Department of Statistics, University of Manitoba

STAT 7250-T03 – Statistical Methods for Biological Data Analysis

Course Outline, Winter Term 2022

Course Details

Course Schedule: Tuesdays and Thursdays, 11:30 to 12:45

Location: Zoom - connect through UMLearn

Instructor Contact Information

Instructor: Aleeza Gerstein

Office Location: 364 Machray Hall

Lab Location: 412 Buller Buildilng

Phone: (204) 272-1681

Email: Aleeza.Gerstein@umanitoba.ca

Office Hours: By appointment only.

Textbook: Modern Statistics for Modern Biology
Full book: Free online
Order: Cambridge or Indigo or Amazon

Data: All data required to run the code in the book is found here.

General Goals and Course Description

This course is focused on applying practical statistics to cutting-edge, heterogeneous biological data. This is a hands-on course, where you will be responsible for directing a lot of your learning. It is not the type of course where a professor will stand at the "front of the room" for 2.5 hours each week and transmit information that you are expected to prove you know on a test. Many sessions are not formal lectures. Different parts of the material are expected to be more or less difficult for different members of the class. Students are expected to thoroughly read and work through assigned material prior to attending class. Course time will often be used to work through the written material and code as a group. Hence a major component of the class is participation and preparation: coming to class with a list of questions or comments on the material and participating in working through problems that arise. In addition to learning modern biostatistical techniques (and hopefully a considerable amount about the type of data that arises in modern biology experiments!), an equal goal of the course is to provide students with an opportunity to practice oral and written communication of statistical concepts and biostatistical data and results. Assessments are designed to meet the course aims, and are centered more around communicating knowledge than proving you have learnt how to code a specific test or solve a theorum. What you will get out of this course will be directly tied to the effort you put in. The goal is that you will walk away with research and communication skills that exceed the specific material that we will cover.

Course Objectives

1. To understand central concepts in modern biostatistics
2. To use the R Programming language to apply these statistics to biological data
3. To build a research toolkit:
   apply techniques for reproducible research
   learn best practices for data visualization
   develop discussion moderation techniques
   practice oral and written science communication

Course Assessment

Final Mark: The final mark for the course will be obtained from the following:

Preparation for and participation in weekly chapter discussions (3% per class, worst two dropped)	30%
Weekly assignments (6 assignments worth 3% each, worst mark dropped)	15%
Discussion moderation*	10%
Data visualization projects (oral)**	5%
Data visualization projects (written)	10%
Points of significance final project (oral)***	15%
Points of significance project final (written)	15%

* If you are ill when you are scheduled to moderate a class the marks will be split between the oral and written final project grades

** If small oral assignments are missed due to illness the marks will be added to the written component

*** If you are ill when you are due to present your final oral project, Aleeza will meet with you separately for the presentation

Outline of Course

Week 1 (Jan 25, 27)	Introduction to course & reproducible research Methods: .Rproj files, the 'here' package, markdown, github, bioconductor
Week 2 (Jan 31, Feb 3)	Generative models for discrete data (Chapter 1, Aleeza moderated) Biology: number of mutations, epitope detection Methods: computing probabilities, random numbers **Due: February 4, Weekly Assignment 1**
Week 3 (Feb 8, Feb 10)	Introduction to genetics
Week 4 (Feb 15, Feb 17)	Statistical modelling (Chapter 2, Student moderated) Biology: DNA base pair counts, Hardy-Weinberg equilibrium, haplotype frequencies Methods: Maximum likelihood, Bayesian estimation **Due: February 18, Weekly Assignment 2**

