New Course

Proposal Reference Number : 23460
PRN Alias : 22-23#1311
Version No : 4
Submitted By : Ms Angela White
Edited By : Ms Angela White

Display Printable PDF

	New Data
Program Affected?	Y
Program Change Form Submitted?	Y
Subject/Course/Term	MATH 527 • two consecutive terms (D1, D2)
Credit Weight or CEU's	6 credits
Course Activities	**Schedule Type** / **Hours per week** A - Lecture / 3 Total Hours per Week : 3 Total Number of Weeks : 13
Course Title	**Official Course Title :** Stat. Data Science Practicum **Course Title in Calendar :** Statistical Data Science Practicum
Rationale	This course will offer a unique experience to senior BA and BSc statistics students as well as to graduate students in statistics. It will fill an important gap in our undergraduate and graduate programs in that it will allow students to learn how put their theoretical skills into practice as well as to develop their teamwork and communication skills. All these abilities are key to students' future professional success. In the A.Stat. accreditation of our programs, the course would count towards the communications module. The course is being developed by the Director of Computational and Data Systems Initiative (CDSI), Professor Kolaczyk; he designed and very successfully run a similar course at Boston University.
Responsible Instructor	Eric Kolaczyk
Course Description	Development of the holistic skills required for doing statistical data science in practice. Data science life cycle from a statistics-centric perspective and from the perspective of a statistician working in the larger data science environment. Group-based projects with industry, government, or university partners. Statistical collaboration and consulting conducted in coordination with the Data Science Solutions Hub (DaS^2H) of the Computational and Data Systems Initiative (CDSI).
Teaching Dept.	0290 : Mathematics and Statistics
Administering Faculty/Unit	SC : Faculty of Science

Prerequisites	MATH 324 Statistics; MATH 223 Linear Algebra; MATH 208 Intro. to Statistical Computing. Web Registration Blocked? : N
Corequisites	MATH 423 Applied Regression or equivalent. Web Registration Blocked? : N
Restrictions	
Supplementary Calendar Info	
Additional Course Charges	
Campus	Downtown
Projected Enrollment	20
Requires Resources Not Currently Available	N
Explanation for Required Resources	
Required Text/Resources Sent To Library?	
Library Consulted About Availability of Resources?	
Consultation Reports Attached?	
Effective Term of Implementation	202309
File Attachments	• MATH 527 Syllabus _final version.pdf View • MATH 527 Simple Program Changes.pdf View
To be completed by the Faculty	
For Continuing Studies Use	

Approvals Summary

Show all comments

Version No.	Departmental Curriculum Committee	Departmental Meeting	Departmental Chair	Other Faculty	Curric/Academic Committee	Faculty	SCTP	Version Status
4					**Approved** Jing Xu Meeting Date: Jan 10 2023 Approval Date: Jan 11 2023 View Comments			**Approved by Curric/Academic Committee** Edited by: Angela White on: Dec 20 2022
3								**Approved by Department Meeting** Edited by: Angela White on: Dec 20 2022

2		**Approved** Djivede Kelome Meeting Date: Dec 05 2022 Approval Date: Dec 16 2022 View Comments						**Approved by Department Meeting** Edited by: Angela White on: Dec 16 2022
1		**Rejected** Djivede Kelome Meeting Date: Dec 05 2022 View Comments						**Rejected by Department Meeting** Created on: Dec 14 2022

December 14, 2022

Faculty: Faculty of Science

Unit: Department of Mathematics and Statistics

MATH 527 Statistical Data Science Practicum; PRN#23460

Simple changes to programs: Yes

Please add this new course to the following program:

1) B.Sc. Honours Statistics and Computer Science -under Complementary Courses-"At least 8 credits selected from."-listed in numerical order.

2) B.A Honours Probability and Statistics - under Complementary Courses -"Part II: 6-11 credits in mathematics and computer science selected from"-listed in numerical order.

3) B.Sc. Honours Probability and Statistics - under Complementary Courses -"Part II: 6-11 credits in mathematics and computer science selected from"-listed in numerical order.

Note: There will be complex changes to 5 programs. 5 Program Revisions forms will be submitted.

