Syllabus

Explanatory and predictive data analysis with multiple explanatory variables. Choosing the right methods to apply based on the statistical question and data at hand. Trade-offs between model-based and non-model based approaches. Emphasis placed on case studies and real data sets, as well as reproducible and transparent workflows when writing computer scripts for analysis and reports.

Prerequisites

STAT 201 and one of MATH 100, MATH 102, MATH 104, MATH 110, MATH 120, MATH 180, MATH 184, SCIE 001.

Instructor:

Dr. Gabriela Cohen Freue (ESB 3146)

- Use CANVAS to send me an email. Please use email only for personal matters that you would want to discuss privately.
- Use office hours and Piazza Discussion Board for questions regarding assignments, projects and class note examples etc.

Teaching & Learning Activities:

The course is structured in weekly lectures and tutorials. The lectures are **mandatory** *and* **in person**. Tutorials are **in person** and a good space to work collaboratively with peers on worksheets, tutorials and project. This course will have plenty of synchronous activities that students must work on during the lectures and tutorials. **Students are expected to attend lectures and tutorials.

The lectures will be expositive with the use of in-class activities. Students will work on activities in Jupyter Notebooks . Students need to bring a laptop to class to work on activities in Jupyter Notebooks and iClickers. If a student does not have their own laptop or chromebook, students may be able to loan a laptop from the UBC library.

Office Hours:

- Instructor: Thursdays, 11am – 12pm (Room TBD)
- TAs office hours: Fridays, 1pm - 2pm (Online in Zoom)
- Alternative times can be added by request

Course pages:

- Please **check the Canvas website regularly** to keep up-to-date with the course. Important course information and announcements will be frequently posted there. Everything you need will be available through Canvas and you should get familiar with all the tabs as soon as possible.
  - If you have any problems related to technical issues, please use ?Help (see the left side menu in the Canvas course page) to report the problem or to contact IT service.
- Post questions on PIAZZA for general discussions and answers!

IMPORTANT cancellations

- Midterm Break: February 17-21
- If the university is closed due to extreme weather conditions or other reasons, lectures will run via Zoom or recorded lectures will be posted on CANVAS. Detailed information will be posted through CANVAS announcements.

Test Dates

- Midterm: Wednesday, February 12, 2025 during class time and at class room.
- Final Exam: TBD
- Final Project Report: individual and group assignments deadlines will be reported in Canvas course page.

Software Platforms

- Students will learn to perform their analysis using the R programming language.
- Worksheets and tutorial problem sets as well as the final project analysis, development, and reports will be done using Jupyter Notebooks and R.
- Students will access the worksheets and tutorials in Jupyter Notebooks through Canvas and work on through the course server.

Textbooks

Textbooks are recommended for extra context and reference but are not strictly required for the course.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R.
  - Available ONLINE at UBC Library: http://resolve.library.ubc.ca/cgi-bin/catsearch?bid=6667014
- Rafael Irizarry. Introduction to Data Science.
  - Available ONLINE at https://rafalab.github.io/dsbook/
- Modern Dive: we used this book in STAT 201. This books also has some useful material for this course, in particular:
  - Chapter 5: Basic Regression
  - Chapter 6: Multiple Regression
  - Chapter 9: Hypothesis Testing
  - Chapter 10: Inference for Regression

Learning Outcomes

By the end of the course, students are expected to be able to:

- Describe real-world examples of explanatory modelling (e.g. A/B testing optimization & regression with variable selection) and predictive modelling problems.
- Explain the trade-offs between model-based and non-model based approaches, and describe situations where each might be the preferred approach.
  - Explain the difference between creating models for explanation vs prediction, in the context of both how you choose and evaluate models as well as how you interpret the results.
- Choose & apply a suitable method (e.g., regression, GLM's, sample size estimation, controlling for multiple testing, peeking, bandit algorithms, variable selection, model diagnostics) based on the statistical question and data at hand. Discuss the advantages and disadvantages of different methods that may be suitable for a given problem.
- Correctly interpret computer output when performing the statistical analyses presented in this course, in the context of the statistical question being asked and the audience being reported to.
- Identify the assumptions / conditions required for each method to produce reliable results. Choose techniques to check (or at least be able to falsify) those assumptions. Discuss the consequence(s) of mapping the wrong methods to the question and/or data type.

