

# **Modélisation statistique**

Plan de cours

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

# Organisation

- + Rencontres hebdomadaires les jeudi en salle Louis-Laberge
- + Deux enseignants
  - + Léo Belzile (CSC 4.850, [leo.belzile@hec.ca](mailto:leo.belzile@hec.ca)) enseigne la première partie
  - + Juliana Schulz (CSC 4.809, [juliana.schulz@hec.ca](mailto:juliana.schulz@hec.ca)) enseigne la deuxième partie

# Contenu du cours

- + Tout le matériel est disponible via le site web du cours: <https://lbelzile.github.io/modstat/> et sur ZoneCours
  - + notes de cours
  - + capsules vidéo
  - + exercices et solutionnaires
  - + code pour démos **R**

# Communications

- + Soumissions de travaux via *ZoneCours*
- + Publiez vos questions de cours sur le canal **Teams**: 0365-MATH60604  
A2024 - code 3v9x63b
- + Pour d'autres questions, écrivez-nous par courriel à [leo.belzile@hec.ca](mailto:leo.belzile@hec.ca)  
et [juliana.schulz@hec.ca](mailto:juliana.schulz@hec.ca) --- mettez-nous toujours ensemble en  
copie-conforme

# Contenu du cours

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

- + Un projet en trois parties valant 30% au total (travail d'équipe), avec remises les vendredi
  - + 11 octobre
  - + 8 novembre
  - + 6 décembre
- + Examen intra (30%) le mardi 29 octobre de 12h à 15h
- + Examen final (40%) le jeudi 12 décembre de 13h30 à 16h30

L'intra et le final sont des examens à livre fermés (pas d'ordinateur ou de programmation). Vous avez droit à une feuille d'aide mémoire format lettre (recto pour intra, recto-verso pour final)

## Quel est le format du cours?

À la maison:

- + lectures des notes de cours
- + visionnement des capsules vidéos
- + exercices et problèmes avec solutionnaires

En classe:

- + cours magistral
- + 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
- + économie 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

## Quel logiciel utilisera-t-on en classe?



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)
- + je recommande l'environnement de développement **RStudio**

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

Oui. Nous apprendrons les bases de **R** pour ajuster les modèles

+ Nous fournirons le code pour les diapos, ateliers et exercices

Vous aurez à produire du code uniquement pour le projet de groupe

+ 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.

# Apprentissage de R

Le Centre d'aide en mathématiques et statistique (CAMS) offre des tutoriels sur **R** (gratuits, mais l'inscription est obligatoire et le nombre de places limité).

- + Introduction à **R**: 1re partie, le vendredi 30 août 2024 de 15h30 à 17h30
- + Introduction à **R**: 2e partie, le vendredi 5 septembre 2024 de 15h30 à 17h30

L'objectif n'est pas de devenir un(e) expert, mais d'être en mesure d'utiliser le langage de programmation pour réaliser les analyses.

- + Des ressources sont disponibles sur le site web du cours.

# Quelles sont les attentes des professeurs?

- + Participation active: présence supposée au cours
  - + 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

# Environnement inclusif et respectueux

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

Ne trichez 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.

## Intelligence artificielle générative (IAG)

Projet: Les étudiant(e)s peuvent utiliser des outils d'IAG pour améliorer et éditer un livrable (texte ou code) dans une perspective de révision, mais le premier jet des analyses statistiques, du code et du rapport écrit doit être un travail original produit par les étudiant(e)s.

Notre vision reflète celle d' [Andrew Heiss](#)



# Utilisation d'IAG

Toute utilisation d'IAG doit être citée adéquatement. Les étudiant(e)s doivent fournir en sus

1. les détails sur l'outil employé (nom et version)
2. la liste des requêtes
3. une copie du code ou texte brouillon soumis
4. une description des modifications apportées aux sorties, le cas échéant
5. une brève réflexion sur leur utilisation de l'IAG