Syllabus

STAT 540: Statistical Methods for High Dimensional Biology

Contents

2022-2023 Winter Term 2 (January 9, 2023 - April 13, 2023)

STAT 540 is a 3 credit course with a mandatory computing seminar

Cross-listed as STAT 540, BIOF 540, GSAT 540

Land acknowledgement

We respectfully recognize that the University of British Columbia Vancouver campus is located on the traditional, ancestral, and unceded territory of the xʷməθkʷəy̓əm (Musqueam) people. Please take a moment to learn about the territory you are occupying by visiting this interactive indigenous land map.

Course-level learning objectives

- Perform exploratory data analysis and visualize genomics data

- Apply tailored statistical methods to answer questions using high dimensional biological data

- Make your work reproducible, reusable, and shareable

- Work with real data in a collaborative model

Selected topics

- Basics of molecular genetics/genomic and data collection assays (methods)

- Basic probability and math foundations

- Exploratory data analysis and data quality control

- Normalization, batch correction

- Causal inference and confounding effects

- Basic statistical inference ("one gene at a time") - linear models

- Large-scale inference ("genome-wide") - multiple testing

- Analysis of microarray, RNASeq, and epigenetics data

- Principal Component Analysis and clustering (unsupervised machine learning)

- Classification and cross validation (supervised machine learning)

- Gene set analysis and gene networks

- Genome-wide association analysis (GWAS)

- Single-cell genomics

Teaching Team

For more info on the Teaching team, including brief bios, see the 'People' pages on this website (linked below).

Instructors

- Keegan Korthauer, PhD (She/Her/Hers)
  Email: keegan@stat.ubc.ca
  Virtual office hours: By appointment

- Yongjin Park, PhD (He/Him/His)
  Email: ypp@stat.ubc.ca
  Virtual office hours:

Teaching Assistants

- Asfar Lathif (He/Him/His)
  Email: asfarlathifbt@gmail.com
  Virtual office hours:

- Ishika Luthra Email: ishika.luthra@ubc.ca
  Virtual office hours:

Schedule

Lectures (Sec 201)

- Time: Tues Thurs 9:00 - 10:30am

- Location: Frank Forward 317

- See Lectures for lecture materials and schedule

Seminars (Sec S2B)

- Time: Mon 12pm - 1pm

- Location: Neville Scarfe 206

- See Seminars for schedule and seminar materials

Course communication

ANNOUNCEMENTS

Course announcements will be posted in the course Slack workspace. You are responsible for checking it regularly.

GENERAL QUESTIONS

Please also use the course Slack workspace for posting questions (e.g. topics discussed in class, questions about course organization, assignment clarifications, if you are stuck on an assignment and need help). This ensures the message can be seen by the entire teaching team, and that others in the class who might have the same question can learn from responses. You are also welcome to share your input on questions posted by others.

PRIVATE MATTERS

For private matters (e.g. requesting an extension, scheduling appointment for office hours), please contact the Teaching team by email (listed above). Please refrain from using email to ask general questions described above.

GROUP WORK

In your final project groups, we expect you to (1) arrange regular meetings either in person or virtually and (2) make use of your group's private Slack channel. Please reach out if you have any questions or

challenges in this space.

Prerequisites and Resources

This interdisciplinary course requires a broad range of skills at the interface of statistics, molecular biology / genomics, and computing. While you may have some background in one or more of the following areas, you are not expected to be an expert in all. That said, to be successful in the course, you may need to spend a bit more time in the areas that you have less experience in. Although there are no official prerequisites for the course, here is a list of useful skills to bring into the course and/or learn along the way.

Statistics:

- You should have already taken a university level introductory statistics course.

- This free online book "Modern Statistics for Modern Biology" by Susan Holmes and Wolfgang Huber is a great reference for introduction or review of many of the statistical concepts that are relevant for this course.

- This free online book "Data Analysis for the Life Sciences" by Rafael Irizarry and Michael Love is another great resource for introduction or review of many of the statistical concepts relevant in this course, with an emphasis on implementation in R.