Week 5 (Mar 1, Mar 3)	High quality graphics (Chapter 3, Aleeza moderated) Biology: gene expression Methods: ggplot **Due: March 4, Weekly Assignment 3**
Week 7 (Mar 8, Mar 10)	Mixture models (Chapter 4, Student moderated) Biology: ChiP-Seq data, next generation sequencing read counts Methods: generate mixture model data, EM algorithm, bootstrap **Due: March 11, Weekly Assignment 4**
Week 6 (Mar 15, Mar 17))	High quality graphics presentations *March 15 will be a prep class for graphics assignment* **Due: March 17, Graphics assignment (oral)** **Due: March 20, Graphics assignment (written)**
Week 8 (Mar 22, Mar 24)	Clustering (Chapter 5, Student moderated) Biology: single cell RNA-seq, flow cytometry, cell clustering, 16S metagenomics Methods: similarity measurements, clustering methods, similarity measurements **Due: Mar 25, Weekly Assignment 5**
Week 9 (Mar 29, Mar 31)	Tidy data demonstration Biology: Gene Expression data Methods: dplyr, broom **Due: April 1, Weekly Assignment 6**
Week 10 (April 5, April 7)	Testing (Chapter 6) Methods: hypothesis testing, types of error, permutation tests, multiple testing
Week 11 & 12 (April 12, 14, 19, 21)	Final project presentations **Due: Final project oral presentations** **Due: April 24, Final project written document**

University Dates	Information
January 25	First lecture
February 4	Last day to drop without penalty
February 22-25	Winter term break
April 25	VW day for Winter courses

Information about assessments

Preparation for and participation in weekly discussions (30%):
Students are expected to come to class prepared to discuss subject material (and code) from assigned textbook chapters in weeks 2, 4, 5, 7, 9, and 11. In these weeks we will closely follow the textbook, which was written to be essentially the course notes for a similar course at Stanford. Half of the grade comes from being prepared for lectures, which likely means you have an .RMD file on your course GitHub site that shows how you have worked through the textbook material, including coding the Question exercises from the chapter. The other half of the participation grade comes from whether you are participating in class discussion and responding to questions without always being prompted.

Discussion moderation (10%):
Students working in threes will be responsible for moderating discussion from one textbook chapter (two class sessions of class). For the sessions you are moderating, you will be evaluated on (1) whether you came prepared with questions or concepts that promote critical thinking. You don't need to have all the answers, but you should facilitate a discussion that reaches an answer. (2) Attempt active participation, which generally means making space for everyone to participate (i.e., not dominating the conversation), ensuring everyone feels comfortable asking questions, and being flexible in how the discussion proceeds. (3) Keeping things on track, that is, paying attention to the time and trying to ensure focus stays on the things that matter rather than minutia. As this will be done in small groups, a final point is to ensure that everyone in the group has space to participate in moderating; if you're someone who talks a lot, consider stepping back, if you're someone who is generally quiet, figure out how to step up.

Data visualization project (15%):
Exploratory data analysis (EDA) and data visualization is an integral part of biostatistical research. There is a textbook chapter that we will read on data visualization, but this is a big and important topic. To delve deeper into this topic, we will complete a data visualization assignment where students will find a dataset, use EDA to understand the structure of the data and explore potential relationships among variables, and then create optimized visualizations that intuitively provides insight into a biological result from the dataset. The deliverables will be a short (5 minute) presentation of your dataset and figure in class and a written report that describes the process you undertook to arrive at your final figure.

Weekly assignments (15%):
Weekly assignments will bridge the textbook reading and the two course projects (Data visualization project and the Points of Significance project). They are primarily intended to be scaffolded assignments that are self-directed learning towards finding, reading, interpreting and evaluating scientific literature.

Point of significance project (30%):
This is the final semester project. Although we will cover a fair amount of material during the semester, the topic of advanced biostatistics is large and we will be unable to discuss many important statistical methods. This project gives students to do a deep dive into a statistical method that we are not able to cover. Students will look through Chapter 7-11 in the Modern Statistics for Modern Biology textbook and pick a topic. They will do a literature search on google scholar to identify at least three biological papers thats use their method of choice. Students will pitch their topics and papers in a short (5 minute) presentation to the class. Based on feedback from the

class they will pick one paper and develop their topic into a 10 minute presentation to the class that briefly describes the biological background of the paper, walks through how the method was applied, and discusses the major biological implications that were reached because of the analysis that was done. Finally, students will submit a written manuscript similar to the style of a "Point of Significance" paper from Nature Methods.

