

IS2002B/DS2000B: Introduction to Data Science

Description:

Introduction to basic concepts and tools used to analyze large and complex data sets, enabling scientists to identify relevant information, while being able to quantify the certainty of their conclusions. The class is centered on three basic concepts: sampling as the process of making inferences about a population, random assignment and experiments to be able to draw causal inferences, and choosing adequate models to be able to make good predictions. For each concept, the class introduces a related technique (bootstrap, randomization test, and cross-validation), providing the students with the toolkit to tackle even difficult and non-standard statistical problems. The class is built on the insight that “you only understand it if you can do it”. Therefore, the main emphasis of the class is to apply the concepts and techniques in the homework assignments, midterm, and final, all of which are centered on real data analysis problems. Students learn to program in Python, use data analysis toolboxes, and Jupyter Notebooks, equipping them with valuable real-world data-analysis skills.

Class time:

Winter term, 2022:^[L]_[SEP]

Lecture: Thursday, 10:30-12:30pm^[L]_[SEP], B&GS-0165

In case of pivoting to online learning, the lectures will be delivered on Zoom synchronously (i.e., at the times indicated in the timetable) and asynchronously (posted on OWL after the lecture). The Zoom link will be provided on OWL.

Labs:

Friday: 12:30-2:30 HSB-11 (or online, see above)

Friday: 2:30-4:30 HSB-9 (or online, see above)

Instructors:

Professor: Daria Proklova

^[L]_[SEP] Email: dproklov@uwo.ca

Office: WIRB 4154 (Mondays 2-3 pm – by appointment, zoom option available)

TA: Ladan Shahshahani

Email: lshahsha@uwo.ca

Office: WIRB 4th floor (Mondays 11am-12 – by appointment, zoom option available)

TA: Duo Xu

Email: dxu258@uwo.ca

Office: MC 275K (Mondays 11am-12 – by appointment, zoom option available)

Prerequisites:

1.0 courses from Mathematics, Calculus, or Applied Mathematics (numbered 1000 and higher) with a minimum mark of 60%. Data Science 1000A/B (with a minimum grade of 60%) can be used to meet 0.5 of the 1.0 mathematics course requirements.

Antirequisite(s): Computer Science 2034A/B, Computer Science 2035A/B.

Course content:

Week starting with	Lecture concepts	Tutorial / Homework
01/13	Intro to class Data Basics (O1.2) Examining numerical data (O1.6) Mean, Median, Mode Standard deviation	HW1 Python and Pandas basics Pandas basics Histograms Descriptive Statistics
01/20	Examining categorical data (O1.7) Probability (O2.1-O2.2, O2.4)	HW2 Plotting and analyzing categorical data Probabilities Descriptive by Category
01/27	Estimation, Confidence Intervals, Bootstrap Types of studies (O1.4, 1.5)	HW3 Bootstrap and confidence intervals
02/03	Foundations of Inference (R2) Hypothesis testing	HW4 Randomization Test
02/10	Chi-square test of independence. Monte-carlo simulations, Decision errors and Power	HW5 More Randomization test and Monte Carlo simulation
02/17	Bayes rule Bayesian Inference	HW6 Bayesian inference + Midterm practice
02/24	Reading week	
03/03	Midterm (open-book, takes place during the Thursday lecture)	
03/10	Simple linear regression: Cost function and optimization parameter estimation, Outliers (O7.3-O7.4)	HW7 Outlier detection and exclusion Optimization of squared error cost function

03/17	Outliers and robust techniques, Polynomial regression, Cross-validation	HW8 Optimization of Mean-absolute deviation cost function (Median regression) Polynomial regression Crossvalidation
03/24	Multiple regression: Model comparison and correlation of regressors	HW9 Multiple regression Bootstrap with multiple regression Co-variance of estimators
03/31	Discrete variables and logistic regression	HW10 Regression with discrete variables: dummy coding Logistic regression
04/07	Model selection Stepwise regression Summary	HW11 Final practice
determined by Exam Central	Final (open-book)	

Course Materials

References

Python for Data Analysis: (P)

<https://www.amazon.ca/Python-Data-Analysis-Wrangling-IPython/dp/1491957662>

Open Intro Statistics (O):

https://www.openintro.org/stat/textbook.php?stat_book=os

Open Intro Stat with Randomization and Simulation (S):

https://www.openintro.org/stat/textbook.php?stat_book=isrs

Equipment (in case of pivoting to online learning)

- a laptop or computer
- a working microphone and webcam
- a reliable internet connection

Evaluation:

Weekly Assignments: 33%

Each assignment accounts for 3.3%. A self-reported absence (submitted before the due date) will allow you a 48hr extension on the submission deadline – but you are still required to submit the assignment for the week. **There will be no makeup for assignments that are turned in late, the 3.3% individual assignment weight can not be redistributed among the remaining assignments.**

Midterm: 31%

Final: 36%

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 synchronously (i.e., at the times indicated in the timetable) and 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.

When deemed necessary, tests and examinations in this course will be conducted using a remote proctoring service. By taking this course, you are consenting to the use of this software and acknowledging that you will be required to provide personal information (including some biometric data) and that the session will be recorded. Completion of this course will require you to have a reliable internet connection and a device that meets the technical requirements for this service. More information about this remote proctoring service, including technical requirements, is available on Western's Remote Proctoring website at: <https://remoteproctoring.uwo.ca>.

Weekly assignments:

New assignments will be posted on OWL Thursday morning latest. *You should start on the assignment before the corresponding lab section on Friday, so you can ask appropriate questions.* Assignments are due on Tuesday 4pm (EST), with submission on OWL. Late assignments will be accepted until Wednesday 4pm (EST), but will be awarded only 50% of the achieved score. The solutions to the homework will be posted Wednesday at 4pm, unless somebody reported an absence. Any homework not submitted by the due date will be scored with 0%. We will strive to provide grading of the homework on Thursday latest.

Accommodation and Accessibility

If you are unable to meet a course requirement due to illness or other serious circumstances, you must provide valid medical or supporting documentation to the Academic Counselling Office of your home faculty as soon as possible. If you are a Science student, the Academic Counselling Office of the Faculty of Science is located in WSC 140, and can be contacted at scibmsac@uwo.ca.

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

https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic_Accommodation_disabilities.pdf

For further information, please consult the university's medical illness policy at http://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_medical.pdf.

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.

Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at this website: http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf.

All required homeworks 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>).

Support Services

Please visit the Science & Basic Medical Sciences Academic Counselling webpage for information on adding/dropping 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 Accessible Education at (519) 661-2147 if you have any questions regarding accommodations.

Students who are in emotional/mental distress should refer to Mental Health@Western (<http://www.health.uwo.ca/mentalhealth>) 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>.

The policy on Accommodation for Students with Disabilities can be found here:
www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_disabilities.pdf

The policy on Accommodation for Religious Holidays can be found here:
http://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_religious.pdf