

## Department of Statistical and Actuarial Sciences SS 3850G – Data Analysis

### Course outline for Winter 2021



Although this academic year might be different, Western University is committed to a **thriving campus**. We encourage you to check out the [Digital Student Experience](#) website to manage your academics and well-being. Additionally, the following link provides available resources to support students on and off campus: <https://www.uwo.ca/health/>.

#### Technical Requirements and Important Dates:



Stable internet connection



Laptop or computer



Working microphone



Working webcam



Classes Start	Reading Week	Classes End	Study day(s)	Exam Period
January 11	February 15 - 19	April 12	April 13	April 14 - 30

\* March 15, 2021: Last day to drop a second-term half course or a second-term full course without penalty

## 1. Course Information

### Course information

Course name – Data Analysis  
Course Number: SS3850G  
Academic Term: Winter 2020/21

Lecture: Online Asynchronous – OWL  
Tutorial/Q&A: Friday 4:30pm to 5:30pm on Zoom or equivalent  
Place: Online

### List of Prerequisites

A minimum mark of 60% in Statistical Sciences 2858A/B or a minimum mark of 65% in Statistical Sciences 3869A/B.

Pass in Statistical Sciences 2864A/B.

## List of Antirequisite(s)

CS 4414 / CS9637A/ CS9114A / SE4460B

Unless you have either the requisites for this course or written special permission from your Dean to enroll in it, you may be removed from this course and it will be deleted from your record. This decision may not be appealed. You will receive no adjustment to your fees in the event that you are dropped from a course for failing to have the necessary prerequisites.

## 2. Instructor Information



Course Coordinator	Contact Information
Cristián Bravo Román	<a href="mailto:cbravoro@uwo.ca">cbravoro@uwo.ca</a>

Instructor(s) or Teaching Assistant(s)	Contact Information
Yishan Zang	<a href="mailto:yzang8@uwo.ca">yzang8@uwo.ca</a>

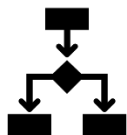
Office Hours	Medium	Names
Wednesdays 11am – 12pm	Zoom	Cristián

Students must use their Western (@uwo.ca) email addresses when contacting their instructors.

## 3. Course Syllabus, Schedule, and Delivery Mode

Introduces machine learning and statistical methods for data analysis through applied examples. The goal of this course to expose students to topics related to statistical learning such as, Linear Regression, Logistic Regression, Discriminant Analysis, Model Selection and Regularization, Cross Validation, Tree Based Methods and Clustering. Particular emphasis is placed on how to rigorously evaluate an analysis of data.

Students will develop a data science analysis project using the methods covered in class.



Topics include:

- Supervised Learning and Model Fitting
- Statistics, Prediction, and Maximum Likelihood
- Introduce test set/out-of-sample idea.
- Classification, Evaluation, Logistic regression Regularization, Multi-class problems
- Estimating Performance, Quantifying Uncertainty on parameter estimates and on model predictions
- Test error, Cross-validation, Model Selection, Bias-Variance tradeoff

- Feature Selection and Regularization (L1 and L2)
- Trees, Random Forest
- Neural Networks, Gradients, learning
- Autoencoders, Dimensionality reduction, PCA, NMF, tSNE
- Clustering, K-means, hierarchical clustering
- Model limitations, Causality.

Type	Mode	Dates	Time	Frequency
Lecture	Asynchronous online	N/A	2 hours	weekly
Tutorial	Synchronous online	Wednesdays	1 hour	weekly

- Asynchronous pre-work must be completed prior to synchronous sessions
- A recording will be provided for synchronous sessions
- Closed captioning will be provided on audio or video recordings (if access to Microsoft Streams is available in your location)

The course has a synchronous and an asynchronous component. The asynchronous component includes the theoretical material taught by a team of lecturers across Western, from the data science committee. The synchronous part of the course will focus on solving any questions arising from the asynchronous portion (Friday at 4.30pm to 5.30pm). Full recordings will be provided for the synchronous portion, as well as pre-recorded tutorials in case you prefer that option.

