

# Modélisation statistique

## Plan de cours

**Dr. Léo Belzile**  
**HEC Montréal**

# Bienvenue



**Léo Belzile**

- + Professeur adjoint, Sciences de la décision
- + Spécialité: modélisation d'événements rares

# Organisation

- + Rencontres hebdomadaires via **Zoom** les lundi 8h30-11h30.
- + Tout le matériel est disponible via le site web du cours:  
<https://lbelzile.github.io/modstat/>
  - + notes de cours
  - + capsules vidéo
  - + quiz
  - + exercices et solutionnaires
  - + démo **SAS/R**
- + Soumissions de travaux via *ZoneCours*
- + Publiez vos questions de cours sur **Piazza**:  
[piazza.com/hec.ca/fall2020/math60604](https://piazza.com/hec.ca/fall2020/math60604)
- + Pour d'autres questions, je suis joignable par *courriel*:  
[leo.belzile@hec.ca](mailto:leo.belzile@hec.ca)

# Contenu du cours

*Tous les modèles sont faux, mais certains sont utiles.*

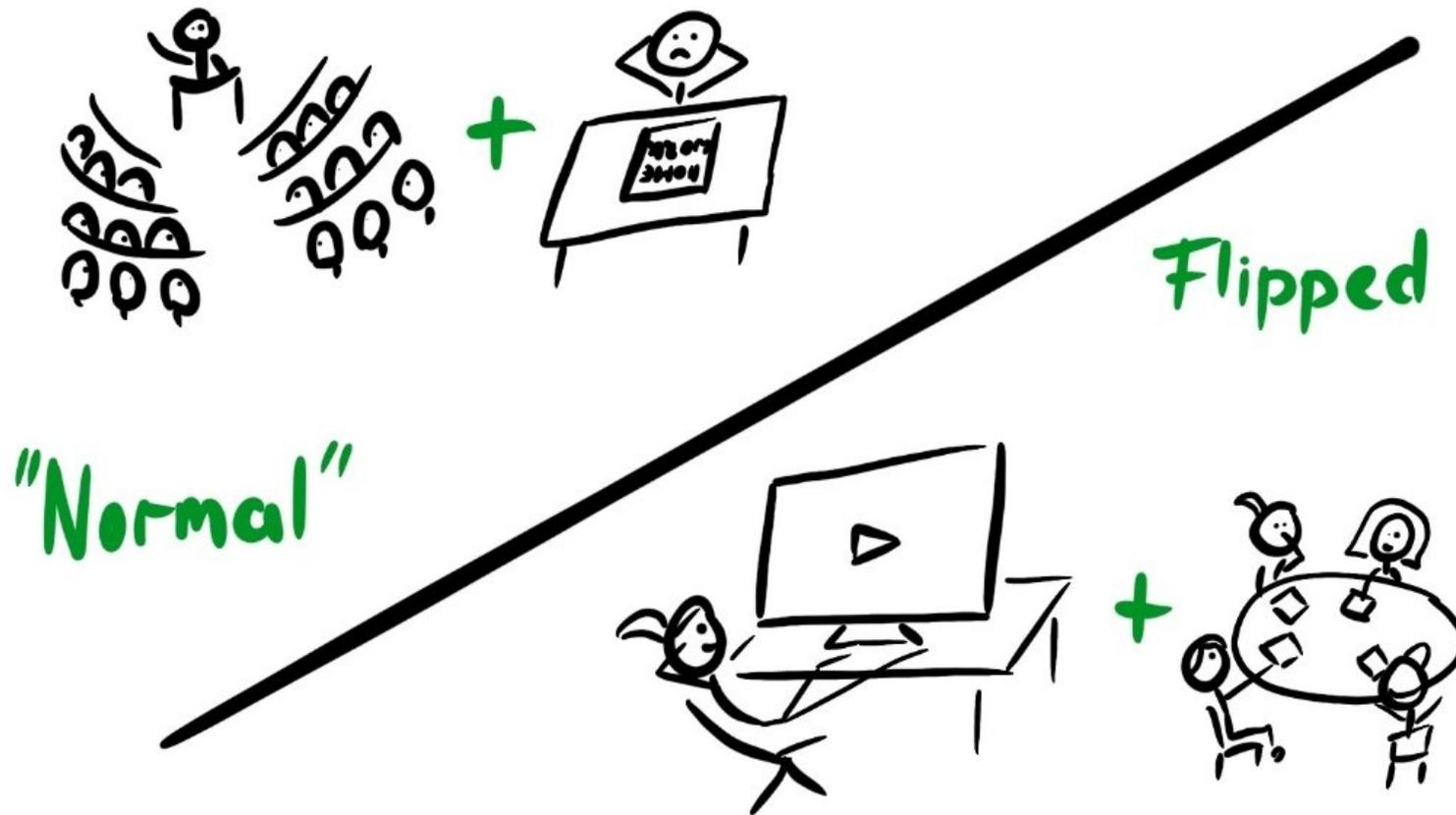
— Georges Box

1. Principes fondamentaux de modélisation et d'inférence statistique
2. Modèles linéaires
3. Modèles linéaires généralisés
4. Modèles pour données longitudinales et corrélées
5. Modèles linéaires mixtes
6. Introduction à l'analyse de survie

# Évaluations

- + Trois devoirs, chacun 15% (parties en équipes?)
  - + questions courtes (choix multiples)
  - + questions à développement avec programmation
  - + analyse de données
- + Un projet d'analyse de données (25%)
  - + créez un devoir!
- + Examen final (30%, format à déterminer selon les directives de HEC)

# Innovation pédagogique, version pandémie!



Classe inversée (illustration de Heidi Seibold, sous licence CC-BY 4.0)

## Quel est le format du cours?

À la maison:

- + lectures avec vidéos d'accompagnement
- + questionnaires à choix multiples sur le contenu
- + exercices et problèmes avec solutionnaires

En classe:

