BCB 731: Critical readings in biomedical statistics and machine learning (Fall 2023, 2 credits)

Course Description: BCB 731 is a survey of recurring statistical errors and pitfalls which are sometimes used to exaggerate the weight of evidence for novel biological claims or inflate the estimated accuracy of proposed predictive biomedical models. This course focuses on misapplied analyses of data sources where a small number of biological samples are quantified into very high dimensional feature spaces, such as in genomics, proteomics, and biomedical imaging.

Prerequisites: Students should be familiar with both predictive modeling and inferential data analysis (BIOS 661, COMP 755, or STOR 665) as well as genomic data derived from DNA and RNA (BCB 716). Permission of the instructor required for students lacking the prerequisites.

When: Monday & Wednesday / 2pm – 3:15pm / October 2nd - December 6th

Where: Marsico 5004

Instructor

Alex Rubinsteyn, PhD Assistant Professor Department of Genetics 11202B Mary Ellen Jones Email: alex.rubinsteyn@unc.edu

Office Hours: Available by appointment.

Course Website: darkarts.bio

Course Format

The course format will be a seminar-style class that meets twice weekly. The first week will consist of lectures by the instructor meant to act as a refresher on statistics, machine learning, and best practices for reproducible model training and evaluation. For the remainder of the course, each week will focus on a particular paper meant to exemplify a particular category of statistical error. The instructor will present necessary background for understanding each paper's assays, techniques, and biological claims. The contents of the papers will be presented by students enacting one of two roles: the *Optimist* and *Critic*. In the first session of each week, the *Optimist* will present a summary of the paper's claims as credulously and uncritically as possible. In the second session for a paper, the *Critic* will deconstruct a core claim of the paper and discuss what statistical error created the appearance of support for this claim. Lastly, the class will discuss the erroneous data analysis pattern, in what cases a similar analysis might be correct, and how to distinguish these scenarios in other research contexts.

Course-at-a-Glance

The instructor reserves the right to make changes to the syllabus, including topics, readings, assignments, and due dates. Any changes will be announced as early as possible. For papers and session-by-session course schedule details, please the course website (<u>darkarts.bio</u>).

Week	Session	Торіс
1	Oct 2 nd	Recap of classical statistics and null hypothesis significance testing
	Oct 4 th	Machine learning and reproducible predictive model evaluation
2	Oct 9 th	Just make it up: image manipulation and tables which don't add up
		(Background & Optimist)
	Oct 11 th	Critic & Discussion
3	Oct 16 th	Try everything and report the small p-values: uncorrected
		multiple hypothesis testing (Background & Optimist)
	Oct 18 th	Critic & Discussion
4	Oct 23 rd	My predictor is perfect on the training set: lack of independent
		validation data over-estimates accuracy (Background & Optimist)
	Oct 25 th	Critic & Discussion
5	Oct 30 th	My predictor is perfect if I tell it the labels: information leakage
		between training and test datasets (Background & Optimist)
	Nov 1 st	Critic & Discussion
6	Nov 6 th	Shrink your p-values by increasing the sample size: deriving
		correlated samples from a small cohort (<i>Background & Optimist</i>)
	Nov 8 th	Critic & Discussion
7	Nov 13 th	Makes lots of predictors and report the best: simultaneous model
		selection and accuracy estimation (<i>Background & Optimist</i>)
	Nov 15 th	Critic & Discussion
8	Nov 20 th	Confounders are a handy source of accuracy: learning artifactual
		differences between conditions (Background & Optimist)
	Nov 22 nd	Critic & Discussion
9	Nov 27 th	Delay until sick patients drop out: pop-random censoring in
5		clinical trial survival analysis (<i>Backaround & Ontimist</i>)
	Nov 29 th	Critic & Discussion
10	Dec 4 th	Novel biology is easy to discover with bad FDR control:
		inconsistent false discovery rate methods (Background & Optimist)
	Dec 6 th	Critic & Discussion

Course Assignments and Assessments

This course will be graded 50% for participation in course discussions and 50% for successfully playing the roles of "Optimist" and/or "Critic" in presenting papers to class.

Course Grading Scale(s)

Final course grades will be determined using the following <u>UNC Graduate School grading scale</u>. The relative weight of each course component is shown in the Graded Assignments section.