### Table of Contents and Schedule



Week	Dates	Lecture	Lab	Assignment
1	Jan 11-15	Supervised Learning and Model Fitting	Pandas and Numpy, Optimization	Linear regression, Loss function Squared error Mean absolute deviation
2	Jan 18 – 22	Probability and Maximum Likelihood	Pandas	Regression
3	Jan 25 – 29	Introduce test set/out-of-sample idea. Classification, Evaluation, Logistic regression Regularization, Multi-class problems	Implementation of topics	Logistic regression, Classification, Evaluation
4	Feb 1 – 5	Estimating Performance, Quantifying Uncertainty	Bootstrap	Bootstrap and confidence intervals
5	Feb 8 – 12	Test error, Cross-validation, Model Selection, Bias-Variance tradeoff	Cross-validation	Cross-validation and model selection

6	Feb 15 – 19	Reading Week	N/A	
7	Feb 22 – 26	Feature Selection and Regularization (L1 and L2)	Regularization	Regularization and nested cross-validation
8	Mar 1 – 5	<b>Midterm</b>		-
9	Mar 8 –12	Trees, Random Forest	Tree Lab	Tree homework
10	Mar 15 – 19	Neural Networks, Gradients, learning	Simple 1-hidden layer network	3-layer network on non-linear problem
11	Mar 22 – 26	Autoencoders, Dimensionality reduction, PCA, NMF, t-SNE	Dimensionality reduction	Dimensionality reduction
12	Mar 29 – Apr 2	Clustering, K-means, hierarchical clustering	Clustering	Clustering
13	Apr 5 – 9	Model limitations, Causality.	Deploying models	N/A
14	Apr 12		Exam Review	

### Online Participation and Engagement



- Students are expected to participate and engage with content as much as possible
- Students can participate during synchronous sessions or post on OWL after watching the recording
- Students can also participate by interacting in the forums with their peers and instructors

## 4. Course Materials

### Core book:

*The Elements of Statistical Learning* by Hastie, Tibshirani and Friedman. [**Available for Free Online at <https://web.stanford.edu/~hastie/ElemStatLearn/>**]

### Recommended complementary book:

Machine Learning: a Probabilistic Perspective by P. Kevin Murphy [**Free online**]



Students should check OWL (<http://owl.uwo.ca>) on a regular basis for news and updates. This is the primary method by which information will be disseminated to all students in the class. Students are responsible for checking OWL on a regular basis.

All course material will be posted to OWL: <http://owl.uwo.ca>. Any changes will be indicated on the OWL site and discussed with the class.

If students need assistance, they can seek support on the [OWL Help page](#). Alternatively, they can contact the [Western Technology Services Helpdesk](#). They can be contacted by phone at 519-661-3800 or ext. 83800.

[Google Chrome](#) or [Mozilla Firefox](#) are the preferred browsers to optimally use OWL; update your browsers frequently. Students interested in evaluating their internet speed, please click [here](#).

## 5. Methods of Evaluation

The overall course grade will be calculated as listed below:

Weekly Assignments (10)	40%
Midterm Exam	25%
Final Exam	35%

### Weekly Assignment:

Assignments will be released each week, with due dates of the following week. There will be no make-up for missed weekly assignments. If the student submits a self-reported absence before the due date of the assignment, an extension of the deadline will be granted in line with the University policies.

### Midterm:

The midterm will be a practical examination in the form of a timed assignment. Students will be given a data set and a set of practical data analytic problems to solve, similar to the structure of the weekly assignments. The exam is “open book & open web”, meaning that students can access any notes or any documents on the web. Electronic communication with other people inside or outside class is prohibited.

### Final Exam:

The final exam will be scheduled by the Registrar. The final exam will cover concepts from the entire course and is in structure similar to the midterm exam. A grade of 50% or higher on the final examination is required to pass the course. The midterm will be a practical examination; each student will need a laptop to complete the midterm.

### Accommodated Evaluations

- Late assessments without illness self-reports will be subject to a late penalty discount of 10%/day (this means if your coursework gets an 80%, and you submit one day late, your final mark will be  $80\% - 10\% = 70\%$ ). The day late starts at the 00:00 of the day after the deadline posted above. There are **NO EXCEPTIONS** to this policy.
- Late assessments with illness self-reports should be submitted within 24 hours of submission of the last illness self-report
- An assessment cannot be submitted after it has been returned to the class. In case of a missed assignment with justified cause, the weight will be transferred to the final grade]
- If permission to waive the requirement that students receive evaluation on work totaling 15% of their final grade at least three days prior to the deadline for withdrawal without academic penalty has been obtained from the Dean’s Office, a statement to this effect must be made.

