

## PLAN DE COURS

DMO 6371

HIVER 2020

MÉTHODES STATISTIQUES EN DÉMOGRAPHIE

3 CR.

COURS: **lundi ET mercredi, de 16h00 à 17h30**  
EXAMEN INTRA: **lundi, 26 février 2020**  
EXAMEN FINAL: **lundi, 27 avril 2020**

- Modification du choix de cours: date limite le **21 janvier 2020** (*tout cours annulé pendant la période active de modification du choix de cours ne sera pas mentionné dans le relevé de notes et ne sera pas facturé à l'étudiant*);
- Abandon d'un cours: date limite le **13 mars 2020** (*entre le 22 janvier et le 13 mars, l'abandon de cours peut se faire en se présentant au Secrétariat de son département; tout cours abandonné fera l'objet d'une mention "ABA" sur le relevé de notes et la facturation des frais de scolarité sera maintenue*).

Professeur: **CHAE, Sophia**  
Courriel: [sophia.chae@umontreal.ca](mailto:sophia.chae@umontreal.ca)  
Local: C-5036  
Disponibilité: Vendredi, de 11h00 à 12h00

Une version électronique de ce plan de cours est disponible sur Internet. On peut y accéder par la page d'accueil du Département de démographie ([www.demo.umontreal.ca](http://www.demo.umontreal.ca)). Cependant, notez que les informations qui suivent peuvent faire l'objet de modifications au cours du trimestre. Le cas échéant, le professeur vous avisera en classe ou, s'il y a lieu, au moyen du calendrier affiché sur le site StudiUM du cours (<https://studium.umontreal.ca/>).

### OBJECTIFS DU COURS

- Faire le survol des méthodes statistiques de base utilisées en démographie sociale (et en sociologie, santé publique...) pour l'analyse explicative, c'est-à-dire l'analyse des causes et conséquences des phénomènes démographiques.
- Sensibiliser les étudiants aux problèmes méthodologiques et conceptuels que l'on rencontre fréquemment dans les études empiriques et donner les outils statistiques pour les contourner ou les éviter.
- Donner une compréhension intuitive et, dans une certaine mesure, mathématique, de comment ces techniques fonctionnent et pourquoi elles sont appropriées (ou non) dans des situations différentes.
- Développer la capacité fonctionnelle et l'habitude d'utiliser ces techniques par les étudiants à travers des exercices sur micro-ordinateur.

Les étudiants apprendront les notions et méthodes statistiques de base les plus importantes utilisées en démographie pour l'analyse explicative: les causes et conséquences des phénomènes. Le cours ciblera les régressions linéaires et non-linéaires binomiales et polychotomiques, et mettra l'accent sur les enjeux conceptuels et les problèmes méthodologiques que l'on rencontre fréquemment dans les études empiriques. Ces connaissances sont essentielles pour comprendre la plupart des études publiées dans les revues scientifiques et, le plus souvent, pour faire un mémoire ou une thèse en démographie. La maîtrise de ces méthodes est un pré-requis pour accéder aux cours de statistiques plus avancés, tels que l'analyse multiniveaux ou les méthodes de risques et durées. Les étudiants auront à faire beaucoup d'exercices pratiques afin de perfectionner leurs compétences à utiliser ces méthodes et d'interpréter correctement les résultats. Les fondations mathématiques de ces méthodes ne seront pas présentées dans ce cours; pour ceux qui s'y intéressent, vous pouvez consulter les cours offerts en statistique ou en économétrie.

Voici quelques-unes des questions auxquelles nous allons répondre au cours du trimestre:

