

# DATA\*6100

## Introduction to Data Science

Fall 2023



---

### 1 INSTRUCTIONAL SUPPORT

#### 1.1 Instructor

Please see [courselink](#) for details

---

### 2 COURSE DESCRIPTION

#### 2.1 Calendar Description

The course includes an introduction to the methods of modern statistics such as splines, general additive models, principal components analysis, and classifiers. Students learn resampling methods such as bootstrap, cross-validation, boosting, and bagging. Methods of model selection include search-and-score and regularization, and students practice communicating technical ideas to a non-technical audience, including via data visualization.

**Credit Weight:** 0.5    **Department:** Mathematics & Statistics    **Campus:** Guelph

**Restriction(s):** Restricted to Master of Data Science students.

#### 2.2 Course Topics

- Review of multiple linear regression
- Fundamentals of model building: data splitting for training, validating and testing
- Nonlinear regression models: logistic regression, regression splines, generalized additive model
- Resampling methods: bootstrapping, permutations, boosting, bagging, parallel computing

- Data reduction through principal components, model scoring, variable selection and regularization

## 2.3 Learning Objectives

Upon successful completion of this course, students will have demonstrated the ability to:

1. Understand full data analysis process: data collection, cleaning, description, exploratory data analysis, model building, model checking, model selection, knowledge transfer.
2. Perform data splitting for model training, model validation and model testing; differentiate estimation from prediction task.
3. Demonstrate understanding of regression model framework; differences between linear, generalized linear and generalized additive models.
4. Build nonlinear regression model based on logistic regression, regression splines and generalized additive models.
5. Perform resampling techniques, such as bootstrapping, cross-validation, boosting, and bagging.
6. Perform data reduction through principal components, model scoring and variable selection.
7. Design non-parametric test based on permutations or bootstrapping.
8. Understand basics of data ethics: fairness, bias, privacy, transparency, security, impact and consequences.

## 2.4 Important Dates

(see also <https://calendar.uoguelph.ca/graduate-calendar/schedule-dates/fall-semester/>)

**First day of lecture:** Thursday, Sept 7 2023.

**Fall Break:** Friday, Oct 7 to Tuesday, Oct 10 2023. (No classes)

**Last day of classes:** Friday, Dec 1 2023.

**Drop Deadline:** Courses that are one semester long must be dropped by the last day of class (**Friday, Dec 1, 2023**).

**Course Feedback Information:** Near the end of the term, you will be given the opportunity to evaluate your instructor and provide comments regarding your experience. Your instructor will inform you of when these are to take place.

---

# 3 LEARNING RESOURCES

## 3.1 Course Website

Course material, news, announcements, and grades will be regularly posted to the DATA\*6100 Courselink website. *You are responsible for keeping up-to-date on this site.*

### 3.2 Textbooks and Resources

*An Introduction to Statistical Learning Second Edition* by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. (Note: We will use the NEW Python version of the book)

Available for free online from the authors at <https://www.statlearning.com/>

Other notes and other resources will be posted on the Courselink website.

### 3.3 Lecture Information

***Note that due to the COVID-19 pandemic, lectures may be rescheduled to online. Instead of the in person lecture time, synchronous classes will meet over Zoom during the lecture time. The Zoom link for the lecture will be accessible from Courselink.*** Please check the Courselink website to make sure you know where lectures are happening.

### 3.4 Mathmatize, Campuswire and karma bonus points

In lectures, we will be using the Mathmatize platform for student participation. Outside of lecture, we will be using the Campuswire platform to help facilitate student questions and student discussion during the term. You are highly encouraged to post questions to Campuswire and to help answer other students questions. Good questions will not just state a problem, but will include what you have tried so far and clarification on what exactly it is you don't understand. Good answers will not just provide the final answer but will explain *why* with a well justified solution.

By being an active participant in lectures, you can earn up to 5 karma points. The instructor will identify good questions and good answers on Campuswire during the term. You can earn up to 10 Karma bonus points during the term for your participation on Campuswire by posting good questions and/or answers and/or meaningfully participating in student discussions.

Karma points increase your course grade by the following prescription: if you earned  $n$  Karma points, then those points are yours and the regular grading scheme will apply to the remaining  $(100 - n)$  of your course grade. (Example: If you earned 10 Karma points, and got 80% on the rest of the course, then your final grade is  $10 + (100 - 10)80\% = 82$ )

### 3.5 Communication & Email Policy

Please use office hours and Campuswire discussion forums as your main opportunity to ask math content questions about the course. Email is to be used for logistical purposes and math questions are not generally answered by email. Major announcements will be posted to the course Courselink website. **It is your responsibility to check the course website regularly.** As per university regulations, all students are also required to check their <uoguelph.ca> e-mail account regularly: e-mail is the official route of communication between the University and its students.