- **H**—High Pass (93-100): Clear excellence
- **P**—Pass (80-92): Entirely satisfactory graduate work
- **L**—Low Pass (70-79): Inadequate graduate work
- **F**—Fail (0-69)

Map of Competencies to Learning Objectives and Assessment Assignments

Competency: As an elective course, BCB 731 does not fulfill any required competencies for any PhD curriculum in the Biological and Biomedical Sciences Program (BBSP).

BCB 731 fulfills the following course-specific competencies: the course helps students develop an intuition for when statistical techniques are being misapplied or when a claim of a model's predictive accuracy may not generalize to real world usage. While the course assumes some familiarity with statistics, it teaches students how to think about the soundness of analyses across classical statistics, Bayesian statistics, machine learning, and more recent large model deep learning techniques applied to biomedical data.

Expectations, Policies, and Resources

Accessibility at UNC Chapel Hill

The University of North Carolina at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing University courses, programs and activities. Accommodations are determined through the Office of Accessibility Resources and Service (ARS) for individuals with documented qualifying disabilities in accordance with applicable state and federal laws. See the <u>ARS Website</u> for contact information or email <u>ars@unc.edu</u>.

Attendance/ Participation

Your attendance and active participation are an integral part of your learning experience in this course. If you are unavoidably absent, please notify the course instructor. No right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences:

- 1. Authorized University activities
- Disability/religious observance/pregnancy, as required by law and approved by <u>Accessibility Resources and Service</u> and/or the <u>Equal Opportunity and Compliance</u> <u>Office</u>.
- Significant health condition and/or personal/family emergency as approved by the <u>Office</u> of the Dean of Students, <u>Gender Violence Service Coordinators</u>, and/or the <u>Equal</u> <u>Opportunity and Compliance Office</u>.

Community Standards in Our Course and Mask Use.

UNC-Chapel Hill is committed to the well-being of our community – not just physically, but emotionally. The indoor mask requirement was lifeted for most of campus on March 7, 2022. If you feel more comfortable wearing a mask, you are free to do so. There are many reasons why a person may decide to continue to wear a mask, and we respect that choice. For additional information, see <u>Carolina Together</u>.

Counseling and Psychological Services at UNC Chapel Hill

CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to the <u>CAPS website</u>, call them at 919-966-3658, or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.

Honor Code

You are not permitted to upload any content from this course to the web in any form, including but not limited to Chegg, Course Hero, Coursera, Google Drive, etc. If you post my course content, you may be violating my intellectual property rights. If you post your own work from this course, you are allowing sites to profit from your intellectual property. In utilizing web sources to upload or download course content, you risk violating the University's Honor Code.

If you have any questions about your rights and responsibilities, consult the <u>Office of Student</u> <u>Conduct</u> or review the following resources: <u>Honor System</u>; <u>Honor System module</u>; <u>UNC Library's</u> <u>plagiarism tutorial</u>; <u>UNC Writing Center's handout on plagiarism</u>. I expect all students to follow the guidelines of the UNC Honor Code. In particular, students are expected to refrain from "lying, cheating, or stealing" in the academic context. You can read more about the honor code at <u>studentconduct.unc.edu</u>. In any course, including mine, what constitutes cheating can change from one activity to another. For example, collaboration may be encouraged for an assignment but qualify as cheating during an exam. Please see my guidelines for each activity, and if you are unsure, please ask me to clarify. In remote classes, there may be many temptations for using online exchange sites, such as Chegg. Note that these sites provide names of students who have used their materials, and they routinely cooperate with institutions around academic integrity issues. Please don't get caught up with honor code issues just because it appears to be simple and untraceable. It is not!

As a student at UNC Chapel Hill, you are bound by the <u>university's Honor Code</u>, through which UNC maintains standards of academic excellence and community values. It is your responsibility to learn about and abide by the code. To ensure an effective Honor System at UNC, in this course students are expected to:

- Conduct all academic work within the letter and spirit of the Honor Code, which prohibits the giving or receiving of unauthorized aid in all academic processes.
- Students may use materials they worked on in other classes or in research, but only if the student themselves produced the work.
- For homework, students may verbally discuss approaches to the problems but each student should independently write up the answer and verify solutions. <u>Do not submit</u> <u>identical solutions!</u>
- For take-home mid-term, students must work completely independently without communicating with other students, tutors, or anyone else about any material related to the test questions. The test is 'open book' and 'open notes.'
- For the final project (and in general in your research) <u>never copy any text from a publication/review/book without attribution</u>. An exception is a direct quotation with quotation marks, which can be used sparingly. Taking a copied section of text, not quoting it, and changing the phrasing here and there is also considered plagiarism, and doesn't help, as plagiarism detection scripts can find a match regardless.

