Data Analysis Syllabus

Department of Mathematics and Statistics

STAT4620/5620 Winter 2024

Dalhousie University acknowledges that we are in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq People and pays respect to the Indigenous knowledges held by the Mi'kmaq People, and to the wisdom of their Elders past and present. The Mi'kmaq People signed Peace and Friendship Treaties with the Crown, and section 35 of the Constitution Act, 1982 recognizes and affirms Aboriginal and Treaty rights. We are all Treaty people.

Dalhousie University also acknowledges the histories, contributions, and legacies of African Nova Scotians, who have been here for over 400 years.

Course Instructor

Name	Email	Office Hours
Dr. Joanna Mills Flemming	Joanna.Flemming@Dal.Ca	January 9 – February 13: TR 2:35pm-3:25pm February 14 – April 8: MWF 10:35am-11:25am

Course Description

This course begins with an introduction to the multi-disciplinary field of *data science*, making clear the role of statistics therein. Issues surrounding *data ethics* and *reproducibility* are investigated followed by a review of tools for *exploratory data analysis* (EDA). *Statistical models* are detailed commencing with *linear models* (LMs) and *generalized linear models* (GLMs). Next, *additive,* and *generalized additive models* (GAMs) are described followed by their *mixed model* extensions. *Tree-based methods*, *longitudinal models* and *spatial statistics* are demonstrated with a view to enhancing each student's ***statistical toolbox***. Emphasis is placed on understanding model assumptions and method implementation. Real and relevant data sets are used throughout the course to demonstrate best practices for data analysis. **The R programming language** is used exclusively.

At the beginning of this course students select a dataset of relevance to their field of study (or interest). Each student analyzes their dataset and prepares a report (due at the end of the term) describing their data, analyses, and findings. Graduate students are required to orally present both their proposal and final report. If preferred, students may work in small (≤ 3) groups.

Course Prerequisites

STAT 3340, STAT 3460, or the instructor's consent.

Course Exclusions

None.

Course Structure

Course Delivery

Course delivery will be **in-person** except for three lectures (February 6, February 8, and March 7) that will be delivered online (synchronous via Zoom video conferencing). No lectures will be recorded.

Lectures

Tuesdays and Thursdays, 1:05pm-2:25pm, LSC-Common Area C208.

Course Materials

1. There is no required textbook but rather two suggested reference textbooks: Generalized Additive Models: An Introduction with R, Second Edition and Core Statistics, both authored by Simon N. Wood.
2. Articles and videos will also be used to supplement learning.
3. Brightspace, R and R Studio are required.
    a. The Brightspace site provides lecture materials, assessment instructions, and communication tools. Students are responsible for checking the site regularly.
    b. Students are required to use the statistical software program R and the integrated development environment, R Studio, for all data analyses. Both software packages are free to download to your personal computer. The dynamic report generator embedded within R-Studio, called R-markdown, is to be used for the course project.

Assessment

Assessment	*Weight (% of final grade)*	*Due Date*
Reading Responses	15%	Jan 15, Jan 29, Feb 12, Feb 26, Mar 11
Project Proposal (5 minutes)	0% Undergrad / 5% Grad	Feb 15
Assignments (4)	40% Undergrad / 30% Grad	Jan 25, Feb 8, Mar 7, Mar 22
Project Presentation (10 minutes)	0% Undergrad / 5% Grad	Mar 26 and 28

Final exam (80 minutes)	15%	Apr 4
Project Report (~5000 words)	30%	Apr 10

Conversion of numerical grades to final letter grades follows the Dalhousie Grade Scale

A+ (90-100)	B+ (77-79)	C+ (65-69)	D (50-54)
A (85-89)	B (73-76)	C (60-64)	F (0-49)
A- (80-84)	B- (70-72)	C- (55-59)	

Course Policies on Missed or Late Academic Requirements

Any missed requirements are automatically be given a grade of 0. This includes failing to complete annotations, assignments, presentations, the final exam, or the final project on time. No make-up opportunities are available.

Course Policies related to Academic Integrity

For group projects, each member is required to submit an individual reflection on your final product, as well as evaluate the individual contribution of each member of your group to this final product. Related presentations require all group members to be present and involved.

Students are allowed to work together on assignments but must submit them individually. Responses to questions on the assignment must be appropriately referenced. Annotations must be completed independently.

Learning Objectives

This course aims to provide (upper level undergraduate and graduate) students with an awareness of important considerations when undertaking data analysis along with working knowledge of a range of statistical methodologies. Students develop the confidence to perform appropriate data analyses to answer scientific (and other) questions of interest.

