STAT 201 - Statistical Inference for Data Science

Syllabus

Course description: Classical and simulation-based techniques for estimation and hypothesis testing, including inference for means and proportions. Emphasis on case studies and real data sets, as well as reproducible and transparent workflows when writing computer scripts for analysis and reports.

Prerequisites: DSCI 100 - Introduction to Data Science.

Required: A laptop.

- If a student does not have their own laptop or chromebook, students may be able to loan a laptop from the UBC library.

Course Website: canvas.ubc.ca

- Please check Canvas regularly to keep up-to-date with the course.

Instructor: Marie Auger-Méthé

Contact: Please use **Piazza** for most questions (course material, re-grading, etc) and come to our **office hours**. Contact instructor at auger-methe@stat.ubc.ca *only* for personal matters.

Office Hours: Please see Canvas course main page for times and locations.

Learning outcomes:

By the end of the course, students will be able to:

- Describe real-world examples of questions that can be answered with the statistical inference methods presented in this course (e.g., estimation, hypothesis testing) and apply inference skills and concepts to answer such questions.
- Explain what random and representative samples are and how they can influence estimation.
- Write computer scripts to perform estimation and hypothesis testing via simulation-based inference approaches, as well as by applying results from exact and approximate distributional theory.
- Interpret and explain results from confidence intervals and hypothesis tests.
- Compare the application of simulation-based inference approaches with the application of exact and approximate distributional results.
- Effectively visualize point estimates and different measures of uncertainty (e.g., confidence intervals, standard errors) by writing computer scripts.
- Discuss the impact of type I & II errors as well as responsible use and reporting of p-values on hypothesis tests.
- Explain estimator bias and uncertainty, and write a computer script to calculate it.
- Discuss how an estimator's bias arises (e.g., sample bias, study design), and its implications in statistical inference.
- Perform all aspects of a statistical analysis (from data consumption to reporting) using reproducible and transparent computer scripts.

Teaching method:

Students are expected to complete recommended readings before class. During the lecture, the instructor will review concepts, deliver course material, and ask Clicker questions to check progress and provide feedback to students. The remainder of the lecture time will be given to students to work on worksheet activities in Jupyter Notebooks. Please bring a device (tablet, laptop) to lectures and tutorials.

Exam dates:

- Midterm: Wednesday, October 22, 2025 (in class)
- Final Exam: To be scheduled by Classroom Services

Clickers:

We will be using the iClicker Cloud in lectures. It is free for students. iClicker is a response system that allows you to respond to questions posed by instructors during class. You will need to have an iClicker account associated with your Canvas profile by Wednesday, Sep 8 4:00 PM.

See canvas **iClicker assignment** for details.

Piazza Discussion Board:

We will use **Piazza** for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TAs, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com

Software platforms:

- Students will learn to perform their analysis using the **R programming language**.
- Worksheets and tutorial problem sets will be done using **Jupyter Notebooks**.
- Students will access the worksheets and tutorials in Jupyter Notebooks through Canvas.

Course assessments:

Assessment	Date	Percentage
Clickers	In class	7%
Lecture worksheets	Weekly (**due dates on Canvas**)	10%
Tutorial worksheets	Weekly (**due dates on Canvas**)	13%
Midterm	**Wednesday Oct 22 (in class)**	25%
Final Exam	To be scheduled by Classroom Services	45%

- I will drop the following automatically:
  - 2 lowest iClicker poll grades
  - 1 lowest lecture worksheet grade
  - 1 lowest tutorial worksheet grade

Late registrants are automatically given extensions on lecture 01 and tutorial 01 worksheets.

Bonus	Date	Percentage
Midterm survey	TBD	0.5%
Top 10 Piazza helpers	TBD	1%

Recommended textbooks:

- ModernDive: Statistical Inference via Data Science developed by Chester Ismay and Albert Y. Kim.
  - This book is freely available: https://moderndive.com/
- OpenIntro Statistics by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr which is available by donation.
  - This book is available by donation (whatever you can afford): https://www.openintro.org/book/os/

Worksheets:

Each week there will be two assignments: (1) lecture worksheet; and (2) tutorial worksheets. The lecture worksheets are fully autograded with visible tests to help you identify points that need more clarification. Reach out to the teaching team if you don't understand why you are getting an answer wrong in the worksheet. On the other hand, the tutorials are not fully autograded, and only a few exercises will have visible tests. You can access the assignments through Canvas assignments.

