

ACSC/STAT 3740, Predictive Analytics
WINTER 2023
Toby Kenney

Instructor:	Toby Kenney Department of Mathematics and Statistics Chase Building, Room 102 email: tkenney@mathstat.dal.ca
Course Website:	Brightspace
Office Hours:	TBA
Lectures:	TT: 10:05-11:25 LSC C220
Topics:	Data exploration; Data visualisation; Modelling data; Model validation; Communication
Textbook:	Regression Modeling with Actuarial and Financial Applications by Edward W. Frees published by Cambridge University Press Chapters 1-3, 5-9, 11-13
Additional References:	Data Visualization: A Practical Introduction Healy Princeton University Press, 2018 Currently available online at https://socviz.co/ Chapters 1-4
	An Introduction to Statistical Learning, with Applications in R, James, Witten, Hastie & Tibshirani Springer, 2013 Available online at https://statlearning.com/ Chapters 2-6, 8, 10

Course Work and method of assessment

There will be a project and a final exam. The project will be completed during the term and submitted within one week of the end of classes (so by 23:59 on 18th April). The final exam will be arranged **online** during the examination period 13-28th April.

There will also be 5 homework assignments, which must usually be submitted on Thursdays. After the deadline, I will put the model solutions on the course website. **No credit can be given for late homework.** The overall homework mark will be made up of an average of the homework marks, with the exception of the worst mark for each student.

The homework sheet will be divided into 2 sections: The *basic questions* section tests the basic concepts covered in the course: everyone should be able to do all these questions. The *standard questions* section has questions where the concepts covered in the course can be applied to more realistic situations, or questions which involve a stronger theoretical insight; these questions are mostly straightforward, though there may be the occasional tricky question included. There may also be some *bonus questions* which are either more challenging, or else raise interesting or important issues that are not central to this course.

Grades will be determined by performance in the exam, project and the weekly homeworks. The project counts for 40%, the final counts for 40%, while the homework counts for the remaining 20%.

Discussion Lectures

Because of the practical nature of this course, an important part of the instruction will involve discussion of your ideas and approaches to problems. To facilitate this, some problems for discussion will be given in lectures, on the course website and on brightspace. You will then be expected to spend a short time (20–30 minutes) approaching these problems (you can work in groups if you prefer), and in the following lecture, we will discuss your approaches. In cases involving coding, the easiest approach is to send the code to me before the lecture, then we can present the code and discuss it during the lecture with minimal technical disruptions.

These discussion lectures will mostly be on Thursdays. The other lectures, mostly on Tuesdays will introduce the relevant topics and approaches.

This is the current schedule of planned lectures and discussion sessions.

Week	Tuesday	Thursday
9th January	Introduction and Preliminaries, Problem Formation, Data Selection, R setup.	Data visualisation
16th January	Making graphics with <code>ggplot</code>	Data visualisation
23rd January	Data Exploration	Data Exploration
30th January	Model Fitting	Model Fitting
6th February	Time Series Models	Time Series Models
13th February	Time Series Models	Project Planning
20th February	STUDY BREAK	
27th February	Tree-based Methods	Tree-based Methods
6th March	Model Validation	Model Validation
13th March	Variable Selection	Variable Selection
20th March	Communication	Communication
27th March	Case study	Case study
4th April	Revision	Revision

Material covered

Rather than covering particular material, the purpose of this course is to practice applying material in realistic scenarios. We will introduce particular topics as needed for the data analysis problems.

Students with disabilities

Students with disabilities are encouraged to register as quickly as possible at the Student Accessibility Services if they want to receive academic accommodations. To do so, please phone 494-2836, email access@dal.ca, drop in at the Killam, G28, or visit our website at www.studentaccessibility.dal.ca.

Plagiarism

Plagiarism is a serious academic offense which may lead to loss of credit, suspension or expulsion from the university. Please read the Policy on Intellectual Honesty contained in the Calendar or on the Dalhousie web site at: <http://www.registrar.dal.ca/calendar/ug/UREG.htm#12>.

Dalhousie Writing Centre

Writing expectations at university are higher than you will have experienced at high school (or if you are entering a master's or PhD program, the expectations are higher than at lower levels). The Writing Centre is a Student Service academic unit that supports your writing development. Make an appointment to discuss your writing. Learning more about the writing process and discipline-specific practices and conventions will allow you to adapt more easily to your field of study.

Internet Resource Policy

As the course is designed to reflect real world problems, I will take a more relaxed approach to use of internet resources in answering homework questions. Using an internet search to find approaches to a problem is possible.