- Comment utiliser les méthodes statistiques pour étudier les déterminants de la survie des enfants?
- Comment savoir si une variable « indépendante » a un impact significatif sur ce poids (par exemple, l'effet de la scolarité de la mère sur le poids à la naissance des enfants) et que veut dire significatif?
- Que sont les variables continues, dichotomiques (*dummies*), polychotomiques, *proxies* et les interactions? Comment les utiliser dans les régressions et ensuite interpréter les résultats?
- Quels sont les problèmes liés à l'utilisation d'une régression linéaire pour étudier le nombre d'enfants nés des femmes dans une population donnée?
- Pourquoi une régression logistique est-elle mieux qu'une régression linéaire pour l'analyse des déterminants de la participation des femmes au marché de travail (et ça veut dire quoi "mieux")?
- Pourquoi la fécondité des femmes ne peut-elle pas être considérée comme une simple cause de leur décision de travailler ou pourquoi les pratiques d'allaitement des enfants ne sont-elles pas considérées comme un déterminant « exogène » de leur santé ou de leur survie? Quelle méthode d'estimation est adéquate pour évaluer les effets causals dans ces situations?
- Pourquoi et de quelle manière les techniques statistiques habituelles donnent-elles les résultats biaisés lorsque vos données proviennent d'une enquête pondérée ou en grappe?
- Pourquoi et comment les régressions de type pas-à-pas ("*stepwise*") donnent-elles des résultats systématiquement biaisés dans les études de type « cause à effet »?
- Comment estimer les divers modèles de régression sur l'ordinateur et quels sont les problèmes que l'on rencontre souvent (valeurs manquantes, préparation de la base de données, etc.)?
- Et le plus important: comment procéder pour faire une étude empirique, soit conceptualiser les liens de causalité, choisir le modèle et ensuite estimer et interpréter les résultats?

## PRÉ-REQUIS

*Pour ceux qui ont besoin de revoir les notions statistiques et mathématiques de base, il est très important d'étudier durant la première semaine:*

Wooldridge, Jeffrey M. (2018), *Introduction à l'économétrie: Une approche moderne*, 2<sup>e</sup> ed., de Boeck Supérieur.

- Annexe A: Outils mathématiques de base
- Annexe B: Éléments de probabilités
- Annexe C: Éléments de statistique mathématique

## LECTURES OBLIGATOIRES

Wooldridge, Jeffrey M. (2018). *Introduction à l'économétrie: Une approche moderne*, 2<sup>e</sup> ed., de Boeck Supérieur.

- Note: ce livre est disponible à la librairie Jean-Brillant (dans la section de sciences économiques) et à la réserve de la BLSH.

Note: il y aura d'autres lectures à faire qui seront ajoutées au plan de cours tout au long du trimestre.

## AUTRES OUVRAGES UTILES POUR CE COURS

- \*Allison, Paul (1998). *Multiple Regression: A Primer*, Thousand Oaks, Sage Publications.
- Bressoux, Pascal (2010). *Modélisation statistique appliquée aux sciences sociales*, 2<sup>e</sup> ed., De Boeck Supérieur.
- Briscoe, Akin & Guilkey (1990). "People are not Passive Acceptors of Threats to Health: Endogeneity and its Consequences", *International Journal of Epidemiology*, 19(1): 146-153.
- Jan Kmenta (1997). *Elements of Econometrics* (2<sup>e</sup> édition: University of Michigan Press).
- Alfred Demaris (1992). *Logit Modeling: Practical Applications*, Beverly Hills, Sage Publications 86.
- V.T. Farewell (1979). "Some Results on the Estimation of Logistic Models Based on Retrospective Data", *Biometrika* 66(1): 27-32.
- J.H. Stock & M.W. Watson (2007). *Introduction to Econometrics* (2<sup>e</sup> édition: Addison-Westley).
- Fred Pampel (2000). *Logit Regression: A Primer*, Beverly Hills, Sage Publications 132.
- Lee, E.S & R.M. Forthofer (2006). *Analyzing complex survey data*, Beverly Hills, Sage Publications 07-071.
- Pétry, F. & Gélinau F (2009). *Guide pratique d'introduction à la régression en Sciences sociales*, Les presses de l'Université Laval.
- Riberdy et al. (2000). "Des concepts aux chiffres", ch. 4 (pp. 131-136) dans *Culture, santé et ethnicité*, (S. Gravel et A. Battaglini, Eds) Régie régionale de la santé et des services sociaux de Montréal-centre.

\*Ce livre est fortement recommandé.

