

Université de Montréal
Faculté des Arts et Sciences
Département de sociologie

SOL 6211 – MODÈLES MULTINIVEAUX ET LONGITUDINAUX
(COURS EN LIGNE ET ASYNCHRONE)

Professeur : **Éric Lacourse**

Session : **Automne 2021**

Bureau : virtuel ;-)

Skype : elac543210

Courriel : eric.lacourse@umontreal.ca

Auxiliaire d'enseignement : **Clémentine Courdi**

Courriel : clémentine.courdi@umontreal.ca

SOL 6211-A MODÈLES DE MÉDIATION, MODÉRATION ET MULTINIVEAUX (1 crédit)

OBJECTIFS DU COURS

Ce cours avancé en statistique sociale appliquée aux données transversales et/ou hiérarchiques a pour but de permettre aux étudiants d'établir de meilleurs liens entre les théories en sciences sociales et de la santé (sociologie, psychologie, criminologie, démographie, médecine, épidémiologie, etc.) et les méthodes et statistiques de régression pour analyser différents modèles théoriques supposant des effets différentiels en fonction de sous-groupes dans la population (variable modératrice) ou spécifiant des mécanismes liant une variable dépendante à une variable indépendante (variable médiatrice). Dans une perspective d'inférence causale, les étudiants devront se familiariser avec les principes fondamentaux nécessaires à la compréhension de l'ajout d'une troisième variable dans des analyses statistiques en sciences sociales et de la santé. Ce cours met l'accent sur les modèles statistiques dont la base est la régression linéaire généralisée pour ultimement aborder les analyses de régression multiniveau.

PRÉREQUIS

Ce cours nécessite une connaissance préalable de la statistique descriptive et inférentielle et quelques notions de base en analyses multivariées (régression linéaire multiple, ANOVA) ainsi que la connaissance d'un logiciel statistique comme SPSS, SAS, R, Mplus ou STATA. Lors de ce cours, le logiciel principalement utilisé sera SAS. Des modules ou macros qui sont spécifiques à ce logiciel seront introduits et approfondis (c.-à-d. PROC REG et PROC MIXED). Les syntaxes pour les autres logiciels, **dont R**, seront disponibles et pourront être utilisées. De plus, ce cours tentera d'une manière appliquée, avec des exemples de recherche concrets, de faire le plus de liens possibles entre les théories et la modélisation statistique des données qui visent à les confirmer.

Plus spécifiquement, ce cours permettra aux étudiants de développer leur jugement critique quant au choix et à l'utilisation des différents modèles statistiques, tant dans la réalisation de leurs travaux de recherche que dans la lecture d'articles et d'ouvrages scientifiques.

Livres recommandés pour les trois cours:

Menard, Scott (2002). *Longitudinal Research*, Second Edition, Newbury Park, CA: Sage Publications.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
<https://gseacademic.harvard.edu/alda/> <https://stats.idre.ucla.edu/other/examples/alda/>

Nagin, D. S. (2005) *Group-Based Modeling of Development*. Cambridge, MA: Harvard University Press. <https://www.andrew.cmu.edu/user/bjones/>

Ces livres sont disponibles à la Bibliothèque Lettres et Sciences Humaines de l'Université de Montréal : <https://bib.umontreal.ca/travailler/les-bibliotheques/lettres-sciences-humaines>

CONTENU DES CHAPITRES DU COURS

Chapitre 1 : Régression linéaire simple avec VI continue

- Modèle et équation de la régression linéaire simple
- Droite de régression et interprétation du R^2
- Tests d'hypothèse et signification statistique
- Résidus, données extrêmes et prédictions

Chapitre 2 : Régression linéaire simple avec VI catégorielle

- Création de variables factices
- Interprétation des coefficients de régression dans ce contexte
- Parallèles avec l'analyse de variance (ANOVA)

Chapitre 3 : Régression linéaire multiple

- Régression linéaire avec deux variables ou plus
- Conditions d'applications et postulats de la régression multiple
- Coefficients bêtas, R^2 , tests d'hypothèse et signification statistique
- Diagnostics et distributions des résidus

Chapitre 4 : Régression non linéaire et transformations linéaires

- Techniques d'identification des relations non linéaires
- Techniques d'identification des relations non normales.
- Les transformations logarithmique, quadratique et cubique

Chapitre 5 : Effets de modération, contexte de régression multiple

- Variables modératrices continues et catégorielles (effets d'interaction)
- Procédures analytiques et tests des effets modérateurs
- Effets principaux et d'interaction dans le contexte de l'ANOVA et de la régression multiple
- Décomposition des effets simples et interprétation des paramètres