- + récapitulatif hebdomadaire
- + séance de réponse aux questions
- + ateliers en groupe et discussion.

## Quelle est la charge de travail de ce cours?

- + 3 crédits = 135 heures de travail
- + une moyenne de 9 heures par semaine
- + ne sous-estimez pas l'investissement initial nécessaire
  - + installation de logiciels
  - + apprentissage de la programmation
  - + mise à jour des connaissances

## Quelle est la clientèle cible du cours?

Étudiant(e)s du programme de sciences des données et analytique d'affaires.

Les étudiant(e)s admis(es) ont normalement un diplôme de premier cycle en

- + ingénierie,
- + physique ou
- + mathématique.

Une certaine familiarité mathématique (algèbre linéaire, calcul) est assumée.

# Quels sont les prérequis pour ce cours?

Un premier cours de statistique couvrant les notions suivantes:

- + axiomes de probabilité
- + variables aléatoires
  - + espérance et variance
  - + lois discrètes: Bernoulli, binomiale, Poisson
  - + lois continues: uniforme, exponentielle, normale
- + statistiques descriptives
- + tests d'hypothèses
- + comparaison de moyennes et proportions (un et deux échantillons)
- + régression linéaire simple et corrélation

## Est-ce qu'il y aura de la programmation?

*Oui.* Nous couvrirons les fonctionnalités de base de **SAS** (et **R**) pour l'ajustement des modèles et la visualisation des données.

✦ Le code sera fourni uniquement pour les exercices et les diapos.

Vous devrez remettre votre code avec vos devoirs/projets.

✦ je dois pouvoir reproduire *exactement* vos analyses.

✦ fichier `.txt`(sinon, *Zonecours* vous empêchera de soumettre votre travail)

✦ encodage UTF-8

✦ suivez les consignes (voir instructions) pour nommer vos scripts (indice: `≠ moncode.sas`)

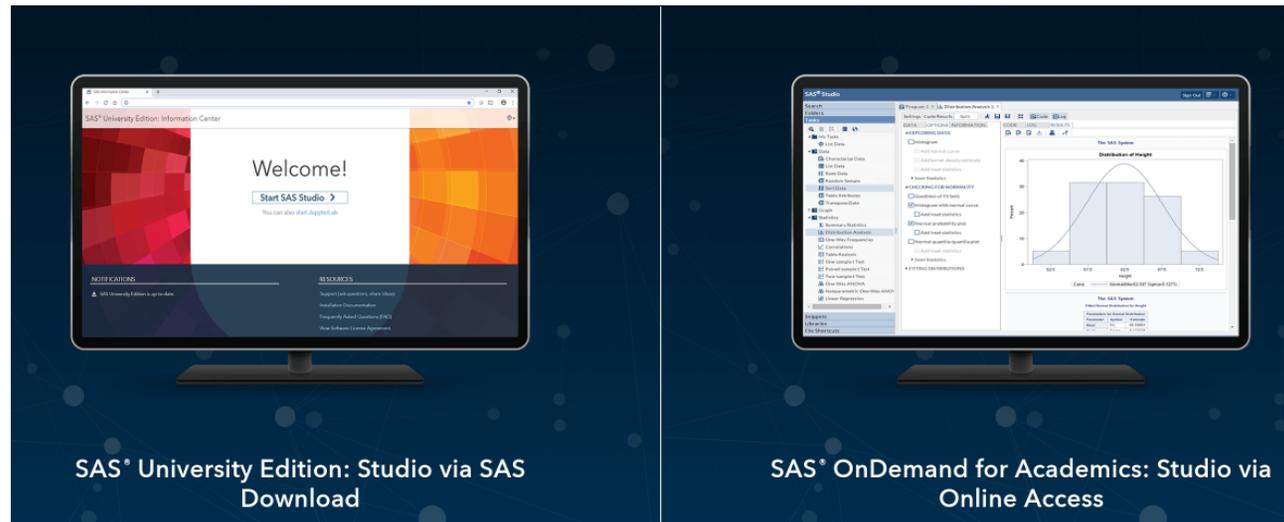
## Quels logiciels utilisera-t-on en classe?