Assessments

Most weeks there will be two assignments: (1) a worksheet; and (2) a tutorial. Deadline TBD.

The worksheets are fully autograded with visible tests to help you identify points that need more clarification. Therefore, 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.

You can access the assignments through Canvas (assignments).

Assessments' Weights

- **Worksheets**: 3%
- **Tutorials**: 5%
- **Clicker Qs**: 2%
- **Project**: 20%
- **Midterm**: 25%
- **Final Exam**: 45%
- **Bonus Piazza**: 1%

Project

- **Team Work Contract**: 1%
- **Individual Assignment 1 (Planning Stage)**: 5%
- **Individual Assignment 2 (Analysis Stage)**: 5%
- **Group Interview with TA**: 2%
- **Group Final Report**: 5%
- **Teammate Evaluation**: 2%

Note: Please refer the Canvas Course page for deadlines

Worksheets and Tutorials

- Worksheets are fully auto-graded with visible tests to help you identify points that need more clarification.
- Tutorials have only a few exercises will have visible tests.
- Make sure your work is saved **on our server** (i.e., accessed using the link from Canvas) before the deadline. Our server will automatically snapshot at the due date/time.
- Please **do not** rename the assignments files.
- While you can work locally on your computer, only work saved on the course's server will be graded. Please, do not use other courses' servers (e.g., Syzygy).

iClicker cloud: We will be using iClicker Cloud in lectures. iClicker Cloud is a response system that allows you to use your own computer or mobile device to respond to questions posed by instructors during class. You need to set up an iClicker Cloud account and add STAT 301 as a course to this account. To do so, please follow https://lthub.ubc.ca/guides/iclicker-cloud-student-guide for details.

Project A project based on a case study that you will work on throughout the term. Details about this assignment will be made available to you on Canvas. You will work on this project both individually and in group. You can (and are encouraged to) discuss all parts of the project with your group members. However, every student will submit their own individual assignments and will receive an individual grade on the individual components.

Exams All the exams will be on Canvas with lockdown browser. You will be able to have 1 letter-size page (double-sided) cheatsheet (all you can write or print). You **are not allowed** to access any webpage or files in your computer or other electronic devices.

The types of questions can vary: reasoning, multiple-choice, multiple-answer, dropdown, true or false. Although most questions will be about the content, you can expect a few simple 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 software development skills!!! Please don't spend energy trying to memorize everything. If you had done the worksheets and tutorial, this should not be a problem for you.

- **Midterm:** administered during the full lecture of Week 8. The midterm will cover Worksheets/Tutorials 1-5. Content in readings are to help you get a full understanding but won't be tested if not covered in course material.
- **Final Exam:** The final exam will be a two-hour exam and it will cover the material of the **entire course**. However, more emphasis will be given to the second part of the course (i.e., Worksheets/Tutorials 6-10).

Piazza Discussion Board and Bonus points: To have access Piazza, go to "Piazza" in the left menu in the Canvas course page and it will open in a new window. Then you can sign up for the class page. You can use "Piazza Discussion Board" to post your questions and also to provide answers/hints to the questions posted there. This is where you can discuss ideas, strategies, and resources for solving the problems with your classmates.

TAs will monitor the questions in the piazza page and post answers. But do not expect we will answer all your questions posted in Piazza page with detailed solutions. If you need more clarification, it's always better to contact TAs or me during our office hours.

Do not provide full answer to questions in Piazza. The idea is to use this platform to generate a fruitful discussion, give hints and tips, but not full solutions. The same applies to other platforms (e.g., Discord).

Students that have answered the most statistics-related questions in Piazza in a way that explains concepts well but does not reveal the answer to an assignment, lab, or webwork question will get a bonus 1% added to their grade. When you answer question, teaching team endorse your answers as "good answer". I add this 1% if you have more than 10 Endorsed Answers.

Course Policies

Missing the midterm:

- There will be no make-up exam.
- Students who miss an exam should notify the instructor prior to (if possible) or immediately after the exam to request an Academic Concession and provide a self-declaration form (available on canvas).
- Failing to contact the instructor may result in a grade of zero on the Midterm.
- If your request is approved by the instructor, your midterm weight will be moved to the final exam.