Chapitre 6 : Effets de médiation, contexte de régression multiple

- Variables médiatrices continues et catégorielles
- Procédures analytiques (path diagram) et tests des effets médiateurs
- Distinction entre processus de modération et médiation

Chapitre 7 : Introduction aux modèles multiniveaux linéaires avec ordonnée à l'origine et pente aléatoire

- Historique des modèles linéaires hiérarchiques
- Parallèles avec la régression et l'ANOVA
- Décomposition de la variance de niveau 1 et de niveau 2
- Équation composite et introduction de covariables aux différents niveaux
- Ajustement des modèles : -2Log-likelihood (-2LL) et Bayesian Information Criterion (BIC)

SOL 6211-B MODÈLES DE TRAJECTOIRES GÉNÉRALISÉS (1 crédit)

OBJECTIFS DU COURS

Ce cours avancé en statistique sociale appliquée aux données hiérarchiques longitudinales a pour but de permettre aux étudiants d'établir de meilleurs liens entre les théories en sciences sociales et de la santé (sociologie, psychologie, criminologie, démographie, médecine, etc.) et les méthodes et statistiques de régression pour analyser

différents modèles théoriques supposant des effets du temps sur les données recueillies. Ce cours met l'accent sur les modèles statistiques dont la base est la régression linéaire généralisée, tout en incluant des variables latentes continues.

PRÉREQUIS

La connaissance des modèles spécifiant des liens différentiels en fonction de sous-groupes dans la population (variable modératrice) ou spécifiant des mécanismes liant une variable dépendante à une variable indépendante (variable médiatrice) sera un prérequis à ce cours. Dans une perspective d'inférence causale, les étudiants devront se familiariser avec les principes fondamentaux nécessaires à la compréhension de l'ajout d'une troisième variable dans des analyses statistiques en sciences sociales et de la santé.

Ce cours nécessite une connaissance préalable de la statistique descriptive et inférentielle et des analyses multivariées (régression linéaire simple et multiple) ainsi que d'un logiciel statistique comme SPSS, SAS, R, Mplus ou STATA. Lors de ce cours, le logiciel principalement utilisé sera SAS. Des modules ou macros qui sont spécifiques à ce logiciel seront introduits et approfondis (c.-à-d. PROC MIXED, PROC NL MIXED et PROC GLIMMIX). Les syntaxes pour les autres logiciels seront disponibles et pourront être utilisées.

CONTENU DES CHAPITRES DU COURS

Chapitre 1 : L'exploration des données longitudinales

- Modèle de niveau 1 pour l'analyse du processus de changement individuel
- Modèle de niveau 2 pour l'analyse des différences interindividuelles systématiques dans le changement
- Interprétation des paramètres estimés (effets fixes et effets aléatoires)

Chapitre 2 : Spécification composite du modèle multiniveau du changement

- Spécification des composantes du modèle multiniveau du processus de changement
- Modèle inconditionnel de la moyenne et modèle inconditionnel de croissance
- Développement et adaptation d'une taxonomie de modèles
- Comparaison de modèles

Chapitre 3 : Variabilité dans la durée

- Variabilité de la durée des intervalles de temps en analyse multiniveau
- Variabilité du nombre de temps de mesure
- Avantages et inconvénients de cette flexibilité

Chapitre 4 : Inclusion de variables indépendantes variant dans le temps

- Spécification du modèle incluant des variables variant dans le temps
- Illustration des trajectoires incluant des variables variant dans le temps
- Effet différentiel, par vagues, d'une variable variant dans le temps

Chapitre 5 : Révision de la régression logistique

- Analyse de régression logistique simple
- Fondements et postulats de la régression logistique
- Équations et tests d'hypothèse

Chapitre 6 : Loi de Poisson, loi binomiale négative et distributions modifiées en zéro

- Données comptées
- Loi de Poisson, binomiale négative et les distributions modifiées en zéros
- Généralisation aux modèles multiniveaux

Chapitre 7 : Introduction à l'analyse des données manquantes

- Types de données manquantes et leurs conséquences sur les analyses
- Postulats et mécanismes générant des données manquantes
- Approches contemporaines aux données manquantes
- Imputation multiple, maximum de vraisemblance et pondération

SOL 6211-C MODÈLES DE TRAJECTOIRES PAR CLASSES LATENTES (1 crédit)

OBJECTIFS DU COURS

Ce cours avancé en statistique sociale appliqué aux données hiérarchiques longitudinales a pour but de permettre aux participants d'établir de meilleurs liens entre les données et les théories des sciences sociales ou de la santé (sociologie, psychologie, criminologie, démographie, médecine, etc.) en mettant en pratique les méthodes d'analyse de données longitudinales par trajectoires latentes de développement. Les participants devront se familiariser avec les principes fondamentaux nécessaires à la compréhension des techniques d'analyse par regroupement (cluster analysis). Ce cours met l'accent sur les modèles statistiques dont la base est la régression linéaire généralisée multiniveau/hiérarchique, tout en incluant des variables latentes continues et catégorielles.