List of knowledge/skills student are expected to have upon completion of this course:

- Capacity to recognize important features of data (e.g., heterogeneity, dependence).
- Understanding of zero-inflation, zero-truncation, and over/under-dispersion.
- Proficiency with fitting GLMs, GAMs and their extensions.
- Knowledge of hierarchical modelling frameworks and interpretation of random effects.

- Understanding of tree-based methods.
- Appreciation for the field of spatial statistics.
- Working knowledge of the R language and environment for statistical computing and graphics.

Course Content

Week	Date	Lesson Topic(s)	Assessment
1	January 9 and 11	Data Science, Data Ethics, and Reproducibility	
2	January 16 and 18	Exploratory Data Analysis	Reading Response 1
3	January 23 and 25	Linear Models (LMs)	Assignment 1
4	January 30 and February 1	Generalized Linear Models (GLMs)	Reading Response 2
5	February 6 and 8	Overdispersion and Zero-inflation	Assignment 2
6	February 13 and 15	Additive Models (AMs)	Reading Response 3 / Proposal Presentations
7	February 20 and 22	STUDY BREAK	
8	February 27 and 29	Generalized Linear Mixed Models (GLMMs)	Reading Response 4
9	March 5 and 7	Spatiotemporal Models	Assignment 3
10	March 12 and 14	Species Distribution Models	Reading Response 5
11	March 19 and 22	Tree Based Methods	Assignment 4
12	March 26 and 28		Project Presentations
13	April 2 and 4	Review and Final Exam	
14	April 10	EXAM Week	Project Report

University Policies and Statements

Recognition of Mi'kmaq Territory

Dalhousie University would like to acknowledge that the University is on Traditional Mi'kmaq Territory. The Elders in Residence program provides students with access to First Nations elders for guidance, counsel, and support. Visit or e-mail the Indigenous Student Centre at 1321 Edward St or elders@dal.ca. Additional information regarding the Indigenous Student Centre can be found at: https://www.dal.ca/campus_life/communities/indigenous.html

Internationalization

At Dalhousie, 'thinking and acting globally' enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders." Additional internationalization information can be found at: https://www.dal.ca/about-dal/internationalization.html

Academic Integrity

At Dalhousie University, we are guided in all our work by the values of academic integrity: honesty, trust, fairness, responsibility, and respect. As a student, you are required to demonstrate these values in all the work you do. The University provides policies and procedures that every member of the university community is required to follow to ensure academic integrity. Additional academic integrity information can be found at: https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Accessibility

The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation. If there are aspects of the design, instruction, and/or experiences within this course (online or in-person) that result in barriers to your inclusion, please contact the Student Accessibility Centre (https://www.dal.ca/campus_life/academic-support/accessibility.html) for all courses offered by Dalhousie with the exception of Truro. For courses offered by the Faculty of Agriculture, please contact the Student Success Centre in Truro (https://www.dal.ca/about-dal/agricultural-campus/student-success-centre.html)

Conduct in the Classroom – Culture of Respect

Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

Diversity and Inclusion – Culture of Respect

Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality. Dalhousie is strengthened in our diversity. We are a respectful and inclusive community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness (Strategic Priority 5.2). Additional diversity and inclusion information can be found at: http://www.dal.ca/cultureofrespect.html

Student Code of Conduct

Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students don't follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal manner - perhaps through a restorative justice process. If an informal resolution can't be reached, or would be inappropriate, procedures exist for formal dispute resolution. The full Code of Student Conduct can be found at: https://www.dal.ca/dept/university_secretariat/policies/student-life/code-of-student-conduct.html

Fair Dealing Policy

The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie. Additional information regarding the Fair Dealing Policy can be found at: https://www.dal.ca/dept/university_secretariat/policies/academic/fair-dealing-policy-.html

Originality Checking Software

The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method. Additional information regarding Originality Checking Software can be found at:
https://www.dal.ca/dept/university_secretariat/policies/academic/student-submission-of-assignments-and-use-of-originality-checking-software-policy-.html

Student Use of Course Materials

Course materials are designed for use as part of this course at Dalhousie University and are the property of the instructor unless otherwise stated. Third party copyrighted materials (such as books, journal articles, music, videos, etc.) have either been licensed for use in this course or fall under an exception or limitation in Canadian Copyright law. Copying this course material for distribution (e.g. uploading to a commercial third-party website) may lead to a violation of Copyright law.