To submit your assignment, make sure your work is saved **on our server** (i.e., using the link from Canvas) before the deadline. Our server will automatically snapshot at the due date/time. Also, please **do not** rename the assignments files.

Policies

Late/Absence

- Students are expected to attend all lectures and tutorials, and attendance during the lecture will be assessed via iclicker. We allow for absences due sickness/personal reasons by automatically dropping the 2 lowest iclicker grades. Students who are absent from numerous lectures because of illness or other reasons should inform the instructor(s) of the course as soon as possible, preferably, prior to the start of the lecture/tutorial.
- There will be no make-up exams. Students who miss the Midterm and want to request an Academic Concession need to contact the Instructor as soon as possible. Failing to present a declaration may result in a grade of zero on the midterm.
- Late submissions of **lecture worksheets** and **tutorial worksheets** will receive a grade of 0. Note that one worksheet and one tutorial will be dropped at the end of the semester. If you have extenuating circumstances and need concessions beyond dropping one worksheet and one tutorial, please email the instructor.

Autograder Policy Many of the questions in assignments are graded automatically by software. The grading computer has exactly the same hardware setup as the server that students work on. No assignment, when completed, should take longer than 5 minutes to run on the server. The autograder will automatically stop (time out) for each student assignment after a maximum of 5 minutes; *any ungraded questions at that point will receive a score of 0.*

Students are responsible for making sure their assignments are *reproducible*, and run from beginning to end on the autograding computer. Creating reproducible code is an important data science skill! *Please ensure that any data that needs to be downloaded is done so by the assignment notebook with the correct filename to the correct folder.* A common mistake is to manually download data when working on the assignment, making the autograder unable to find the data and often resulting in an assignment grade of 0. Even small mistakes such as using the wrong sample size will justify an incorrect response for that (and possibly downstream) questions.

In short: whatever grade the autograder returns after 5 minutes (assuming the teaching team did not make an error) is the grade that will be assigned.

Tip: when you're done the assignment, click "Restart Kernel and Run All", and check that the autograder returns the results you are expecting.

Regrading: If you have concerns about the way your work was graded, please open a request **within one week** of having the grade returned to you. After this one-week window, we may deny your request for re-evaluation. Also, please keep in mind that your grade may go up or down as a result of re-grading. To open a regrade requests, please follow the steps below:

1. Go to Piazza and click on *New post*.
2. In *Post Type*, select *Question*.
3. Make the post private to instructors and TAs only. In *Post to* select *Individual Students(s)/Instructor(s)*. A text box will appear, where you must type *Instructors*.
4. In *Select Folder(s)* select the folder *regrading*
5. In *Summary* say the Assignment you want to be regraded, followed by the question and your name and student number. For example, *lab 3 -> Q3 – Rodolfo Lourenzutti (9982313)*
6. Provide a brief reason for why the regrade is needed.
7. The TAs will see the request and will take a look at the assignment. If necessary, they will involve the instructors. Finally, once the TA is finished reassessing the assignment:

- If the grade deserves more marks: the TA will update the mark on Canvas and comment on the question so everyone can see that the question has been addressed.
- If your grade goes down or stays the same: the TA will answer the post on Piazza, giving the student a reason for their final decision

Device/Browser Students are responsible for using a device and browser compatible with all functionality of Canvas. Chrome or Firefox browsers are recommended; Safari has had issues with Canvas quizzes in the past.

Missed Final Exam Students who miss the final exam must report to their faculty advising office within 72 hours of the missed exam, and must supply supporting documentation. Only your faculty advising office can grant deferred standing in a course. You must also notify your instructor prior to (if possible) or immediately after the exam. Your instructor will let you know when you are expected to write your deferred exam. Deferred exams will ONLY be provided to students who have applied for and received deferred standing from their faculty.

