 5:30 pm - 7:30 pm and Wednesdays 5:30 pm - 7:30 pm in FXB G10 (except 11/6 lab: FXB G03).

Lab 1A: Introduction to R
Lab 1B: Biology Refresher
Lab 2: Concepts in linkage/association
Lab 3: GWAS
Lab 4: PRS, MR, rare variants

Though the labs are optional, we encourage you to attend. Note the lab sections are not every week. Lab 1A is an introduction to R. Lab 1B is a biology refresher focused on DNA, transcription, translation, mitosis, and meiosis. Lab 2 on linkage and association corresponds to HW1. Lab 3 on GWAS corresponds to HW2. Lab 4 on rare variants corresponds to HW3. If you have never used plink before, we **highly recommend** that you attend Lab 3 since plink is needed for HW2.

Credits
2.5 credits

Course Competencies:
This course covers the following Program/Field of Study Competencies:
SM 80 CBQG-Computational Biology Track:
Analyze genomic, transcriptomic, and epigenomic data to explore health and disease
SM 80 CBQG-Statistical Genetics Track:
Examine the basics of techniques modern population genetics such as genome wide association study methodology

Course Description

This course will provide an introduction to statistical methods for genetic studies. The emphasis of this course is on understanding basic concepts and methods and how they are applied in the analysis of population genetic data. The course starts out with an overview of methods for identifying disease-associated variants, and then goes on to cover approaches for interpreting the role of these variants in disease. A background in genetics is not necessary. The basic concepts of human and molecular genetics relevant to the course will be covered in class and the biology refresher lab.

- **Pre-Requisites**
  BST210 (Concurrent Enrollment Allowed) or PHS2000A (Concurrent Enrollment Allowed)

Learning Objectives

Upon successful completion of this course, you should be able to:

- Understand the basic concepts and methods in the analysis of population genetic data
- Run basic Genome Wide Association Studies (GWAS)
- Understand methods for sequencing data, rare variants, and Epigenome-Wide Association Studies (EWAS)

Course Readings: There is no required text book. There are supplemental articles on the canvas website.

- Suggested Textbook:
  The Fundamentals of Modern Statistical Genetics, Nan Laird and Christoph Lange, Springer 2011.
- Additional Reference Books
  Statistics in Human Genetics by Pak Sham
  Statistical Methods for Genetic Epidemiology by Duncan Thomas

Course Structure

Canvas Course Website: *https://canvas.harvard.edu/courses/62208*

Grading, Progress and Assessment

The final grade for this course will be based on:

- Homework assignments (60%)
- Final project (40%)

Homework assignments (60%)

There will be 3 homework assignments due on Fridays (except HW 2) to be submitted via the online course dropbox in Canvas by midnight EST. Students should feel free to discuss approaches to solving the problems in working through homework problems, but each student must turn in their own solution, written entirely in their own words and not copied from another source. In addition, you should not share your final solutions with another student. Plagiarism is not allowed. Late homeworks will not be accepted unless permission was given in advance for the late assignment. Use of AI (such as ChatGPT or similar programs) is not allowed for the homeworks.

Final project (40%)

The class will be divided into teams which will complete a project, write a report, and present their results the last week of class. Plagiarism is not allowed. Use of AI (such as ChatGPT or similar programs) is not allowed for the final project.

Course Schedule & Assessment of Student Learning

- Please note, session topics, and activities may be subject to change during the course

Week	Date	Topic	Lecturer	Lab/Notes	Assignments
Week 1	10/28	Introduction and overview of genetic disease	Dr. Lutz	**Lab 1A:** Intro to R Monday 10/28 **Lab 1B:** Biology refresher Wednesday 10/30	
	10/30	Mendel's laws & Hardy Weinberg Equilibrium	Dr. Lutz		
Week 2	11/4	An overview of linkage and association	Nick Su & Yuchen Zhao	**Lab 2:** Concepts in linkage/association 11/4&6	
	11/6	Population substructure & family-based association tests (FBAT)	Dr. Georg Hahn		**HW 1:** Due Friday 11/8
Week 3	11/11	**Veteran's Day**	**No lecture**		
	11/13	Genome wide association studies (GWAS)	Dr. Lutz	**Lab 3:** GWAS 11/13 Wednesday	
Week 4	11/18	GWAS: Meta-analyses & multiple testing	Dr. Lutz	**Lab 3:** GWAS 11/18 Monday	
	11/20	Pleiotropy & Mendelian randomization (MR)	Dr. Lutz		**HW 2:** Due Friday 11/22
Week 5	11/25	Polygenic risk scores (PRS)	Dr. Lutz		
	11/27	**Thanksgiving Recess**	**No lecture**	No office hours 11/27-11/29	
Week 6	12/2	Analysis of rare variants and non-SNP variation	Dr. Lutz	**Lab 4**: PRS & Rare variants 12/2 & 12/4	
	12/4	Variant calling from high-throughput sequencing data	Dr. Lutz		**HW 3:** Due Friday 12/6
Week 7	12/9	Metabolomics	Dr. Lutz		
	12/11	Epigenome-wide Association Studies (EWAS)	Dr. Lutz		
Week 8	12/16	Special topics in statistical genetics	Dr. Kurt Christensen		
	12/18	Project presentations			**Project presentation:** Due 12/18