Ce cours nécessite une connaissance préalable de la statistique descriptive et inférentielle ainsi que des analyses multivariées (régression linéaire simple et multiple). Les connaissances introduites dans le SOL 6211-A et -B sont également souhaitables. Plus précisément, la connaissance des modèles spécifiant des liens différentiels en fonction de sous-groupes dans la population (variable modératrice). Il est également important de maîtriser un logiciel statistique comme SPSS, SAS, R, Mplus ou Stata. Les démonstrations de chacune des analyses sont effectuées à partir du logiciel SAS. **Une macro développée pour SAS par Daniel Nagin et Bobby Jones à l'Université Carnegie Mellon est également utilisée dans ce cours ([PROC TRAJ](#)).**

Malgré qu'il existe d'autres bibliothèques et logiciels pour faire des analyses de manière similaire, il n'existe pas d'outils pour modéliser des groupes de trajectoires de manière identique à ceux développés dans cette procédure SAS. On doit donc faire cette partie du cours dans SAS.

CONTENU DES CHAPITRES DU COURS

Chapitre 1 : Modèles multiniveaux non paramétriques appliqués aux données longitudinales

- Introduction et historique du modèle avec variable latente catégorielle (classes latentes et modèle de mélange fini)
- La sélection des modèles (Bayesian Information Criterion BIC)
- Analyse des estimations et interprétations des probabilités postérieures
- Valeurs de départ, facteurs de risque et trajectoires jointes

Chapitre 2 : Perspectives récentes dans les modèles multiniveaux non paramétriques

- L'estimation des intervalles de confiance autour des trajectoires
- L'ajout simultané de variables latentes continues et catégorielles (growth mixture model)
- L'estimation simultanée de groupes de trajectoires en incluant des données manquantes
- L'estimation de trajectoires jointes avec des variables prédictives
- L'estimation de trajectoires multiples
- Pondération, échantillon et temps d'exposition

DÉMARCHE PÉDAGOGIQUE

Aspects matériels

1. Fureteur Internet Explorer 6 avec service pack 2 ou supérieur, Firefox (version 11 ou ultérieure), Chrome (version XX ou ultérieure).
2. Les logiciels Adobe Reader et PowerPoint doivent également être installés.
3. Vous aurez besoin d'une connexion internet haut débit (DSL ou câble) pour consulter aisément les capsules vidéo.
4. ATTENTION :
 - a. Le cours s'appuie sur des capsules vidéo et consommera donc une certaine quantité de bande passante tout au long de la session.

Assistance technique

Ce cours utilise la plateforme [StudiUM](#) de l'Université de Montréal. Vous devrez utiliser votre [identifiant UNIP](#) et son mot de passe pour vous connecter à la plateforme. Si vous n'avez jamais utilisé la plateforme StudiUM, vous pouvez consulter la page de [Soutien Étudiant](#) pour une présentation de la plateforme.

Approche pédagogique

L'approche pédagogique de ce cours met l'emphase sur l'écoute de capsules vidéo et la résolution de problèmes en ligne. Chacun des chapitres suit la séquence suivante: une lecture, une présentation théorique (vidéo), un mini-quiz, un exemple d'application (vidéo), des exercices pratiques et se termine avec un quiz final.

Des rencontres synchrones auront lieu à quelques reprises au cours de la session en fonction des besoins des étudiants. Ces rencontres seront le moment d'obtenir les réponses à vos questions. Le logiciel *Zoom* sera utilisé lors de ces rencontres.

Interactions et encadrement

Rencontres “en direct”

Les rencontres individuelles ou de groupe avec le professeur et l'auxiliaire d'enseignement se feront au moyen de *Zoom*.

Forums de discussion

Les forums de discussion sont organisés selon trois thèmes. Un forum est réservé aux questions théoriques, un second aux questions d'application dans SAS et le troisième sera consacré aux questions de soutien technique. Le professeur abordera les questions qui seront publiées par les étudiants sur le forum théorique lors des séances synchrones. Les questions posées par les étudiants dans le forum théorique seront répondues sur le

forum ou lors des rencontres « en direct ». Il est important de mettre vos questions sur le forum avant les rencontres « en direct », de cette façon, le professeur et l'auxiliaire peuvent mieux se préparer pour donner des réponses claires et précises.

