

**DÉPARTEMENT DE MATHÉMATIQUES ET DE STATISTIQUE
UNIVERSITÉ DE MONTRÉAL
PLAN DE COURS: AUTOMNE 2024**

Sigle du cours:	STT-2000
Titre du cours:	Échantillonnage
Professeur:	Pierre Duchesne
Bureau:	4251, Pavillon André-Aisenstadt
Courriel:	pierre.duchesne@UMontreal.CA
Horaire des travaux pratiques:	mercredi, 16h30-18h30
Horaire du cours:	mardi, 13h30-14h30 (Salle: Z-300, Pavillon Claire-McNicholl). jeudi, 12h30-14h30 (Salle: Z-300, Pav. Claire-McNicholl; sauf 28 nov.). jeudi, 12h30-14h30 (Salle: 1177, Pav. André-Aisenstadt; 28 nov. 2024).
Disponibilité:	Sur rendez-vous.

Plan du cours

1. Structure d'une enquête par sondage.

Conception et planification: population cible; variable(s) d'intérêt; questionnaire; base de sondage. Exécution du plan préconçu: (1) tirage de l'échantillon, (2) collecte de données, (3) vérification des données et imputation, (4) estimation et analyse, (5) publications des résultats.

2. Méthodes d'échantillonnage.

Tirage à l'aide d'une base de sondage. Probabilités d'inclusion. Estimation utilisant les poids d'échantillonnage. Plans d'échantillonnage à un et deux degrés. Tirage de Bernoulli. Tirage simple (avec remise, sans remise). Tirage stratifié simple: choix des strates; répartition optimale de l'échantillon sur les strates; gain de précision dû à la stratification. Tirage systématique. Tirage en grappes: variance intra-grappes; mesure d'homogénéité des grappes; perte de précision causée par l'effet des grappes; Tirage à deux degrés (tirage simple à chaque degré).

3. Utilisation de variables auxiliaires.

Sources d'information auxiliaire. Régression linéaire entre la (les) variable(s) d'intérêt et les variables auxiliaires. Estimateur par la différence. Estimateur par le ratio. Estimateur par la régression. Estimateur de calage. Gain en précision. Estimation avec les probabilités inégales.

4. Types d'erreur de sondage.

Erreur due à l'échantillonnage; erreur non due à l'échantillonnage. Non-réponse: biais dû à la non-réponse: méthodes d'ajustement pour la non-réponse. Erreurs dues aux méthodes de mesures: biais de l'intervieweur.

5. Méthode de Monte Carlo.

Nombres aléatoires et pseudo-aléatoires. Introduction aux méthodes de simulation de variables aléatoires continues et discrètes. Création d'échantillons artificiels. Étude du comportement des estimateurs par la méthode de Monte Carlo.

6. Sondages empiriques.

Introduction à la méthode des quotas et aux méthodes non probabilistes. Avantages et inconvénients. Limites de ces méthodes.

Barème

Le barème proposé est le suivant:

Examen intra:	30% (date de l'intra: 30 octobre 2024; 16h30-18h30; Salle: 1175, Pav. André-Aisenstadt)
Examen final:	40% (date du final: 18 décembre 2024; 8h30-11h30; Salle: 1175, Pav. André-Aisenstadt)
Projet:	7.5%
Travaux :	$7.5\% \times 3 = 22.5\%$

La démarche pour les travaux sera la suivante: les énoncés des travaux pratiques sont sur StudiUM, et des exercices seront résolus et mis disponibles. Notez que tous les énoncés sont déjà sur StudiUM. Quelques travaux d'une envergure un peu plus grande seront aussi proposés (une lecture, 2 devoirs et un projet). À l'exception de la lecture, qui est un travail individuel, vous êtes fortement encouragé à remettre les travaux par équipes de deux.

Pour réussir le cours, l'étudiant doit obtenir plus de 35 points dans la somme pondérée de l'intra et du final (valant au total 70%).