Complex program revisions for the following programs:
B.A Major Concentration Statistics
B.A Supplementary Minor Concentration Statistics
B.Sc. Major Statistics
B.Sc. Liberal Program - Core Science Component Statistics
B.Sc. Joint Major Statistics and Computer Science

Statistical Data Science Practicum[1]
MATH 527D1/D2

Course Summary: This is a 6-credit, two-semester (Fall/Winter) course aimed at developing the holistic skills required for doing statistical data science in practice. This goal will be accomplished through an emphasis on statistical collaboration and consulting throughout, conducted in coordination with the Data Science Solutions Hub (DaS^{2}H) of the Computational and Data Systems Initiative (CDSI), layered on top of a pedagogical backbone informed by the data science life cycle.

Staff:

- Course Instructor: Prof. Eric Kolaczyk, Burnside 917, eric.kolaczyk@mcgill.ca
- Co-Instructor(s) (at scale, this course requires team teaching)
- Graduate student TA's (operate as leads for ~15 students each, in teams of 3-4)

Class Meetings: Twice per week for 80 minutes each.

Prerequisites: MATH 324 (Statistics); MATH 223 (Linear Algebra); MATH 208 (Intro to Statistical Computing)

Corequisites: MATH 423 (Applied Regression) or equivalent.

Organization and Flow of the Course: The central organizing principle of the course is the data science life cycle, encountered first in the fall semester from a statistics-centric perspective and revisited in the winter semester from the perspective of a statistician working in the larger data science environment. We will use a flipped classroom, requiring pre-class preparation for in-class discussions and group work. Class meetings will be of two types weekly. One class will focus mainly on reading/discussion/active learning, in parallel with group-based projects with industry/government/university partners. As we move from one topic to another throughout each semester, you will regularly be asked to put into practice what you're learning using your project. The other class will focus on statistical consulting and collaboration, working on problems brought to us via the CDSI Data Science Solutions Hub (DaS^{2}H), where the nature of the problems and the solutions required will be more fluid in nature.

Learning Outcomes: Students can expect to learn the principles and rationale behind, as well as the integrated practice of, the following topics and skills: project lifecycle and management; communication; data wrangling; exploratory data analysis and visualization; modeling, inference, and prediction; (mis)interpretation of results; reporting and archiving.

Assessment: Student work will be evaluated in this course based on four factors – reading quizzes, homework assignments, project work, and consulting work.

---

[1] This course is an adaptation of the MA675-676 Statistics Practicum sequence and concept piloted by the instructor previously at Boston University, in his role as founding director of the MS in Statistical Practice program there. A comprehensive description of the practicum concept, and its place and potential within the larger data science pedagogy, can be found in the following article: Kolaczyk, E. D., Wright, H., & Yajima, M. (2021). Statistics Practicum: Placing "Practice" at the Center of Data Science Education (with discussion). *Harvard Data Science Review*, *3*(1). https://doi.org/10.1162/99608f92.2d65fc70

● *Reading Quizzes (25%)*: Most classes will have required reading, accompanied by an on-line reading quiz that must be completed prior to class.
● *Homework Assignments (15%):* Homework will be assigned periodically throughout the semester, particularly during the first half of the semester.
● *Project Work (25%):* While any consulting project can be expected to touch on at least *some* of the elements of the data science life cycle, student work on a curated industry/government-partner project (or, in the absence of such, reconstituted case studies) will ensure at least one project whose practical application ties to *all* elements. These will be semester/year-long, team-based projects with timing closely connected to that of the reading and pedagogical side of the course.
● *Consulting Work (35%):* Each student can expect to be involved in teams with several consulting projects throughout the year.

Resources: There is no formal textbook for this course. Rather, the "text" will be the classroom, combined with various select articles for reading, and the learning will be largely experience-based.

Why Two Semesters? The proposed two-semester nature of this course, while unusual in Faculty of Science (although less so in certain other parts of the university), is important for student success. Two key reasons: (i) work on real-world projects with McGill researchers through DaS^{2}H (being a key value-added of this Practicum approach -- as opposed to using "canned" projects) frequently does not break down to fall cleanly within a semester but can be managed to largely do so over the course of an academic year; and (ii) students can expect to achieve a certain level of maturity and self-sufficiency in doing this type of work if they see multiple projects (e.g., 2-3), but a semester is too short to guarantee that they will get that critical level of repetition.