Activités d'apprentissage

Les chapitres ont tous la même forme et contiennent deux parties principales : une première partie qui est plus théorique et une seconde partie qui est appliquée.

La partie théorique débute par des lectures proposées et est suivie par le visionnement de capsules vidéo théoriques qui sont accompagnées d'une présentation PowerPoint. Les capsules théoriques ont été enregistrées avec la collaboration du [Laboratoire multimédia \(LaM\)](#) de la Faculté des Arts et Sciences. Le rythme des capsules est relativement rapide et vous devrez probablement les visionner plus d'une fois. Un mini-quiz clôt la partie théorique. Les quiz sont formatifs et visent à consolider les apprentissages.

La partie appliquée débute par quelques capsules présentant un exemple d'application au sein du logiciel SAS. Cet exemple est accompagné des syntaxes et des bases de données pour permettre à l'étudiant de reproduire ce qu'il visionne dans les capsules. La partie appliquée se termine par une série d'exercices où l'étudiant doit lui-même appliquer les apprentissages théoriques et méthodologiques à de nouvelles données qui lui seront fournies. Les exercices demandent souvent de mobiliser des compétences acquises, par exemple, lors d'un précédent chapitre. Les solutionnaires à ces exercices sont également fournis pour que l'étudiant puisse valider ses apprentissages.

ÉVALUATION

Comme vous l'avez remarqué, le cours est divisé en trois cours/modules successifs d'un crédit chacun, pour un total de 3 crédits. La majorité des étudiants inscrits suivent les trois modules, de sorte que l'évaluation se fera comme s'il s'agissait d'un cours régulier de trois crédits. Dans ce contexte, le module 1 sert plutôt de révision ou d'introduction et vous avez ensuite un TP à remettre pour le module 2 (SOL 6211-B) et un pour le module 3 (SOL 6211-C), chacun valant pour 50% de la note globale. Votre moyenne pour ces deux travaux sera votre note finale aux trois crédits. La date de dépôt des TPs sera déterminer au début de la session.

Les étudiants qui ne sont inscrits qu'au module 1 (SOL 6211-A) ont un TP à remettre pour ce module, qui vaut pour 100% de la note finale de ce cours d'un crédit.

Les quiz sont obligatoires, mais uniquement à des fins formatives (pas évalués).

ANNEXE 1

Directives et informations utiles pour les travaux pratiques (TP)

Guide pour la rédaction de la section « Résultats » d'un rapport de recherche.

1. Les TP se font à partir de banques de données hiérarchiques qui peuvent être fournies par l'étudiant (idéalement). Sinon, le professeur peut également fournir des données pour faire les TP (moins idéal).
2. Vous devez **indiquer les étapes préliminaires** à l'exécution des analyses et **justifier l'utilisation des tests** effectués en montrant que les conditions (**postulats**) d'utilisation ont été respectées. Utilisez les bons tests en fonction des types de variables en jeu (**nominales, intervalle/ratio**).
3. Vous devez répondre à vos questions en présentant **les statistiques et les résultats à l'appui** (faire les tableaux qui sont nécessaires à la bonne compréhension de votre démarche d'analyse). Les résultats doivent aussi **être présentés et interprétés en texte**.

Vous devez émettre les hypothèses relatives aux liens pouvant exister entre vos variables.

Les analyses doivent se faire **en 3 étapes**.

1. Analyse descriptive **univariée** (mesures de tendance centrale, histogramme, etc.);
2. Réalisation des analyses où le chercheur examine tous les liens **bivariés** entre les variables;
3. Réalisation des analyses **multivariées** (analyse de régressions multiniveaux)

Voici une suggestion plus détaillée des étapes de l'analyse : 1) faire une liste des variables d'intérêts dans la base de données; 2) déterminer quelles sont les variables indépendantes et dépendantes; 3) déterminer la nature des variables (intervalle, ordinale, nominale); 4) développer des hypothèses pour que l'on comprenne clairement le lien entre les variables indépendantes et dépendantes; 4) faire les analyses **univariées, bivariées et multivariées**.

10 % des points seront accordés pour la qualité de la présentation.

