Statistical modelling

Syllabus

Dr. Léo Belzile
HEC Montréal

Welcome

Léo Belzile

- Assistant professor of Decision Sciences
- PhD degree from EPFL
- I study extreme value analysis (floods, heatwaves, etc.)

Organization

- Weekly meetings via Zoom (Wednesday, 15:30-18:30 Montreal time).
- All of the course material can be downloaded from the course website: `https://lbelzile.github.io/statmod/`
  - course notes
  - video records
  - R package with datasets and quiz
  - exercises and solutions
  - SAS/R demos
- Submit your assignments/projects via *ZoneCours*
- Ask course-related questions on Piazza: `piazza.com/hec.ca/fall2020/math60604a`
- For other matters, I am best reached by *email*: `leo.belzile@hec.ca`

Course content

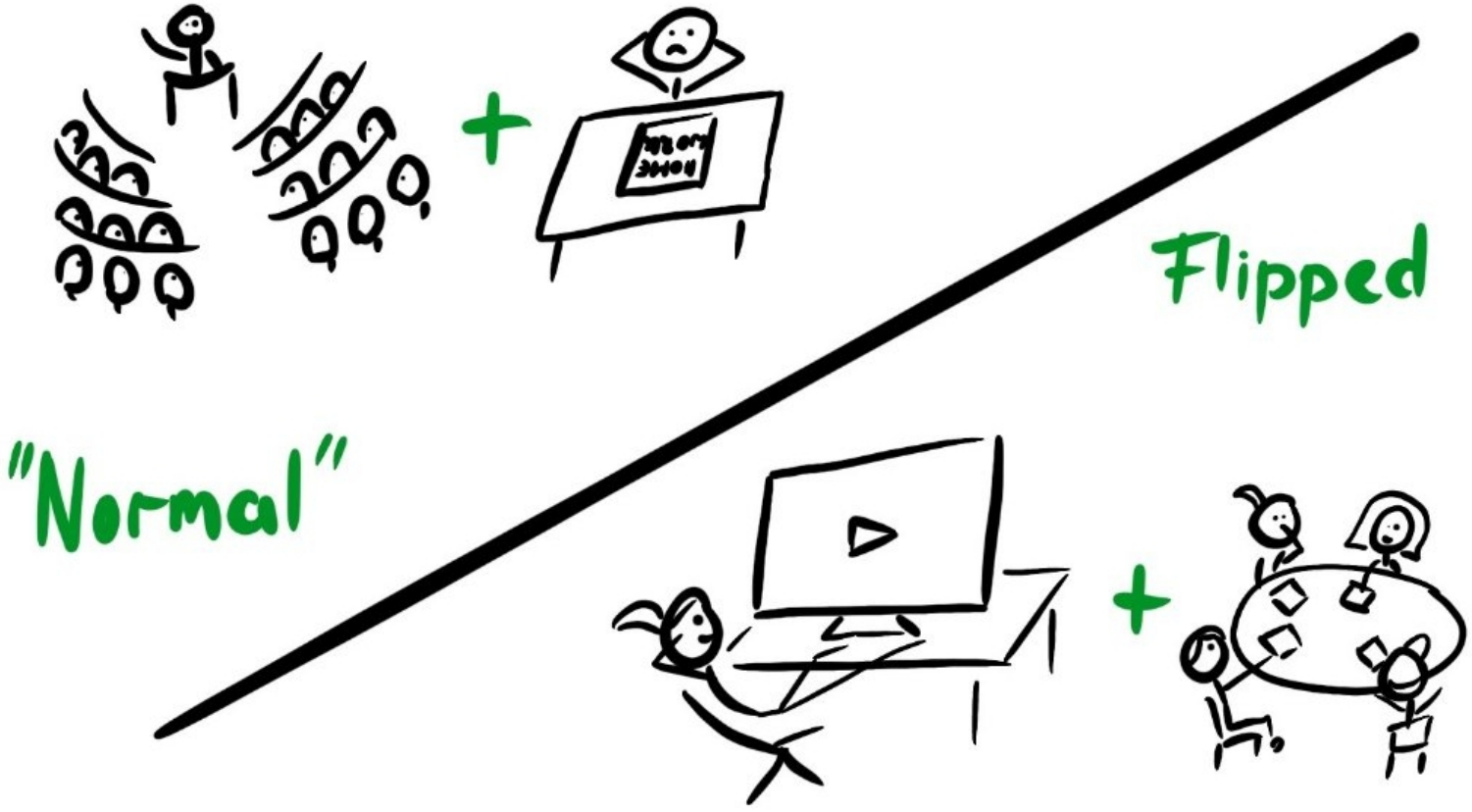
> *All models are wrong, but some are useful.*
>
> — Georges Box

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

Grading

- Three assignments, each worth 15% (partly in teams)
  - short questions (multiple choices)
  - programming questions
  - data analysis
- Data analysis project (25%)
  - create your own assignment!
- Final exam (30%, format TBD depending on HEC directives)

Flipped classroom



Flipped class (drawing by Heidi Seibold, under CC-BY 4.0 license)

What is the format of the course?

At home:

- reading course notes
- slides with videos
- quiz
- exercises (with solutions)

In class:

- weekly summary
- question period
- group workshop and discussions

What is the workload for this course?

- 3 credits = 135 hours of work
- an average of 9 hours per week
- do not underestimate the initial time investment:
  - installing required software
  - learning programming basics
  - getting up to speed with prerequisites

What is the target audience?

Students enrolled in the Data Science and Business Analytics M.Sc program.

Students admitted normally have a bachelor in

- engineering
- physics or
- mathematics.