Letter Grade: I normally use the following cutoffs when assigning letter grades:

Letter Grade	Mark out of 100
A+	90-100
A	80-90
B+	75-80
B	70-75
C+	65-70
C	60-65
D	50-60
F	below 50

Class Communications

The University requires all students to activate an official U of M email account, which should be used for all communications between yourself and the university (including all your instructors). All these email communications should comply with the University's policy on electronic communication with students, which can be found at: http://umanitoba.ca/admin/governance/governing_documents/community/electronic_communication_with_students_policy.html

Questions of an administrative nature should be directed to me via email. Questions related to the course content should be directed to the Discussion Groups on UM Learn (which I will regularly visit). This is in order to provide an opportunity for learning and collaboration between the students.

Student Accessibility Services

If you are a student with a disability, please contact Student Accessibility Services (SAS) for academic accommodation supports and services such as note-taking, interpreting, assistive technology and exam accommodations. Students who have, or think they may have, a disability (e.g. mental illness, learning, medical, hearing, injury-related, visual) are invited to contact SAS to arrange a confidential consultation.

Academic Integrity

Students are encouraged to discuss course material, including assignments and the final project. However, each student must hand in his or her own copy of assignments. Copying from anywhere, including other students, books, or the internet constitutes a case of academic dishonesty and could have serious consequences.

The goal in this class (as in all academic pursuits) is to learn. If you are unclear on what is acceptable, please ask me or visit the Faculty of Science page on cheating and plagiarism at: http://umanitoba.ca/faculties/science/undergrad/resources/webdisciplinedocuments.html

Professional Conduct

It is recognized that these are unusual circumstances, and some adjustments need to be made when working virtually. At the same time, you are reminded that University policies, such as the Respectful Work and Learning Environment policy, still apply, as do basic expectations around how students will engage with each other and all members of the University. This means that when participating in classes, online meetings, etc., students are expected to behave professionally and follow the same basic norms as they would in person. Essentially, if you wouldn't do it in an in-person class, don't do it in a virtual setting. Please familiarize yourself with the UM Respectful Work and Learning Environment (RWLE) Section 2.5(c) of the Student Non-Academic Misconduct and Concerning Behaviour Procedure describes types of inappropriate or disruptive behaviour.

Like all other University-wide announcements, it is anticipated that the activity plans (be they remote or in-person) for after February 26 will be made by email as well as posted on the University website.

While we are online, we will try and mimic the in-person environment as much as possible. If your internet bandwidth permits, students are asked to attend class with their videos on. If you video points to an open room, you are encouraged to blur the background. Please keep the audio muted unless you are speaking. This class is delivered synchronously and lectures will not be recorded.

Image Recordings: Screen shot capture and/or video recording of material presented by the instructor and/or classmate is strictly for personal use only due to copyright and/or privacy concerns. Posting of images that include lecture material and/or instructor and/or classmates on the internet is strictly prohibited. Recording and/or distribution of exam material in any format is strictly prohibited.

Medical notes: Students who are unable to meet a course requirement due to medical circumstances are currently not required to submit medical notes. Purposefully built-in to the course assessment is some slack In grading and a back-up plan for students who miss major course deliverables due to illness. However, please note that circumstances that result in missing multiple course components (e.g. assignments/exams/classes) may require medical documentation (e.g., Authorized Withdrawal, Tuition Fee Appeal, Leave of Absence, or accessibility-related accommodations). Students are advised to speak with an advisor in their faculty/college/school of registration in this case.

ROASS Schedule A

Schedule "A" of the *Responsibilities of Academic Staff with regards to Students* (ROASS) policies of the University of Manitoba lists resources and policies for students. It is important that you familiarize yourself with these resources and policies. This document is available from the Department of Statistics web page at: https://sci.umanitoba.ca/statistics/.

Students enrolled in this course must ensure they satisfy the following minimum technological requirements:

1. A computing device where one can create and edit documents,
2. An internet connection capable of streaming videos and downloading software, and
3. Access to a web-cam and microphone.

University of Manitoba Acknowledgement of Traditional Territories

The University of Manitoba campuses are located on original lands of Anishinaabeg, Cree, Oji-Cree, Dakota, and Dene peoples, and on the homeland of the Métis Nation.

We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of reconciliation and collaboration.