ANNEXE 2

Intégrité et plagiat

Le plagiat, sous toutes ses formes, est pris très au sérieux par le Département de sociologie et par la Faculté des Arts et sciences. C'est du plagiat si:

- vous copiez des phrases ou parties de phrases, mot à mot ou presque, d'un texte rédigé par quelqu'un d'autre, sans suivre les règles de la citation mentionnées plus loin;
- vous présentez certaines idées trouvées dans des articles ou livres, comme si elles sont les vôtres, sans référer aux auteurs originaux selon les règles mentionnées plus loin;
- vous copiez le travail (ou une partie du travail) d'un autre, présentez le travail d'un autre comme s'il est le vôtre ou permettez à quelqu'un d'autre de copier votre travail et ceci, même si vous avez travaillé ensemble.
- Compris dans la définition du plagiat: tout copiage, fraude et tricherie, que ce soit dans le cadre de travaux préparés à domicile ou des examens en salle de cours.
- Même une seule phrase copiée constitue un plagiat.

L'Université de Montréal a pris des engagements pour assurer l'intégrité intellectuelle des travaux qui y sont produits. Les professeurs et chargés de cours sont vigilants par rapport au plagiat et utilisent des moyens efficaces pour le retracer. Le plagiat peut se solder par la note « F », soit échec, et même aller jusqu'à la suspension ou le renvoi de l'Université. Consultez [le site web sur l'intégrité et le plagiat](#) pour prendre connaissance du règlement et tester vos connaissances sur le sujet.

Pour éviter le plagiat dans vos travaux et examens, il faut:

- Rédiger les travaux et examens vous même, dans vos propres mots – c'est d'ailleurs la preuve que vous comprenez ce qu'un auteur a écrit;
- Le cas échéant, insérer les mots d'un autre auteur entre guillemets, en italiques ou dans un bloc de texte décalé et toujours indiquer l'origine de la citation, selon un style bibliographique reconnu (voir le site web de l'UdeM mentionné ci-bas).
- Citer l'auteur ou les auteurs à l'origine des idées présentées dans vos travaux et ceci, même si vous utilisez vos propres mots;
- Inclure dans votre bibliographie, les œuvres citées ET celles qui ont contribué directement au développement des idées exprimées dans votre travail.

Pour des renseignements complets sur la manière de citer vos sources :

<https://bib.umontreal.ca/citer/comment-citer>

Les modes de communication

Ce cours se fonde essentiellement sur une communication en ligne. Vous aurez l'opportunité d'interagir avec vos collègues à travers le forum et les hangouts. Pour que cette communication en ligne soit efficace et agréable, il est important que vous fassiez plus que simplement publier vos questions et vos idées. Vous devriez également lire les publications de vos collègues et y répondre lorsque vous avez des suggestions ou des commentaires à y amener.

Gardez à l'esprit que l'environnement de communication en ligne contient des espaces privés et d'autres publics. Les courriels échangés sont privés, mais peuvent être facilement rendus publics en le transférant. Les forums de discussion sont des espaces publics.

Il est aussi important de se souvenir que vous ne pourrez pas obtenir les signaux non verbaux envoyés par le professeur. Le ton de voix insistant ou la répétition d'une phrase par le professeur pour marquer un passage important ne vous sera pas accessible. Pas plus que le professeur n'aura accès à votre expression confuse ou frustrée par un délai ou une exigence. **Vous devez donc communiquer avec le professeur ou les auxiliaires d'enseignement pour qu'ils vous aident.**

Pour que l'expérience soit agréable, gardez à l'esprit :

- D'initier plus de contacts qu'à l'habitude.
- D'être direct, persistant et clair lorsque vous ne comprenez pas quelque chose.
- De communiquer souvent et tôt dans le semestre.

Vous devez garder en tête que bien que l'Internet soit disponible 24h/24h, vos collègues, les auxiliaires d'enseignement et le professeur ne le sont pas. Vous pouvez vous attendre à une réponse du corps enseignant dans les 24 heures lors de questions posées la semaine sur le forum ou par courriel. Les questions posées durant les weekends peuvent recevoir une réponse uniquement le lundi.

Professionnalisme

L'environnement académique est un lieu d'apprentissage et de débats. Le professeur fera tous les efforts nécessaires pour que vous vous sentiez à l'aise de partager vos idées et opinions, de poser toutes les questions que vous aurez et d'exprimer vos croyances. En retour, il est attendu que vous adoptiez un comportement similaire lors des échanges avec vos collègues, que ceux-ci soient en ligne ou en face à face.

Les étudiants ont le droit d'étudier dans un contexte libre de harcèlement, ce cours suit les directives établies dans la Politique contre le harcèlement de l'Université de Montréal. Les étudiants peuvent consulter le site du [Bureau d'intervention en matière de harcèlement](#) pour obtenir un résumé de la politique de l'université, une définition et des exemples de manifestation de harcèlement.