Biology:

- No requirements, but you are expected to learn things like, e.g. the difference between DNA and RNA, and the difference between a gene and a genome.

- See this video and this slideset for some basic introductory material.

- Go through the (optional, not for a grade) molecular biology quiz on Canvas

- This free online book "Concepts of Biology" by Fowler, Roush & Wise is a great resource for biological concepts, in particular chapters 6 and 9

- This free online book "Biology" by Clark, Douglas & Choi goes more in-depth, see Chapters 14, 15, and 16 for material on genetics that is particularly relevant for this course.

- No matter your prior experience, when you come across a new biological concept during the course or in your research, you might need to spend a bit of time 'learning as you go' - this is expected and something I still do regularly in my day-to-day research!

R:

- No experience required but be prepared to do a lot of self-guided learning if you haven't taken other courses on R or used it in your research.

- Start now by installing R and the HIGHLY RECOMMENDED "integrated development environment" (IDE) RStudio - both are free and open source.

- You should be able to run R on your own computer during each seminar session.

- If you are new to R, check out this blog post on getting started with R

- If you are new to R, check out this blog post on getting started with R.

- This free online book "Introduction to Data Science" by Rafael Irizarry is also a great resource for getting more in-depth with R, programming basics, and the tidyverse. In particular see Chapters 1-5:
  - Chapter 1: Getting Started with R and R Studio

  - Chapter 2: R Basics

  - Chapter 3: Programming Basics

  - Chapter 4: The tidyverse

  - Chapter 5: Importing data

Git/GitHub and R Markdown:

- In this course we'll be using the version control software Git and its web-based hosting and collaborative platform GitHub.

- The online resource "Happy Git and GitHub for the useR" from Jenny Bryan is a great reference for these tools as we learn them.

- Another helpful git resource is Hadley Wickham's webinar "Collaboration and time travel- version control with git, github and RStudio"

- We'll learn about using R markdown to generate readable and reproducible reports with code and text, and you'll be using that a lot in this course - see Chapter 18 of the 'Happy Git' resource: "Test drive R markdown".

Evaluation

You will have three individual assignments, six seminar submissions (one divided into two parts), and one group project. Deadlines are all by 11:59 pm (Pacific time) on the due date. Any submission or modification after the due date will not be graded unless you have requested an extension. If you anticipate having trouble meeting a deadline and need to request an extension/academic concession please reach out via email.

For more detail on each of these assignments, see the assignments page (the header of each assignment on this page points to the relevant section of the assignments page). Also refer to this visual overview of the timeline.

For detailed instructions on how to work and submit assignments using GitHub and Canvas for this course, please see the Submission Guide.

Intro Assignment (5%)

- An introductory assignment designed to assess basic knowledge of GitHub, R and Rmarkdown

Assignment	Due Date
Intro Assignment	Tue 24 January 2023

Seminar completion (10%)

Seminar completion (10%)

- You will submit short "deliverables" for seminars 1, 2 (split into two parts - Seminar 2a and 2b together count as one deliverable), 3, 4, 5, and 7

- Each of these six Seminar session deliverables is weighted equally, but the lowest score will be dropped (so that the 5 deliverables with highest score will each count for 2% of the final grade).

- These deliverables give practical experience applying the knowledge that will be helpful on the homework assignment, final project, and (hopefully) your future research.

- Each deliverable is due on the Friday following the TA-led session for that seminar

Assignment	Due Date
Seminar 1	Fri 13 January 2023
Seminar 2a & 2b	Fri 20 January 2023
Seminar 3	Fri 27 January 2023
Seminar 4	Fri 03 February 2023
Seminar 5	Fri 10 February 2023
Seminar 7	Fri 03 March 2023

Paper critique (5%)

- Read, summarize and critique this paper

- For detailed instructions, refer to the rubric

Assignment	Due Date
Paper Critique	Tue 07 February 2023

Analysis assignment (30%)

- Involves detailed analysis of real data using R

- This assignment will assess your ability to understand and apply the methods learned in class