Additional Notes for Course Proposal Approval Process:

1. Since work in projects and consulting will be explicitly (and necessarily) managed to be successful, it will be evaluated using a self-, peer-, and co-assessment system[2] with a detailed rubric, iterated and/or adjusted for subjectivity, and implemented through the Eduflow software platform.

2. At 3 credits per semester, corresponding to 9hrs per week (including contact hours), the expected distribution of student time is as follows: (i) Twice per week full-class meeting @80min per meeting, or ~3hrs/week; (ii) once per week consulting project management meeting with TA in discussion section format, @1hr/week; (iii) standing consulting hours, facilitating scheduling of clients to fixed time slots, @1-2hr/week; and (iv) the remaining 3-4hrs/week working on reading and homework earlier in the semester, and then increasingly consulting and partner projects as the semester progresses.

3. The student:TA ratio proposed here is lower than typical in the Faculty of Science, but found appropriate through experience, in light of the mentee:mentor nature of the role. It is expected that CDSI and Science will reach an understanding for an appropriate funding model.

4. Several visualizations can be helpful in communicating how the course unfolds over a semester. At a high level, the pedagogical backbone lent by the data science life cycle is depicted in Figure 1 below, as a function of topic and week, for the first semester. The second semester is organized similarly, but with a shift from the statistics-centric perspective of the first semester to a broader

[2] This approach has roots in Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, *24*(3), 331–350. https://doi.org/10.1080/03075079912331379935

data science perspective (e.g., big data, predictive analytics, etc.) and without the modules on communication.

Figure 1: Weekly topics for MATH 527, as a function of key elements of the data science life cycle. [Taken from Kolaczyk, Wright, and Yajima (2021).]

Second, Figure 2 below shows the distribution of consulting and project elements of the course from a student perspective, in the form of a timeline.

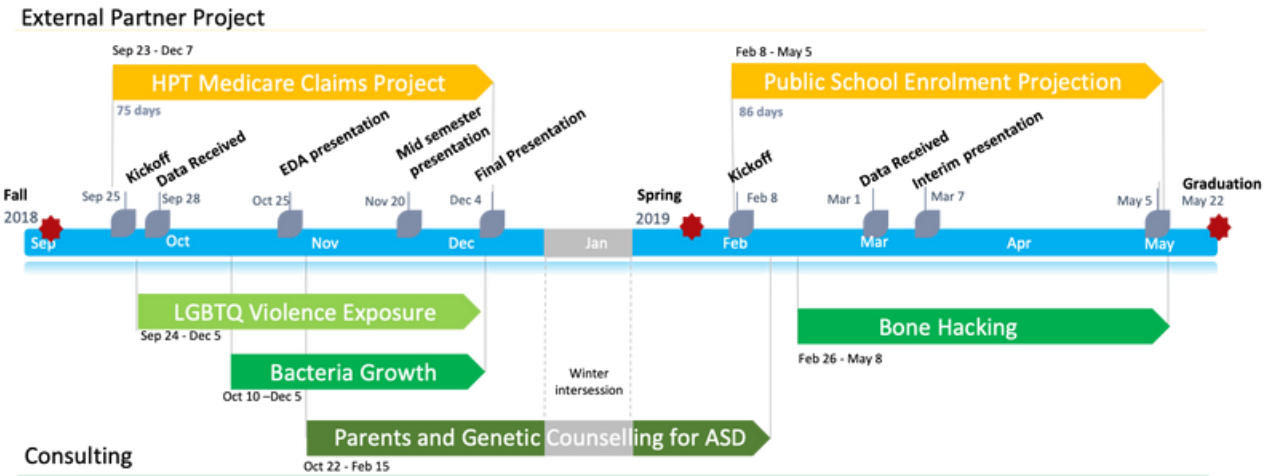

Figure 2: Visual depiction (in the form a two-semester timeline) of project efforts for typical student, based on real experiences at Boston University. [Taken from Kolaczyk, Wright, and Yajima (2021).]