## ÉVALUATION

Votre note finale dépendra de vos résultats:

- Participation en classe\* ..... 10%
- Travaux pratiques (5)\*\* ..... 25%
  - 22 janvier
  - 5 février
  - 19 février
  - 18 mars
  - 1<sup>er</sup> avril
- Examen intra (**26 février – 16h à 18h**) ..... 30%
- Examen final\*\*\* (**27 avril – 16h à 19h**) ..... 35%

\* Votre participation en classe sera mesurée par votre présence et votre participation aux activités en classe. Ces activités peuvent inclure des quiz, des questions à court développement et des travaux de groupe. Si vous êtes présents et que vous participez, vous recevrez un crédit complet pour l'activité; dans le cas contraire, vous en obtiendrez la note minimale « 0 ». Je choisirai au hasard 10 séances pour mesurer votre participation.

\*\* Pour faire vos travaux pratiques, vous aurez besoin d'utiliser le logiciel Stata. Stata est disponible dans plusieurs laboratoires micro-informatiques sur le campus. Je ferai une introduction à Stata dans le laboratoire micro-informatique au 3<sup>e</sup> étage du pavillon Lionel-Groulx (C-3115) le 13 janvier.

\*\*\* L'examen final couvrira toute la matière, mais en particulier la matière vue depuis l'examen intra. Les notes A, B, C... seront déterminées selon une grille « notes numériques - notes en lettres » définie par la professeure et non pas selon la grille standard de l'université.

Selon le règlement pédagogique (article 9.9 reproduit ci-dessous), l'étudiant doit motiver toute absence à une évaluation; pour ce faire, il faut s'adresser au Secrétariat de son département et non au professeur. Seul un motif imprévu et hors du contrôle de l'étudiant peut être acceptable.

« L'étudiant doit motiver, par écrit, toute absence à une évaluation ou à un cours faisant l'objet d'une évaluation continue **dès qu'il est en mesure de constater qu'il ne pourra être présent à une évaluation** et fournir les pièces justificatives. Dans les cas de force majeure, il doit le faire le plus rapidement possible par téléphone ou courriel **et fournir les pièces justificatives dans les cinq jours ouvrés suivant l'absence**.

Le doyen ou l'autorité compétente détermine si le motif est acceptable en conformité des règles, politiques et normes applicables à l'Université.

Les pièces justificatives doivent être dûment datées et signées. De plus, le **certificat médical doit préciser les activités auxquelles l'état de santé interdit de participer, la date et la durée de l'absence, il doit aussi permettre l'identification du médecin.** »

## PLAN GÉNÉRAL DU COURS

1. Introduction
  - a. Brève révision des statistiques
  - b. Introduction à Stata
2. Le modèle classique de régression linéaire
  - a. Régression simple
  - b. Régression multiple
  - c. Types de variables indépendantes
  - d. Utilisation du F-test et interactions
  - e. Multicolinéarité
3. Problèmes avec les hypothèses de base du modèle
  - a. Problèmes de spécification de l'équation
    - i. Non-linéarité
    - ii. Variables indépendantes pertinentes omises de l'équation
    - iii. Variables indépendantes superflues incluses dans l'équation
  - b. Problèmes de mesures des variables
    - i. Censure de la variable dépendante et le modèle Tobit classique
    - ii. Erreurs de mesures aléatoires et de l'utilisation des variables de substitution
  - c. Données manquantes
  - d. Endogénéité
  - e. Les problèmes liés à la variation aléatoire " $\epsilon$ " (l'hétéroscédasticité et l'autocorrélation)
  - f. Introduction aux pondérations et aux effets des grappes
  - g. Données de panel
4. Mener à bien un projet empirique
5. Les modèles non linéaires
  - a. Modèles simples binomiaux: probabilité linéaire, logit et probit
  - b. Modèles polychotomiques non ordonnés et ordonnés
  - c. Modèles de survie à temps discret

## 1. Introduction

6 janvier Introduction: les objectifs et le plan du cours  
 ✓ *Wooldridge*: Annexe A: Outils mathématiques de base

8 janvier Brève révision des statistiques  
 ✓ *Wooldridge*: Annexes B et C

13 janvier Atelier sur l'utilisation du logiciel *Stata* au laboratoire informatique, au C-3115 du pavillon Lionel-Groulx