Vous devez remettre les devoirs en format électronique sur StudiUM, incluant les codes et autres documents auxiliaires. Vous ne devez pas rendre les devoirs sur mon adresse courriel. Ces courriels seront effacés sans avertissement, et ne seront pas corrigés.

Les étudiants inscrits au Bureau de Soutien aux étudiants en Situation de Handicap (BSESH) désirant bénéficier de mesures d'accommodement aux examens (intra et final) sont priés de consulter le lien suivant pour connaître la procédure à suivre: <https://safire.umontreal.ca/reussite-et-ressources/mesures-daccommodement-aux-examens-pour-les-etudiants-en-situation-de-handicap/>

Objectif du cours

Les sondages sont très utilisés à notre époque. Par exemple, dans la production des statistiques officielles d'un pays ou d'une province (Statistique Canada, Institut de la statistique du Québec), pour mesurer l'opinion publique (intention de vote, marketing), dans les études de recherche (démographie, sociologie, psychologie, médecine). Le but de ce cours est de présenter quelques aspects importants de la **méthodologie d'enquêtes par sondage**, compte tenu des outils mathématiques et statistiques dont on dispose en début de deuxième année.

On étudiera les plans d'échantillonnage les plus souvent utilisés en pratique par les agences gouvernementales de la statistique et par les instituts de sondage. Parmi les techniques étudiées, on note l'échantillonnage stratifié, l'échantillonnage systématique et la sélection par grappes. Le rôle de l'information auxiliaire sera abordée, ainsi que son insertion dans les plans d'échantillonnage et dans les estimations. On discutera des avantages pratiques de chacune des diverses méthodes, et on examinera les gains ou les pertes de précision des divers estimateurs, comparativement à un tirage naïf comme le tirage aléatoire simple. On étudiera la méthode de Monte-Carlo et son utilisation dans la simulation de problèmes statistiques complexes.

Le cours repose largement sur la troisième édition de l'ouvrage de Sharon Lohr (2022). Les deux premières éditions ont été populaires et sont encore très utilisées. Le livre de Sharon Lohr est fortement recommandé.

On note dans la liste des références l'ouvrage de Särndal, Swensson & Wretman (1992), dont l'usage est très répandu, et qui est un des manuels de base pour les méthodologistes dans les grandes agences statistiques, comme par exemple à Statistique Canada. Vous allez trouver en librairie d'autres éditions de 2003 ou 2013, mais ce ne sont que des stratégies de marketing, le livre n'a pas changé et demeure

pertinent maintenant. Ce manuel est aussi recommandé.

L'ouvrage de Cochran (1977) est une référence secondaire. Ce livre demeure une référence classique, qui a marqué de manière significative ce domaine de la statistique qu'est la théorie des sondages. L'ouvrage demeure utilisé compte tenu de la qualité d'écriture et de la pertinence des problèmes.

Le guide non-mathématique de Satin et Shastry (1993), n'est pas obligatoire, mais peut être très utile pour ceux qui pourraient envisager de faire le concours annuel de Statistique Canada (qui a lieu habituellement dans les alentours de novembre). Par expérience avec des amis et collègues, ce guide peut être d'une aide précieuse lors d'un examen ou d'une entrevue à Statistique Canada. Il en est de même de Méthodes et pratiques d'enquête (2010).

Toutes ces références sont disponibles à la bibliothèque.

Références

Cochran, W. G. (1977), *Sampling Techniques*, troisième édition, New York : Wiley.

Lohr, S. L. (2022), *Sampling : Design and Analysis*, troisième édition, New York : CRC Press.

Méthodes et pratiques d'enquête (2010), éditrices : Sarah Franklin et Charlene Walker, publication 12-587-X, Statistique Canada. (Le manuel peut être téléchargé ici: <https://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-fra.pdf>)

Särndal, C. E., Swensson, B. et Wretman, J. H. (1992), *Model Assisted Survey Sampling*, New York : Springer-Verlag.

Satin, A. et Shastry, W. (1993), *L'échantillonnage, un guide non-mathématique*, seconde édition, Ottawa : Statistique Canada.