Finally, a week-to-week schedule is shown in the table below, adapted from the Fall 2019 offering of the precedent of the proposed course at Boston University, illustrating how everything is interwoven week to week throughout the semester. Note that the "Thursday" class of the first month is spent preparing consulting teams. Consulting opens Oct 1, after which this class each week is devoted to mini-presentations of consulting projects by student teams as a function of several milestones (typically (i) an intro to the project upon in-take; (ii) an update(s) as key milestones are hit (e.g., an initial exploration of the data); and (iii) a final presentation upon close of project).

ILLUSTRATIVE SYLLABUS

Week	Topic	Reading	Classroom Focus	Discussion Section Focus	Due	Assigned
Week 1 (Tues, Sept 3)	What Is Statistics?	Three readings, on (i) statistics, (ii) consulting, and (iii) data science.	Discussion: Career goals & the role of "Practice".	Setting up the necessary software	None	Reflection on career goals.
Week 1 (Thu, Sept 5)	The Project Lifecycle	Two readings, on (i) project management lifecycle and (ii) reproducibility.	Consulting Overview	N/A	None	None
Week 2 (Tue, Sept 10)	Matching Questions with Analyses	Two readings, on (i) what is the question, and (ii) exploratory data analysis.	Introduction to Partner Projects	Nuts & Bolts of Consulting	"Revised" Resumes	"Why You Should …!" Talks (5min)
Week 2 (Thu, Sept 12)	Data: Types and Sources	Three readings, on (i) data collection, (ii) structured vs unstructured data, and (iii) data mining.	Consulting Role Playing	N/A	None	None
Week 3 (Tue, Sept 17)	Communication: Speaking & Emails	Two readings, on (i) good oral communication, and (ii) emails.	Class presentations on "Why You Should …!"	Case studies with past consulting projects	"Why You Should …!" Talks	None
Week 3 (Thu, Sept 19)	Communication: Writing	Two readings on writing.	Real-time Writing Exercise	N/A		Writing Assignment
Week 4 (Tue, Sept 24)	Data Provenance, Cleaning, & Manipulation	One reading on data cleaning; video on data privacy and confidentiality.	Discussion of Project Data.	Mini-consulting analysis exercises	Writing Assignment	Clean and prepare project data.
Week 5 (Tue, Oct 1)	Data Visualization	Three readings on data visualization "do"s and "don't"s.	Visualization of Project Data	Project management of ongoing consulting		Project EDA
Week 6 (Tue, Oct 8)	Statistical Modeling & Inference	One reading on statistical modeling.	Planning EDA for Projects	Project management of ongoing consulting	None	None

Week 7 (Tue, Oct 15)	Statistical Modeling & Inference (cont)	One reading on statistical modeling.	Refining EDA for Projects	Project management of ongoing consulting	None	None
Week 8 (Tue, Oct 22)	Interim Presentations	No Readings	Presentations	Project management of ongoing consulting	Project EDA	None
Week 9 (Tue, Oct 29)	Interim Presentations	No Readings	Presentations		None	Modeling & Inference for Project Data
Week 10 (Tue, Nov 5)	(Mis)interpretation of Statistical Inferences	*The ASA's Statement on Pvalues* and related reading.	Partner Projects	Project management of ongoing consulting	None	None
Week 11 (Tue, Nov 12)	Assessing and reporting statistical results.	Three readings: (i) why research findings can be false; (ii) ethics in reporting research results; and (iii) intellectual property.	Partner Projects	Project management of ongoing consulting	None	
Week 12 (Tues, Nov 19)	Interim Presentations	No Readings	Planning Project Analysis Writeup	Project management of ongoing consulting	None	Project Analysis Writeup
Week 13 (Tues, Nov 26)	Interim Presentations	No Readings	Planning Project Final Presentations	Project management of ongoing consulting	None	Project Final Presentation
Week 14 (Tue, Dec 3)	Presentation of Projects (Dry-run)	No Readings	In-class Presentation	Project management of ongoing consulting	Final Presentation (Dry-run)	None
Week 15 (Tue, Dec 10)	Presentation of Projects	No Readings	In-class Presentation	Project management of ongoing consulting	Final Presentation & Analysis Writeup	None