Assignment	Due Date
Analysis Assignment	Tue 28 February 2023

Group project (50%)

- A semester-long data analysis group project where you will apply the techniques covered in class to a research question of your choosing

- Groups of target size of 4 students will be formed at the beginning of the course

- Important checkpoints during the term (with deliverables):
  - Project Proposal Lightning Talks

  - Written Project Proposal

- Progress Report

- Oral presentation

- Written Report & GitHub repository

- Individual and Group Evaluation

- For more details on the project components and how groups are selected see the [assignments page](#)

- For detailed grading rubrics of the final project components, see the [final project rubric page](#) and the [final project presentation rubric](#)

Assignment	Due Date
Proposal Lightning Talks	Mon 30 January 2023
Written Proposal	Thu 09 February 2023
Progress Report	Tue 14 March 2023
Final Report	Thu 06 April 2023
Presentation Day 1	Thu 06 April 2023
Presentation Day 2	Tue 11 April 2023
Individual & Group Evaluation	Thu 13 April 2023
Presentation Day 3	Thu 13 April 2023

Academic Concession

If you anticipate having trouble meeting a deadline and need an [academic concession](#), please reach out in advance via email to the instructors. [Here](#) is a template you can use for a self-declaration.

If you miss class, we suggest you to:

- Consult the class resources on the course website

- Use the class Slack workspace to discuss missed material

- Come to virtual office hours

- Seek academic concessions, if applicable

Academic Integrity

Not only is academic integrity is essential to the successful functioning of our course, but adopting best practices will benefit you in your research practice. Make sure you understand UBC's definitions of academic misconduct and its consequences. Policy guidelines can be found [here](#).

What does academic integrity look like in this course?

- **Do your own work**. All individual work that you submit should be completed by you and submitted by you. Do not receive or share completed coursework with students who take the course in another term.

- **Acknowledge others' ideas**. Scholars build on the work of others, and give credit accordingly. This refers to both outside sources (e.g. from the literature), and inside sources (e.g from your peers).

- **Learn to avoid unintentional plagiarism.** Visit the Learning Commons' guide to academic integrity to help you organize your writing as well as understand how to prevent unintentional plagiarism.

At any time: if you are unsure if a certain type of assistance is authorized, please ask us.

Privacy

This course requires students to work and communicate on github.com and Slack. Please be advised that the material and information you put on GitHub and Slack will be stored on servers outside of Canada. Data used for these tools may not be protected, as they have not gone through a Privacy Impact Assessment (PIA) and been identified as FIPPA compliant. When you access GitHub or Slack, you will be required to create an account. While both tools have a privacy policy (GitHub privacy policy, Slack privacy policy), UBC cannot guarantee security of your private details on servers outside of Canada. Please exercise caution whenever providing personal information. You may wish to use a pseudonym to protect your privacy if you have concerns. If you choose to do so, you will need to inform the instructional team of your alias. Please feel free to contact UBC (access.and.privacy@ubc.ca) or the support team for GitHub or Slack if you have any questions.

Learning Environment & Support

We strive to provide a learning environment where all students can succeed. Please join me in contributing to a classroom culture where everyone feels valued. If you encounter an issue that presents a barrier to your learning, please reach out to us. You can also contact the Ombudsperson for help: https://ombudsoffice.ubc.ca. If you have a documented disability that affects your learning, you may contact the UBC Centre for Acessibility: https://students.ubc.ca/about-student-services/centre-for-accessibility, and contact us as soon as possible if you think you may require accommodation options for course work.

Your mental health and wellbeing impacts your academic performance. Sometimes it is possible to manage challenges on your own, while other times you may need support. UBC is committed to providing student mental health and wellbeing resources, strategies, and services that help you achieve your goals. Visit https://students.ubc.ca/health for more information.

Safety in the Classroom

Please follow the current UBC COVID-19 Campus Rules regarding self-monitoring, and staying home if you are sick. Although masks are no longer required on campus, please respect the choices of your fellow students and the instructional team who may continue to wear masks.