Missed Final Exam Students who miss the final exam must report to their faculty advising office within 48 hours of the missed exam, and must apply for deferred standing: https://students.ubc.ca/enrolment/academic-learning-resources/academic-advising. 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.

If you're a Science student, you must apply for deferred standing (an academic concession) through Science Advising no later than 48 hours after the missed final exam/assignment. Learn more and find the application online:https://science.ubc.ca/students/advising/concession.

Students who are granted deferred standing write the final exam/assignment at a later date.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.

Late/Absence

- Regular attendance to lecture and tutorials is expected of students. Students who are unavoidably absent 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.
- Late submissions of *worksheets* and *tutorials* will receive a grade of 0.
- For other assessments, late submission is defined as any work submitted after the deadline. Late submissions will receive a 50% deduction penalty of the original grade in the first occurrence if submitted within 48 hours of the deadline. Hence a maximum attainable grade for the first piece of work submitted late is 50%. Any additional pieces of work that are submitted late will receive a grade of 0 for subsequent occurrences. Any submission after 48 hours of the original deadline will receive a grade of 0.

Excused assignment

- We are aware that sometimes life gets in the way of getting things done. For example, you might feel sick in a given week, have a family emergency, or feel overwhelmed with your other courses. For this reason, we will drop the lowest grade worksheet, tutorial and clicker (one of each) at the end of the semester. This will be done automatically, so you don't need to let us know.

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.

Tip: when you're done the assignment, click "Restart 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. Lockdown browser will be used for exams so students must address any issues with their laptop

prior to the exam. We will test it with a mock quiz before the exam in the review session. Chrome or Firefox browsers are recommended; Safari has had issues with Canvas quizzes in the past.

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.

Academic Integrity Discussion of ideas leaned in class is encourage (with other students, TAs or the instructor). This helps the leaning process. But individual work turned in by each student should be your own work.

The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work.

- Do not copy or paraphrase solutions from other students or from other sources.
- Do not provide your solutions to another student.

Violations of academic integrity (i.e., misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in a mark of zero on the assignment or exam and more serious consequences may apply if the matter is referred to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences.

A more detailed description of academic integrity, including the University's policies and procedures, may be found in the Academic Calendar.

Plagiarism Students must correctly cite any code or text that has been authored by someone else or by the student themselves for other assignments. Cases of plagiarism may include, but are not limited to:

1. the reproduction (copying and pasting) of code or text with none or minimal (e.g., changing the name of the variables).
2. the translation of an algorithm or a script from a language to another.
3. the generation of code by automatic code-generation software.

An "adequate acknowledgement" requires a detailed identification of the (parts of the) code or text reused and a full citation of the original source code that has been reused.

The above attribution policy applies only to assignments. **No code or text may be copied (with or without attribution) from any source during a quiz or exam**. Answers must always be in your own words. At a minimum, copying will result in a grade of 0 for the related question.

Repeated plagiarism of any form could result in larger penalties, including failure of the course.

For more information about UBC Policies

- Academic Honesty and Standards: https://vancouver.calendar.ubc.ca/campus-wide- policies-and-regulations/academic-honesty-and-standards
- Academic Integrity: https://academicintegrity.ubc.ca/about-academic-integrity/
- Academic Misconduct: https://academicintegrity.ubc.ca/regulation-process/academic- misconduct/
- Resources and Support: https://academicintegrity.ubc.ca/resources/

Reach Out for Success

- University students often encounter setbacks from time to time that can impact academic performance. Discuss your situation with your instructor or an academic advisor. Learn about how you can plan for success at: www.students.ubc.ca

- For help addressing mental or physical health concerns, including seeing a UBC counsellor or doctor, visit: https://students.ubc.ca/health-wellness

UBC policies and resources to support student success:

- 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 at https://senate.ubc.ca/policies-resources-support-student-success.

Land acknowledgement:

We acknowledge that the UBC Vancouver campus is situated within the traditional, ancestral and unceded territory of the Musqueam First Nation.