## **Rounding of Marks Statement**

Across the Sciences Undergraduate Education programs, we strive to maintain high standards that reflect the effort that both students and faculty put into the teaching and learning experience during this course. All students will be treated equally and evaluated based only on their actual achievement. **Final grades** on this course, irrespective of the number of decimal places used in marking individual assignments and tests, will be calculated to one decimal place and rounded to the nearest integer, e.g., 74.4 becomes 74, and 74.5 becomes 75. Marks WILL NOT be bumped to the next grade or GPA, e.g. a 79 will NOT be bumped up to an 80, an 84 WILL NOT be bumped up to an 85, etc. The mark attained is the mark you achieved, and the mark assigned; requests for mark “bumping” will be denied.

## **6. Accommodation and Accessibility**

### **Accommodation Policies**

Students with disabilities work with Accessible Education (formerly SSD) which provides recommendations for accommodation based on medical documentation or psychological and cognitive testing. The Academic Accommodation for Students with Disabilities policy can be found at:

[https://www.uwo.ca/univsec/pdf/academic\\_policies/appeals/Academic\\_Accommodation\\_disabilities.pdf](https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Accommodation_disabilities.pdf)

### **Academic Consideration for Student Absence**

Students will have up to two (2) opportunities during the regular academic year to use an on-line portal to self-report an absence during the semester, provided the following conditions are met: the absence is no more than 48 hours in duration, and the assessment for which consideration is being sought is worth 30% or less of the student’s final grade. Students are expected to contact their instructors within 24 hours of the end of the period of the self-reported absence, unless noted on the syllabus. Students are not able to use the self-reporting option in the following circumstances:

- for exams scheduled by the Office of the Registrar (e.g., December and April exams)
- absence of a duration greater than 48 hours,
- assessments worth more than 30% of the student’s final grade,
- if a student has already used the self-reporting portal twice during the academic year

If the conditions for a Self-Reported Absence are *not* met, students will need to provide a Student Medical Certificate if the absence is medical, or provide appropriate documentation if there are compassionate grounds for the absence in question. Students are encouraged to contact their Faculty academic counselling office to obtain more information about the relevant documentation.

Students should also note that individual instructors are not permitted to receive documentation directly from a student, whether in support of an application for consideration on medical grounds, or for other reasons. **All documentation required for absences that are not covered by the Self-Reported Absence Policy must be submitted to the Academic Counselling office of a student's Home Faculty.**

For policy on Academic Consideration for Student Absences - Undergraduate Students in First Entry Programs, see:

[https://www.uwo.ca/univsec/pdf/academic\\_policies/appeals/Academic\\_Consideration\\_for\\_absences.pdf](https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Consideration_for_absences.pdf)

and for the Student Medical Certificate (SMC), see:

[http://www.uwo.ca/univsec/pdf/academic\\_policies/appeals/medicalform.pdf](http://www.uwo.ca/univsec/pdf/academic_policies/appeals/medicalform.pdf)

### **Religious Accommodation**

Students should consult the University's list of recognized religious holidays, and should give reasonable notice in writing, prior to the holiday, to the Instructor and an Academic Counsellor if their course requirements will be affected by a religious observance. Additional information is given in the Western Multicultural Calendar:

<https://multiculturalcalendar.com/ecal/index.php?s=c-univwo>

You may also be eligible to write the Special Exam if you are in a "Multiple Exam Situation" (see [http://www.registrar.uwo.ca/examinations/exam\\_schedule.html](http://www.registrar.uwo.ca/examinations/exam_schedule.html)).

## **7. Academic Policies**

The website for Registrarial Services is <http://www.registrar.uwo.ca>.

In accordance with policy, <http://www.uwo.ca/its/identity/activatenonstudent.html>, the centrally administered e-mail account provided to students will be considered the individual's official university e-mail address. It is the responsibility of the account holder to ensure that e-mail received from the University at his/her official university address is attended to in a timely manner.

### **Contingency plan for an in-person class pivoting to 100% online learning**

In the event of a COVID-19 resurgence during the course that necessitates the course delivery moving away from face-to-face interaction, all remaining course content will be delivered entirely online, either synchronously (i.e., at the times indicated in the timetable) or asynchronously (e.g., posted on OWL for students to view at their convenience). The grading scheme will **not** change. Any remaining assessments will also be conducted online as determined by the course instructor.

### **All of the remote learning sessions for this course will be recorded.**

The data captured during these recordings may include your image, voice recordings, chat logs and personal identifiers (name displayed on the screen). The recordings will be used for educational purposes related to this course, including evaluations. The recordings may be disclosed to other individuals participating in the course for their private or group study purposes. Please contact the instructor if you have any concerns related to session recordings.