Academic Concession Policy Please see **UBC's concession policy** for detailed information on dealing with missed coursework, quizzes, and exams under circumstances of an acute and unanticipated nature.

Reach Out for Success:

University students often encounter setbacks from time to time that can impact academic performance. Learn about how you can plan for success at **students.ubc.ca**. For help addressing mental or physical health concerns, including seeing a UBC counsellor or doctor, visit **students.ubc.ca/health/wellness-centre**. Science students can also refer to **science.ubc.ca/students/wellbeing** If you require accommodations and/or support due to disabilities and ongoing medical conditions, contact the **Centre for Accessibility**. Discuss your situation with an academic advisor or your instructor.

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access, including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious and cultural observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest academic standards in all of their actions. Details of the policies and how to access support are available on the **UBC Senate website**.

Preferred/Chosen Names:

At UBC, students can update their given name to reflect their preferred/chosen name across many university systems. Learn more about this process here: **Preferred/Chosen Names at UBC**. If you have any concerns about ID verification during exams or other questions, please feel welcome to reach out.

Academic Integrity:

All students are expected to follow UBC's Academic Honesty and Standards policy. We encourage students to work together on assignments and labs, however all of your work must be written in your own words. Students must correctly cite references if you quote or use outside sources in your work, including any AI tools. Breach of the academic integrity policy may, at a minimum, result in a grade of 0 on the relevant assessment or may result in more serious consequences. Please see UBC's Academic Calendar for detailed policies on **Academic Misconduct**.

Use of Generative AI (GenAI)

Students are responsible for completing all assessments independently, demonstrating their understanding and mastery of the course material. Generative AI (GenAI) tools (e.g., ChatGPT) can be useful for tasks such as gathering information, brainstorming ideas, reviewing concepts, finding analogies, revising written work, or creating study questions. However, their use should be limited, and **all submitted work must reflect the student's own efforts and comprehension**. Ultimately, your goal is to master the material, so ensure that using these tools supports your learning rather than hinders it. **Important: copy-and-pasting or paraphrasing the answer given by a genAI tool is not acceptable as an answer to assigment's questions (e.g., lecture and tutorial worksheets).**

If you choose to use GenAI tools in your work, you must:

- Cite the GenAI tool you used, including the name of the tool and how it was applied.
- Provide an annotation with your citation explaining the extent to which the GenAI tool contributed to your work and a brief description of its usage.

Important: **The use of GenAI to bypass learning or misrepresent your effort is considered academic dishonesty.** It also deprives you of the opportunity to fully engage with the material and grow as a learner and individual. If an instructor suspects that a submission relies excessively on AI, you may be required to explain your work in more detail. If your explanation is unsatisfactory, your submission will receive a penalty which is scaled according to the severity of the violation and may result in a score of zero on the entire assignment.

Check out **UBC's approach to generative artificial intelligence tools in teaching and learning**.

Health and Safety:

Please follow **UBC Communicable Disease Prevention Framework** 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. I will do my best to stay well, but if I am ill, I will not come to class. If that happens, here's what you can expect:

- Another instructor will substitute.
- Class may take place over Zoom (in this case, Zoom link will be posted on Canvas). I will send any updates via Canvas Announcements as soon as I can.

Extreme Environmental Conditions Contingency Plan:

In-person, on campus activities may need to be cancelled due to issues such as weather conditions (e.g., snow). The most up-to-date information about cancellations will be posted on **ubc.ca**. Please check ubc.ca often during times when an extreme weather event could disrupt our course activities. Here is what you can expect in the event an in-person lecture or lab session is cancelled: Depending on the nature of the planned in-class activities, class may take place over Zoom (in this case, Zoom link will be posted on Canvas), or an alternate activity may be posted on Canvas for you to complete before the next scheduled class. I will communicate over Canvas to announce the specifics for each case that arises as soon as I can.

Copyright:

Attribution:

This syllabus was originally based on the syllabus from MDS and DSCI 100, and has been modified through the years by many STAT 201 instructors.