STAT 201 – Statistical Inference for Data Science
Syllabus

COURSE INFORMATION

Course Title	Course Code Number	Credit Value
Statistical Inference for Data Science	STAT 201	3

PREREQUISITES

DSCI 100

CONTACT

- All course questions **must** be posted on Piazza. Direct contact only for personal matters.
- If you want/need a private chat, make an appointment using Canvas Message.

Course Instructor(s)	Contact Details	Office Location	Office Hours
Rodolfo Lourenzutti	Canvas messages (**preferred**); Email (**avoid if possible**)	ESB 3158	

OTHER INSTRUCTIONAL STAFF

Course TA(s)	Contact Details	Office Hours
Jonathan Agyeman	Piazza	TBD (check canvas)
Wakeel Kasali	Piazza	TBD (check canvas)
Sasha Sharma	Piazza	TBD (check canvas)

COURSE STRUCTURE

The course is structured in weekly lectures and tutorials.

The lectures will be expository with in-class activities.

Students are expected to attend lectures AND tutorials.

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).
- 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.
- 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.
- Perform all aspects of statistical analysis (from data consumption to reporting) using reproducible and transparent computer scripts.
- Explain sequential testing and principled peeking and how they can be used for early stopping of an experiment.

SCHEDULE OF TOPICS

This is a *tentative schedule* for the course. It is subject to changes.

Module	Date	Due Dates
Module 1 (Intro to Statistical Inference) **Module 2** (Bootstrapping)	Jul-01 (Tuesday) Jul-03 (Thursday)	- Worksheet 1 - Tutorial 1 - Worksheet 2 - Tutorial 2
Module 3 (Mathematical Models and CLT) **Module 4** (Simulation-Based Conf. Intervals)	Jul-08 (Tuesday) Jul-10 (Thursday)	- Worksheet 3 - Tutorial 3 - Worksheet 4 - Tutorial 4
Module 5 (Traditional Conf. Intervals) **Module 6** (Simulation-Based Hypothesis Tests)	Jul-15 (Tuesday) Jul-17 (Thursday)	- Worksheet 5 - Tutorial 5 - Worksheet 6 - Tutorial 6
Module 7 (Traditional Hypothesis Tests) **Midterm**	Jul-22 (Tuesday) Jul-24 (Thursday)	- Worksheet 7 - Tutorial 7 - Midterm
Module 8 (Errors in Inference) **Module 9** (ANOVA and Multiple Hyp. Tests)	Jul-29 (Tuesday) Jul-31 (Thursday)	- Worksheet 8 - Tutorial 8 - Worksheet 9 - Tutorial 9

Module 10 **(A/B Testing)** **Course Review**	Mar-12 (Tuesday) Mar-17 (Thursday)	- Worksheet 10 (due end of Thursday).

LEARNING ACTIVITIES

This course will have plenty of synchronous activities that students will have to work on during the lectures and tutorials. Students will work on activities in Jupyter Notebooks and iClickers.

LEARNING MATERIALS

I strive not to require students to buy textbooks. That's not always possible. For this course specifically, we have:

- ModernDive: Statistical Inference via Data Science (MD)
- Openintro to Statistics (OIS)
- Introduction to Modern Statistics (IMS)

ASSESSMENTS OF LEARNING

In this course, we will use an effort-based assessment. The assessment of learning will have two components:

- **Mid-course assessments (50%):**
  - **Lectures' Worksheets (8%):** fully auto-graded with visible tests to help you identify points that need more clarification.
  - **Tutorials' Worksheets (12%):** Only a few exercises will have visible tests.
  - **Midterm (40%)**
- **Final Exam (50%)**
- **Pass the final policy** (weaker version)**:** A score below 40% in the final exam results in an automatic fail.

- **Note 1:** As you can see, the mid-course assessment is worth 50%, but the sum of the components is worth 60%. This means that you can miss half of the worksheet/tutorials and still get full marks in the mid-course assessment – and, because of this, no concessions will be granted for worksheet/tutorials.

- **Note 2**: The types of questions in the exams can vary: reasoning, multiple-choice, multiple-answer, dropdown, true or false. Although most questions will be about the content, you can expect a few coding questions. That being said, the coding question will not be overly complicated, and we will only check your familiarity with the main functions and packages we use in the course. We are not trying to test your memory!!! Please don't spend energy trying to memorize everything. You will have access to an R environment during the exams, where you can check the functions' documentation.

UNIVERSITY POLICIES

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise, 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 observances. UBC values academic honesty, and students are expected to acknowledge the ideas generated by others and 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.**

Some resources

- Academic Concession
- Academic Honesty and Standards
- Attendance
- Grading Practices
- Student Conduct and Discipline
- Viewing Marked Work

COURSE POLICIES

Late/Absence

- Students are expected to attend lectures and tutorials regularly.
- There will be no make-up exams. Students who miss Midterm and want to request an Academic Concession need to contact the instructor as soon as possible and provide a self-declaration form. Failing to present a declaration may result in a grade of zero on the assessment.
- Late submissions will receive a grade of 0. Remember, this is an effort-based assessment.

- If you have extenuating circumstances and need concessions **beyond** the flexibility provided by the effort-based assessment, contact the instructor to discuss your situation.

Missed Final

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.

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 request, 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.

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. Furthermore, students are responsible for making sure their assignments are reproducible and run from beginning to end on the auto-grading computer. 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 with the assignment, click "Restart and Run All" and check that the visible tests are all working and that your notebook runs in less than 5 minutes.

COPYRIGHT