Description: Statistical Analysis and Inference (3 credits). The goal of this course is to provide basic notions of statistical analysis and inference as well as advanced statistical methods. In addition to the theoretical concepts, this course will focus on the practical applications of these methods.

Instructor: Léo Belzile (✉ `leo.belzile@hec.ca`).

Lectures: Wednesday, 15:30–18:30 on Teams/Zoom.

- Students are expected to install SAS and view the online recording by Week 2.

Course content:

1. Basic principles in inference and statistical modeling
2. Linear models
3. Generalized linear models
4. Models for longitudinal data and correlated data
5. Introduction to survival analysis

Target audience: The course is mandatory for the "Business Analytics and Data Science" M.Sc. specialization.

Incompatible course: MATH60619(A): you cannot register for this course if you are enrolled or have completed MATH60619(A), or if MATH60619(A) is part of your study plan.

Learning objectives: at the end of the course, the student will be capable of

- understanding fundamentals of statistical inference and analysis as well as advanced statistical methods used in data science;
- implementing these methods using the SAS software;
- to set-up the statistical methods within the context of data science.

Coursework: written work must be submitted as PDF online on *ZoneCours* alongside with the SAS code used to perform the statistical analyses (as `.txt` files).

Grading scheme: your grade will be based on

- Assignment 1, worth 15% and due October 14th.
- Assignment 2, worth 15% and due November 11th.
- Assignment 3, worth 15% and due December 2nd.
- Data analysis project (teams of three or four), worth 25%, due Wednesday, December 16th.
- Final examination (individual), worth 30% of the final grade.

Team work counts toward your final grade only if you obtain more than 50% on the individual evaluations. Otherwise, the final grade is calculated on a pro rata basis using only individual evaluations.

References: The course notes and slides are available online on the course website. Additional bibliographic references are provided on *ZoneCours*.

Tentative schedule:

Course	Date	Content
1	Sept. 2	Introduction (Chapter 1)
2	Sept. 9	Linear models (Chapter 2)
3	Sept. 16	—
4	Sept. 23	—
5	Sept. 30	—
6	Oct. 7	Likelihood-based methods (Chapter 3)
7	Oct. 14	Generalized linear models (Chapter 4)
	Oct. 21	Fall recess
8	Oct. 28	—
9	Nov. 4	Correlated and longitudinal data (Chapter 5)
10	Nov. 11	—
11	Nov. 18	Mixed effects models (Chapter 6)
12	Nov. 25	—
13	Dec. 2	Introduction to survival analysis