Basic knowledge of calculus and linear algebra is assumed.

What are the prerequisites?

A first course in probability/statistic covering the following notions:

- probability axioms and combinatorics
- random variables
  - moments (expectation, variance, correlation)
  - discrete distributions: Bernoulli, binomial, Poisson
  - continuous distributions: uniform, exponential, normal
- descriptive statistics
- hypothesis tests
- comparison of means and proportions (one and two samples)
- simple linear regression

Will there be programming?

Yes. We will cover the basics of SAS (and R) to fit models and visualized data.

- I will provide code only for exercises and slides.

You must provide your code otherwise for assignments/group projects

- I should be able to reproduce *exactly* your analyses.
- submit as a `.txt` file (otherwise, you won't be able to submit your work on *Zonecours*)
- use UTF8 encoding
- follow the instructions for naming scripts/files (hint: $\neq$ `mycode.sas`)

What software will we use in class?

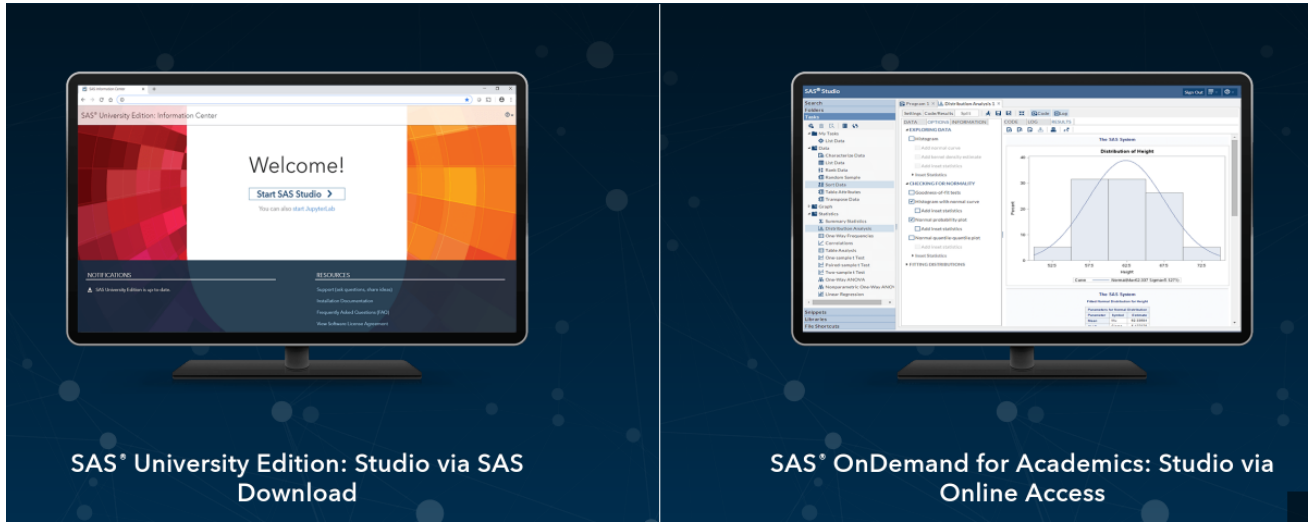
According to the official course outline, SAS (Base/Stat). A tutorial (video recorded) will teach you the basics.

- This is the only course in your program that uses this programming language. Why bother with it, then?
  - comparative advantage if you know multiple programming languages.
  - frequently used by business
    - they maintain databases with SAS
    - the software costs a lot of money ($ $ $)
    - stable, the company offers support (unlike open source)
    - code legacy blocks evolution towards modern languages (Python, R, etc.)

We will work with SAS® OnDemand for Academics: Studio.

- [No need to install! works out of the box + cross-platform support]

Alternatively, you could install SAS University Edition, but this service is being discontinued.



IT services at HEC provide a copy of the software via their platform for Windows (free license, but they charge ~35$ for the download).

I am avid support of open-access software and of R, a programming language written by the community

- its free!
- multi-platform support
- download from cran.r-project.org
- used in the course notes
- I recommand the IDE RStudio

What are the prof expectations?

- Active participation in class: students are expected to be in class (even virtually)
  - if this is impossible, let me know
  - small-group discussions
  - ask questions! there is no silly question
- Autonomy: you are sole responsible for your learning.
  - stay up to date and do your readings
  - don't stay in the dark: ask questions (to me or your peers)!
- Feedback: problems or unclear explanations? let me know asap
  - online semester = friction points (internet connection, lack of human contact, etc.)
  - many tools = longer period to adapt to uncharted territory

Inclusive and respectful environment

Even if the class is virtual, be respectful. Harassment, discriminatory views, etc. are not tolerated.

Let me know if

- I can do something to improve the course experience for you or other students
- a statement or attitude makes you uncomfortable
- your name/prefered pronoun/gender differs from the information provided on *HEC en ligne*
- your performance is affected by external factors: I will do my past to help you or direct you to external resources.

Plagiarism

Please don't. There are consequences and its an insult to your intelligence. Ask for help if necessary!

- if you take and adapt code from elsewhere (e.g., `StackOverflow`), cite your sources!
- you must program yourself your code for individual assignments (discussion with peers is okay, but code sharing, copy paste or similar wording is punishable)

Plagiarism

Among potential offences, are (from UofT?):

- Working together to answer questions.
- Looking at someone else's answers.
- Letting someone else look at your answers.
- Sharing or posting the exam questions.
- Discussing answers or the exam questions with anyone else in or outside the course.
- Misrepresenting your identity or having someone else complete your exam.