Selon le plan de cours, **SAS** (version de base). Une capsule vidéo est mise à disposition pour vous familiariser avec son utilisation.

- + C'est le seul cours de votre programme qui utilise ce langage de programmation.
  - + avantage comparatif si vous maîtrisez plusieurs langages
  - + principalement utilisé en entreprise
    - + de nombreuses entreprises gèrent leur base de données avec **SAS**
    - + coûte bonbon (\$ \$ \$)
    - + stabilité et soutien
    - + le grand important de code en entreprise freine l'évolution

On va utiliser **SAS®** OnDemand for Academics: Studio.

- + [Pas d'installation requise (via fureteur) + support multi-plateforme]
- + Autre option: **SAS University Edition** -- le service sera discontinué l'an prochain.



Les TIs fournissent aussi le logiciel pour Windows via leur plateforme (~35\$ pour le téléchargement).



Je suis un (ardent) partisan du libre-accès et de **R**, un langage de programmation/logiciel écrit par la communauté

- + c'est gratuit!
- + support multi système d'exploitation
- + téléchargement via [cran.r-project.org](https://cran.r-project.org)
- + utilisé dans les notes de cours
- + je recommande l'environnement de développement **RStudio**

# Quelles sont les attentes de l'enseignant?

- + Participation active: présence supposée au cours (virtuel)
  - + si pas possible, faites-le moi savoir
  - + discussions en petits groupes
  - + posez des questions! il n'y a pas de question stupide
- + Autonomie: vous êtes responsables de votre apprentissage.
  - + être à jour et faire les lectures avant le cours
  - + ne laissez pas l'obscurantisme vous gagner: posez des questions (à moi, à vos pairs)!
- + Rétroaction: des problèmes? faites-en part rapidement
  - + semestre en ligne = points de friction (connexion, manque de rapports humains, etc.)
  - + nombreux outils = période d'adaptation

# Environnement inclusif et respectueux

Bien que virtuel, l'environnement de classe doit être respectueux. Les remarques déplacées, le harcèlement, etc. ne seront pas tolérés.

Faites-moi savoir si

- + je peux faire quelque chose pour améliorer le cours pour vous, pour les autres étudiant(e)s et le groupe
- + une remarque vous rend inconfortable
- + vous avez un nom/pronom/genre qui diffère de celui présent sur *HEC en ligne*
- + votre performance est affectée par des expériences externes au cours: je ferai mon possible pour vous aider ou vous diriger vers des ressources extérieures

# Plagiat

N'essayez pas! Les conséquences seraient fâcheuses et ce serait faire insulte à votre intelligence.

- + si vous vous inspirez de code écrit par d'autres (par ex., [StackOverflow](#)), citez adéquatement vos sources.
- + vous devez programmer vous-même votre code pour les travaux individuels (discussion avec vos pairs acceptée, mais tout partage de code et copier-coller est une infraction au code de conduite de l'école)

# Exemples de plagiat

Voici quelques exemples de plagiat (inspirés d'une liste de l'Université de Toronto):

- + Travailler ensemble pour répondre à des questions.
- + Regarder les réponses de quelqu'un d'autre.
- + Laisser quelqu'un copier ou regarder vos réponses.
- + Partager ou publier les questions d'examen.
- + Discuter les questions ou les réponses de l'examen avec d'autres personnes (qui suivent ou pas le cours).
- + Faire compléter un examen ou devoir par quelqu'un d'autre .