### 3.6 Online Behaviour

Inappropriate online behaviour will not be tolerated. Examples of inappropriate behavior include:

- Posting inflammatory messages about your instructor or fellow students or any other topic
- Using obscene or offensive language online
- Copying or presenting someone else's work as your own
- Adapting information from the Internet without using proper citations or references
- Buying or selling term papers or assignments
- Posting or selling course materials to course notes websites
- Having someone else complete your homework or completing homework for another student
- Making false claims about assignment submissions or other course content
- Threatening or harassing a student or instructor online
- Discriminating against fellow students, instructors or TAs
- Using the course website to promote profit-driven products or services
- Attempting to compromise the security or functionality of the learning management system
- Sharing your user name and password
- Recording lectures without the permission of the instructor

## 4 ASSESSMENT

### 4.1 Grade Distribution

Midterm Project 1 (due 11:59pm Fri Oct 6, 2023)	25%
Midterm Project 2 (due 11:59pm Fri Nov 10, 2023)	25%
Final Project (due 11:59pm Fri Dec 1, 2023)	40%
Peer Feedback (Participation in evaluating <i>other</i> students' projects) (Always due <i>1 week</i> after project deadline)	10%

### 4.2 Projects

The course projects will have both a written mathematical component that must be typed up using LaTeX as well as programming components that will be written in Python and submitted to Gradescope. You will also be required to record a short video presentation explaining your work. Detailed instructions and expectations will be posted on Courselink. The due dates are listed in the grade distribution table.

Late submissions will not be accepted.

### 4.3 Peer Feedback

Your projects will also receive peer review from other students in the course using the University of Guelph PEAR (Peer Evaluation, Assessment and Review) Tool. Other student's evaluation

will not effect your grade on projects, but are a chance for you to reflect and improve your work. (You will be given an opportunity to make changes based on the feedback before final grading)

Your participation in the review will constitute 10% of your final grade process (i.e. you are graded on giving meaningful feedback to *other* students in the course). Your Peer Review for other student projects are always due **1 week** after the due date for each project.

---

## 5 MENTAL HEALTH SERVICES

One out of every five students in Canada experiences some sort of mental health issue at some point in their academic career. If you find yourself facing a mental health crisis, or just need to talk to someone, please consider taking advantage of one of the following resources available to University of Guelph students:

***Counselling Services:*** Visit the Counselling Services website (<https://wellness.uoguelph.ca/counselling>) to get information on resources available to you, both online and in-person. You can also visit them at Health Services (J.T. Powell Building, ext 53244) where they offer individual and group counselling sessions by appointment or walk-in.

***Student Support Network:*** is located in the Wellness & Education Promotion Centre in the J.T. Powell Building and offers confidential, peer-based, drop-in support.

***Good2Talk:*** ([1-866-925-5454](tel:1-866-925-5454)) is a free, 24/7 student hotline that provides professional counselling and referrals for mental health, addictions and well-being.

***Here 24/7:*** ([1-844-437-3247](tel:1-844-437-3247)) specializes in assessment, referral and appointment booking and is available 24/7 for crisis support.

You are not alone and you will not be judged for asking for help.

---

## 6 EMAIL COMMUNICATION

As per university regulations, all students are required to check their <uoguelph.ca> e-mail account regularly: e-mail is the official route of communication between the University and its students.

---

## 7 WHEN YOU CANNOT MEET A COURSE REQUIREMENT

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons, please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. See the Undergraduate Calendar for information on regulations and procedures for [Academic Consideration](#).

---

## 8 DROP DATE

Courses that are one semester long must be dropped by the end of the last day of classes; two-semester courses must be dropped by the last day of classes in the second semester. Please see the Graduate Calendar for detailed policies.

---

## 9 COPIES OF OUT-OF-CLASS ASSIGNMENTS

Keep paper and/or other reliable back-up copies of all out-of-class assignments: you may be asked to resubmit work at any time.

---

## 10 ACCESSIBILITY

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.

When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required, however, interim accommodations may be possible while that process is underway. Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to make a booking at least 14 days in advance, and no later than November 1 (fall), March 1 (winter) or July 1 (summer). Similarly, new or changed accommodations for online quizzes, tests and exams must be approved at least a week ahead of time.

More information: [www.uoguelph.ca/sas](http://www.uoguelph.ca/sas)

---

## 11 ACADEMIC MISCONDUCT

The University of Guelph is committed to upholding the highest standards of academic integrity and it is the responsibility of all members of the University community – faculty, staff, and students – to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff and students have the responsibility of supporting an environment that discourages misconduct. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

The [Academic Misconduct Policy](#) is outlined in the Undergraduate Calendar.

---

## 12 RECORDING OF MATERIALS

Presentations which are made in relation to course work—including lectures—cannot be recorded or copied without the permission of the presenter, whether the instructor, a classmate or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

---

## 13 RESOURCES

The [Academic Calendars](#) are the source of information about the University of Guelph's procedures, policies and regulations which apply to undergraduate, graduate and diploma programs.

---

## 14 DISCLAIMER

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings, changes in classroom protocols, and academic schedules. Any such changes will be announced via Courselink and/or class email.

This includes on-campus scheduling during the semester, mid-terms and final examination schedules. All University-wide decisions will be posted on the COVID-19 website (<https://news.uoguelph.ca/2019-novel-coronavirus-information/>) and circulated by email.

---

## 15 ILLNESS

Medical notes will not normally be required for singular instances of academic consideration, although students may be required to provide supporting documentation for multiple missed assessments or when involving a large part of a course (e.g., final exam or major assignment).

---

## 16 COVID-19 SAFETY PROTOCOLS

For information on current safety protocols, follow these links:

- <https://news.uoguelph.ca/return-to-campuses/how-u-of-g-is-preparing-for-your-safe-return/>
- <https://news.uoguelph.ca/return-to-campuses/spaces/#ClassroomSpaces>

Please note, that these guidelines may be updated as required in response to evolving University, Public Health or government directives.