Participants in this course are not permitted to record the sessions, except where recording is an approved accommodation, or the participant has the prior written permission of the instructor.

**Scholastic offences** are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site:

[http://www.uwo.ca/univsec/pdf/academic\\_policies/appeals/scholastic\\_discipline\\_undergrad.pdf](http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf).

All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>).

Tests and examinations in this course will be conducted using Zoom. **You will be required to keep your camera on for the entire session, hold up your student card for identification purposes, and share your screen with the invigilator if asked to do so at any time during the exam.** The exam session will **not** be recorded.\*

More information about the use of Zoom for exam invigilation is available in the Online Proctoring Guidelines at the following link:

<https://www.uwo.ca/univsec/pdf/onlineproctorguidelines.pdf>

Completion of this course will require you to have a reliable internet connection and a device that meets the system requirements for Zoom. Information about the system requirements are available at the following link:

<https://support.zoom.us/hc/en-us>

\* Please note that Zoom servers are located outside Canada. If you would prefer to use only your first name or a nickname to login to Zoom, please provide this information to the instructor in advance of the test or examination.

## Professionalism & Privacy

Western students are expected to follow the [Student Code of Conduct](#). Additionally, the following expectations and professional conduct apply to this course:



- Students are expected to follow online etiquette expectations provided on OWL
- All course materials created by the instructor(s) are copyrighted and cannot be sold/shared
- Recordings are not permitted (audio or video) without explicit permission
- Permitted recordings are not to be distributed
- Students will be expected to take an academic integrity pledge before some assessments
- All recorded sessions will remain within the course site or unlisted if streamed

## Copyright Statement

Please be aware that all course materials created by the instructor(s) are copyrighted and cannot be **sold nor shared**. Those include materials used in tests/quizzes, midterms, and



finals. Any posting/sharing of such materials in part or whole without owner's consent is considered as violation of the Copyright Act and will be considered as a scholastic offence.

In addition, online services such as Chegg are actively monitored. Any questions that are coming out during midterms and finals and are posted to an online service will be searched. Such an activity will be considered as a scholastic offence and will result in academic penalty.

## 8. Professional Accreditation

This course is accredited under the Canadian Institute of Actuaries (CIA) University Accreditation Program (UAP) for the 2020-21 academic year. Achievement of the established exemption grade in this course may qualify a student from exemptions from writing certain preliminary exams.

Please note, a combination of this course and Statistics 3859A, is required to achieve an exemption for preliminary exam SRM (minimum of 80% in each course is required). Please see the following link for full details:

<http://www.cia-ica.ca/membership/university-accreditation-program---home>

NOTE: This course is currently being reviewed by the Canadian Institute of Actuaries (CIA) as a possible course that is accredited under the University Accreditation Program (UAP) for the 2020-21 academic year. A decision will not be known until later in the semester about whether it will be accepted (along with Statistics 3859A) as an exemption for the preliminary SRM exam (minimum of 80% in each course is required). The decision of the CIA will be communicated to students later in the semester.

## 9. Support Services

Please visit the Science & Basic Medical Sciences Academic Counselling webpage for information on add/drop courses, academic considerations for absences, appeals, exam conflicts, and many other academic related matters: <https://www.uwo.ca/sci/counselling/>

Please contact the course instructor if you require lecture or printed material in an alternate format or if any other arrangements can make this course more accessible to you. You may also wish to contact Student Accessibility Services (SAS) at (519) 661-2147 if you have any questions regarding accommodations.

Western University is committed to a thriving campus as we deliver our courses in the mixed model of both virtual and face-to-face formats. We encourage you to check out the Digital Student Experience website to manage your academics and well-being: <https://www.uwo.ca/se/digital/>.

Learning-skills counsellors at the Student Development Centre (<http://www.sdc.uwo.ca>) are ready to help you improve your learning skills. They offer presentations on strategies for improving time management, multiple-choice exam preparation/writing, textbook reading, and more. Individual support is offered throughout the Fall/Winter terms in the drop-in Learning Help Centre, and year-round through individual counselling.

Students who are in emotional/mental distress should refer to Mental Health@Western ([http://www.health.uwo.ca/mental\\_health](http://www.health.uwo.ca/mental_health)) for a complete list of options about how to obtain help.

Additional student-run support services are offered by the USC, <http://westernusc.ca/services>.