

Description : Modélisation statistique (3 crédits). Ce cours aborde notamment les modèles linéaires, les modèles linéaires généralisés et la régression pour données longitudinales avec effets aléatoires et structures de corrélation. Pour chacune de ces méthodes, les principes théoriques de l'inférence, tels que les tests d'hypothèse, statistiques de tests et méthodes d'estimation seront abordés avec rigueur. Une attention particulière sera aussi portée aux applications en gestion à l'aide du logiciel SAS.

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

Cours : les lundi de 8h30–11h30 (sauf le **mardi** 13 octobre) sur Teams/Zoom. L'étudiant(e) doit visionner les capsules vidéos hebdomadaires avant la rencontre de groupe.

- Les étudiant(e)s doivent installer SAS et maîtriser l'introduction à SAS présentée dans la capsule vidéo avant la deuxième semaine de cours en vue du premier devoir. Le Centre d'aide en mathématiques et statistique (CAMS) offrira des tutoriels dont l'horaire sera publicisé sur *Zonecours*.

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. Introduction à l'analyse de survie

Clientèle cible : Cours obligatoire de la spécialisation « Science des données et analytique d'affaires ».

Cours incompatible : Vous ne pouvez pas vous inscrire à ce cours si vous avez postulé ou réussi le cours MATH 60619(A), ou si MATH 60619(A) fait partie de votre structure.

Objectifs pédagogiques : à la fin du cours, l'étudiant sera en mesure :

1. de comprendre les notions fondamentales et théoriques de l'inférence statistique ainsi que les fondements des méthodes de modélisation utilisées en science des données,
2. d'utiliser ces méthodes en pratique avec le logiciel SAS,
3. de positionner ces méthodes statistiques dans le domaine de la science des données.

Travaux : une copie PDF et le code utilisé pour produire les résultats (sous forme de fichier .txt) doivent être remis sur *ZoneCours* avant la date butoir. Vous devez compléter vos analyses à l'aide de SAS ou R.

Évaluations :

- Devoir 1, comptant pour 15% de la note finale, à remettre au plus tard le 19 octobre.
- Devoir 2, comptant pour 15% de la note finale, à remettre au plus tard le 16 novembre.
- Devoir 3, comptant pour 15% de la note finale, à remettre au plus tard le 7 décembre.
- Projet (équipes de trois ou quatre), comptant pour 25% de la note final, à remettre au plus tard le mercredi 16 décembre.
- examen final récapitulatif (individuel), comptant pour 30% de la note finale.

La note obtenue lors des travaux d'équipe contribue à la note globale seulement si la note combinée des évaluations individuelles est supérieure à 50%. Autrement, la note finale de l'étudiant(e) est calculée au pro rata des évaluations individuelles.

Références : Les notes de cours officielles et les diapositives et vidéos sont en ligne sur le site web du cours. Une liste d'ouvrages de références est également disponible sur *ZoneCours*.

Calendrier provisoire :

Cours	Date	Contenu
1	lundi 31 août	introduction (Chap. 1)
	lundi 7 sept.	fête du travail
2	lundi 14 sept.	modèles linéaires (Chap. 2)
3	lundi 21 sept.	—
4	lundi 28 sept.	—
5	lundi 5 oct.	—
6	mardi 13 oct.	vraisemblance (Chap. 3)
7	lundi 19 oct.	modèles linéaires généralisés (Chap. 4)
	lundi 26 oct.	semaine de relâche
8	lundi 2 nov.	—
9	lundi 9 nov.	données corrélées et données longitudinales (Chap. 5)
10	lundi 16 nov.	—
11	lundi 23 nov.	modèles mixtes (Chap. 6)
12	lundi 30 nov.	—
13	lundi 7 déc.	introduction à l'analyse de survie (Chap. 7)