Inclusive Excellence

We are committed to expanding diversity and inclusiveness across the School — among faculty, staff, students, on advisory groups, and in our curricula, leadership, policies and practices. We measure diversity and inclusion not only in numbers, but also by the extent to which students, alumni, faculty, and staff members perceive the School's environment as welcoming, valuing all individuals, and supporting their development.

For more information about how we are practicing inclusive excellence at the Gillings School, visit the following webpages: <u>Inclusive Excellence</u>, <u>Inclusive Excellence Action Plan</u>, <u>Minority Health</u> <u>Conference</u>, and <u>National Health Equity Research Webcast</u>. Additional campus resources include: the <u>LGBTQ Center</u>; <u>Non-Discrimination Policies at UNC</u> <u>Chapel Hill</u>; <u>Ombuds</u>; and <u>Prohibited Discrimination</u>, <u>Harassment</u>, and <u>Related Misconduct at</u> <u>UNC Chapel Hill</u>.

In this class, we practice the Gillings School's commitment to inclusion, diversity, anti-racism and equity in the following ways.

- Develop classroom participation approaches that acknowledge the diversity of ways of contributing in the classroom and foster participation and engagement of *all* students.
- Structure assessment approaches that acknowledge different methods for acquiring knowledge and demonstrating proficiency.
- Encourage and solicit feedback from students to continually improve inclusive practices.
- Treat all members of the Gillings community (students, faculty, and staff) as human persons of equal worth who deserve dignity and respect, even in moments of conflict and disagreement.
- Contribute to creating a welcoming and inclusive classroom environment, where all are able to learn and grow from one another.
- Acknowledge and respect the diversity of experiences that others bring to the classroom and the ways in which this richness enhances everyone's learning
- Strive to maintain a spirit of curiosity and generosity, particularly in the face of new and/or seemingly contradictory information and perspectives Encourage and solicit feedback from students to continually improve inclusive practices.

Land Acknowledgement

Please read The Gillings School's Land Acknowledgement.

Student Feedback and Equity Concerns

The Gillings School has in place a mechanism for students to provide feedback, including specifically equity concerns and bias-related issues. You can use this form to describe feedback, both positive and negative, about anything including issues related to your experience as a student at Gillings, administrative processes, and classroom activities. This form will also allow you to specifically describe incidents in which racial or other equity-related bias, or microaggressions, occurred. You may submit this form anonymously. However, for us to follow up and provide the necessary support, we encourage you to include your contact information. For further information, please visit the <u>Student Feedback and Equity Concerns FAQ</u>. Please note that this form does not take the place of any University process or policy. If you would like to report an incident under the University's policy on <u>Prohibited Discrimination</u>, <u>Harassment, and Related Misconduct including Sexual and Gender Based Harassment, Sexual Violence, Interpersonal Violence, and Stalking</u>, please visit <u>Safe At UNC</u> or the <u>Equal Opportunity and Compliance Office</u> (EOC) for additional information, including resources, contact, and reporting options.

Technical Support

The best way to help prevent technical issues from causing problems for assignments and quizzes is to submit them at least 24-36 hours before the due date and time. Your instructor cannot resolve technical issues, but it's important to notify them if you are experiencing issues. If you have problems submitting an assignment or taking a quiz in Sakai, immediately do the following:

- 1. Contact the UNC Information Technology Services (ITS) department with the time you attempted to do your course action and what the course action was.
- 2. Email your instructor with the information you sent to ITS and what time you sent the information.

The ITS department provides technical support 24-hours per day, seven days per week. If you need computer help, please contact the ITS Help Desk by phone at +1-919-962-HELP (4357), or by <u>online help request</u>, or by <u>UNC Live Chat</u>.

Title IX at UNC Chapel Hill

Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitations, or stalking is encouraged to seek resources on campus or in the community. Please contact the Director of Title IX Compliance / Title IX Coordinator (Adrienne Allison), Report and Response Coordinators in the Equal Opportunity and Compliance Office, Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (confidential) to discuss your specific needs. Additional resources are available at the <u>"Safe at UNC" website</u>.