## 2. Le modèle classique de régression linéaire

15 janvier Régression simple  
 ✓ *Wooldridge*: chapitre 2  
 \*\*\* T.P. #1: remise de l'énoncé

20 janvier Régression simple (suite)

22 janvier Régression multiple: introduction  
 ✓ *Wooldridge* : chapitre 3 (p. 96-116; p. 121-123; p. 127-134)  
 \*\*\* T.P. #1: remise des copies

27 janvier Régression multiple: introduction (suite)

29 janvier Régression multiple: inférence  
 ✓ *Wooldridge*: chapitre 4  
 \*\*\* T.P. #2: remise de l'énoncé

3 février Régression multiple: inférence (suite)

5 février Types de variables indépendantes  
 ✓ *Wooldridge*: chapitre 7 (p. 276-289)  
 \*\*\* T.P. #2: remise des copies

10 février Types de variables indépendantes (suite)  
 Interactions  
 ✓ *Wooldridge*: chapitre 7 (p. 290-298)

## 3. Problèmes avec les hypothèses de base du modèle

12 février Multicolinéarité  
 ✓ *Wooldridge*: chapitre 3 (p. 123-127)  
 \*\*\* T.P. #3: remise de l'énoncé

17 février Problèmes de spécification de l'équation  
 ■ Non-linéarité  
 ✓ *Wooldridge*: chapitre 9 (p. 364-369)

19 février Problèmes de spécification de l'équation (suite)  
 ■ Les variables indépendantes pertinentes omises de l'équation  
 ■ Les variables indépendantes superflues incluses dans l'équation  
 ✓ *Wooldridge*: chapitre 3 (p. 116-121)  
 \*\*\* T.P. #3: remise des copies

24 février Révision pour l'examen intra

26 février **Examen intra (16h à 18h)**

### 2 et 4 mars: Semaine de relâche

- 9 mars Problèmes de mesure des variables  
▪ Censure de la variable dépendante et le modèle Tobit classique  
✓ *Wooldridge*: chapitre 17 (p. 693-701)
- 11 mars Problèmes de mesures des variables (suite)  
▪ Erreurs de mesures aléatoires et de l'utilisation des variables de substitution  
Données manquantes  
✓ *Wooldridge*: chapitre 9 (p. 369-377 ; p. 379-394)  
\*\*\* T.P. #4: remise de l'énoncé
- 16 mars Endogénéité  
✓ *Wooldridge*: chapitre 15
- 18 mars Les problèmes liés à la variation aléatoire " $\varepsilon$ " (l'hétéroscédasticité et de l'autocorrélation)  
Introduction aux pondérations et aux effets des grappes  
✓ *Wooldridge*: chapitre 8  
\*\*\* T.P. #4: remise des copies
- 23 mars Données de panel  
✓ *Wooldridge*: chapitre 13 (p. 537-554); chapitre 14

### 4. Mener à bien un projet empirique

- 25 mars Mener à bien un projet empirique  
**Conférencier invité: Thomas LeGrand (professeur émérite, Département de démographie)**  
✓ *Wooldridge*: chapitre 19  
✓ Regnerus, Mark (2012). "How different are the adult children of parents who have same sex relationships? Findings from the New Family Structures Study", *Social Science Research* 41(4):752-770.  
✓ Gates, Gary J. et al (2012). "Letter to the editors and advisory editors of *Social Science Research*", *Social Science Research*, 41(4):752-770.  
\*\*\* T.P. #5: remise de l'énoncé

### 5. Les modèles non linéaires

- 30 mars Modèles simples binomiaux: probabilité linéaire, logit et probit  
✓ *Wooldridge*: chapitre 17 (p. 680-693)
- 1<sup>er</sup> avril Modèles simples binomiaux: probabilité linéaire, logit et probit (suite)
- 6 avril Modèles polychotomiques non ordonnés et ordonnés  
\*\*\* T.P. #5: remise des copies
- 8 avril Modèles de survie à temps discret
- 13 avril Congé – Lundi de Pâques
- 15 avril Révision pour l'examen final
- 27 avril Examen final (16h00 à 18h50)**