Math 3293FA01
Statistical Learning - Fall 2022
Course Outline (June 20, 2022 version)

Instructor: Hugh Chipman (he/him/his), HSH 157, hugh.chipman@gmail.com / hugh.chipman@acadiau.ca (either email goes to the same place)

Office Hours: to be announced on Acorn.

Lectures: T/Th 2:30 - 4:00

Instruction: This course is being offered as a "Maple League" course. Lectures are hybrid. I will teach in person on campus, and simultaneously broadcast via MS Teams.

Textbook: "An Introduction to Statistical Learning, with Applications in R", 2nd Edition, by James, Witten, Hastie & Tibshirani. Free pdf at https://www.statlearning.com/

Graduate course: For Acadia students, a graduate version of the course (Math 5193) may be offered. The description below is for both offerings of the course.

Website: Both the undergraduate and graduate course will use the undergraduate (Math 3293) acorn page.

Marking Scheme (The higher mark in the two options is the one awarded)

Option	tests	assignments	project
1	40	25	35
2	30	35	35

5193 Marking Scheme:
The independent study report (below) will be worth 20 points. All other elements have the same values as in above, and the better of the two options will be awarded. Course mark = (100/120)*(your mark out of 120).

Assignments: There will be 4 assignments during the term (dates below). Part of each assignment will be not-for-credit. A few questions will be for credit. Solutions to all questions will be posted after the due date. Late assignments will generally not be accepted.

Tests: There will be two tests, given during scheduled class times (dates below). If you miss a test for a legitimate reason, accommodations will be made.

Project: Students will be expected to apply the methods learned in the course to the analysis of a substantial data set, and write a report summarizing their findings. The project will be divided into 3 parts:
1. A proposal, in which you identify a data set and conduct a simple analysis,
2. A draft, which I will provide feedback on,
3. A final version, which incorporates feedback from the draft.

Expectations for 5193:
The graduate offering of this course will be held with 3293. Additional expectations are:
1. Some assignment and test questions may be different from 3293;
2. Independent study of a more advanced topic. Prepare a report that summarizes the main ideas of the assigned readings and demonstrates the application of this technique to data.

It is anticipated that graduate students will meet (virtually) with the instructor 2 - 3 times a month to discuss assigned readings (research papers). Note that this report is in addition to the project.

Tentative course schedule:

Thu Sep 8	First class
Thu Sep 22	Assignment 1 due
Thu Oct 6	Assignment 2 due
Thu Oct 13	Test #1
Thu Oct 27	Assignment 3 due
Mon Oct 31 - Fri Nov 4	Acadia reading week - no classes
Tue Nov 8	Project proposal due
Tue Nov 15	Assignment 4 due
Tue Nov 29	Test #2
Tue Dec 6	Project draft due (and last class)
Tue Dec 20	Project due (as take-home exam)

Computer Use:
The course will make extensive use of the (free) R software environment for statistical computing and graphics, and the RStudio interface to R. Students are encouraged to install R and RStudio on their laptop before classes begin. Copies of R may be downloaded from http://www.r-project.org/ . R runs on a wide variety of UNIX platforms, Windows and MacOS. RStudio can be downloaded from www.rstudio.com.You should install R first, then RStudio.

Note: It is your responsibility to be aware of course requirements and deadlines, and to be present for tests and exams. If you are unable to attend a lecture, it is your responsibility both to make up any material covered in your class, as well as find out any announcements made during the lecture missed. It is a courtesy to your instructor to advise him when you must be absent from lectures. Missing tests or assignment deadlines because of illness or personal emergencies should be documented by signed excuses from physicians or counsellors.

Although collaboration in solving assignments is acceptable, the copying of assignments is inappropriate and will be considered cheating. Submitting someone else's work as your own, copying on tests, and other forms of cheating will be dealt with under the clause concerning Academic Integrity in the Acadia Academic Calendar.

Accessible Learning
Acadia University is dedicated to improving access to campus life for all students with disabilities. While we attempt to ensure that all courses are accessible, we recognize that there are barriers that need to be addressed on an individual basis. Students who require accommodations to complete coursework or otherwise fully participate in class should contact Accessible Learning Services directly as soon as possible. Please visit Accessible Learning Services website or email them at disability.access@acadiau.ca for more information.

I will also make similar accommodations for accessible learning at other Maple League institutions.

PREREQUISITES: Math 2223 or 2243 with a minimum of C-, or permission of the instructor.

Calendar description: Modern statistical methods for supervised and unsupervised learning with large and complex data. Topics include: linear regression, classification, resampling methods, model selection and regularization, smooth regression, tree-based models, support vector machines, principal components and dimension reduction, clustering and statistical graphics.

Topic	Chapter	Week
Introduction	1	1
Statistical Learning	2	1
Linear Regression	3	2-3
Classification	4	3-4
Resampling methods	5	5
Linear Model Selection and Regularization	6	6
Moving beyond linearity	7	7
Tree-Based methods	8	8-9
Deep Learning & Neural Networks	10	10
Unsupervised Learning	12	11