

**MAT3777**

**Échantillonnage et sondages**

Plan de cours

# Description du cours

Introduction à la conception et l'analyse des sondages. Échantillonnage simple aléatoire, stratifié, proportionnel, post-stratifié, par grappes et à niveaux multiples. Développement complet de la théorie; exemples. Analyse de données à l'aide des logiciels de statistique.

Préalable: **MAT 2775** ou MAT 2777 ou MAT 2778 ou MAT 2779.

## Objectifs:

- accréditation de la Société statistique du Canada (A.Stat)
- préparation à l'examen de StatCan
- compréhension des concepts
- application des concepts

# Détails

**Enseignant:** Patrick BOILY

## Horaire

- LUN 14:30-15:50 | LPR 155
- JEU 16:00-17:20 | LPR 155
- disponibilités: sur rendez-vous (ZOOM, SLACK, STEM541)

## Sites web

- [data-action-lab.com/mat3777](https://data-action-lab.com/mat3777)
- [mat3777-h22.slack.com](https://mat3777-h22.slack.com)

## Ouvrages suggérés:

- *Méthodes d'échantillonnage à l'usage des sondages*, S. Alalouf
- *Méthodes et pratiques d'enquête*, Statistique Canada
- *Sampling: Design and Analysis*, 3rd edition, S. Lohr (en anglais)

# Plan de cours

## 1. Introduction

### 1.1 Analyse des données

*1.1.1 Système de collecte de données*

*1.1.2 Formulation du problème*

*1.1.3 Types de données*

*1.1.4 Stockage et accès aux données*

### 1.2 Échantillonnage statistique

*1.2.1 Modèle d'échantillonnage*

*1.2.2 Facteurs déterminants*

*1.2.3 Bases de sondages*

*1.2.4 Concepts fondamentaux*

*1.2.5 Modes de collecte des données*

*1.2.6 Types d'échantillonnage*

## 2. Échantillonnage aléatoire simple

### 2.1 Motivation

### 2.2 Concepts fondamentaux

### 2.3 Estimation et intervalles de confiance

*2.3.1 Estimation de la moyenne*

*2.3.2 Estimation du total*

*2.3.3 Estimation d'une proportion*

### 2.4 Détermination de la taille de l'échantillon

*2.4.1 Moyenne*

*2.4.2 Total*

*2.4.3 Proportion*

# Plan de cours

## 3. Echantillonnage aléatoire stratifié

3.1 Motivation

3.2 Estimation et intervalles de confiance

*3.2.1 Estimation de la moyenne*

*3.2.2 Estimation du total*

*3.2.3 Estimation d'une proportion*

3.3 Répartition et détermination de la taille de l'échantillon

*3.3.1 Taille de l'échantillon, selon une marge d'erreur*

*3.3.2 Taille de l'échantillon, selon un budget*

3.4 Comparaison entre EAS et STR

## 4. Estimation par le quotient, par la régression, et par la différence

4.1 Motivation

4.2 Estimation par le quotient

*4.2.1 Estimateur du quotient*

*4.2.2 Biais de l'estimateur*

*4.2.3 Variabilité de l'estimateur*

*4.2.4 intervalle de confiance*

*4.2.5 Estimation de la moyenne et du total*

*4.2.6 Détermination de la taille de l'échantillon*

# Plan de cours

## 4. Estimation par le quotient, par la régression, et par la différence (cont.)

### 4.3 Estimation par la régression

*4.3.1 Estimateur de régression*

*4.3.2 Biais de l'estimateur*

*4.3.3 Variabilité de l'estimateur*

*4.3.4 Intervalle de confiance*

*4.3.5 Détermination de la taille de l'échantillon*

### 4.4 Estimation par la différence

### 4.5 Comparaisons entre ...

*4.5.1 EAS et la méthode du quotient*

*4.5.2 EAS et la méthode de la régression*

*4.5.3 Comparaison entre EAS et la méthode de la différence*

*4.5.4 Comparaison entre les trois méthodes*

## 5. Conception de questionnaire et collecte automatisée

### 5.1 Conception de questionnaires

*5.1.1 Principes fondamentaux*

*5.1.2 Types de questions*

*5.1.3 Considérations relatives à la formulation*

*5.1.4 Ordre des questions*

### 5.2 Collecte automatisée

*5.2.1 Liste de contrôle*

*5.2.2 Considérations d'ordre éthique*

*5.2.3 Qualité des données de la toile*

*5.2.4 Technologies du web 101*

*5.2.5 Boîte à outils*

# Plan de cours

## 6. Échantillonnage par grappes

6.1 Motivation

6.2 Estimation et intervalles de confiance

*6.2.1 Estimation de la moyenne - tailles identiques*

*6.2.2 Estimation de la moyenne - tailles différentes*

*6.2.3 Estimation du total*

*6.2.4 Estimation d'une proportion*

6.3 Détermination de la taille de l'échantillon

*6.3.1 Moyenne*

*6.3.2 Total*

*6.3.3 Proportion*

6.4 Comparaison entre EAS et EPG

## 7. Échantillonnage systematique

7.1 Motivation

7.2 Échantillonnage systématique en tant qu'EAS

*7.2.1 Estimation de la moyenne*

*7.2.2 Estimation du total*

*7.2.3 Estimation d'une proportion*

7.3 Échantillonnage systématique en tant qu'EPG

# Plan de cours

## 8. Sujets choisis

### 8.1 Échantillonnage avec probabilité proportionnelle à la taille

*8.1.1 Méthodes de sélection d'un PPT avec remise*

*8.1.2 Estimation des paramètres*

### 8.2 Échantillonnage à plusieurs degrés

*8.2.1 EAS à 2 degrés*

*8.2.2 Estimation des paramètres*

### 8.3 Échantillonnage à plusieurs phases

*8.3.1 EAS à 2 phases*

*8.3.2 Estimation de la variance d'échantillonnage*

## 8. Sujets choisis (cont.)

### 8.4 Méli-mélo

*8.4.1 Effet de plan*

*8.4.2 Ajustement pour la non-réponse*

*8.4.3 Estimation de la taille d'une population*

*8.4.4 Réponse aléatoire*

*8.4.5 Échantillonnage de Bernoulli*

# Évaluation

- 60 problèmes – 50% [1 par jour, cf. horaire]
- Examen final – 50% [aide-mémoire de 2 pages et calculatrice de poche permis]

Je ne compterai que les 50 meilleurs problèmes. Vous pouvez donc ne soumettre que 50 des problèmes sans nuire à votre évaluation.

Dans le cas d'une maladie, d'un retard, d'un oubli, ou d'un congé mental, vous êtes automatiquement couvert: pas besoin de billet de médecin ou d'explications.

**Mais faites bien attention:** c'est une police d'assurance, pas une carte de "sortie de prison". Si vous en avez déjà profité à 10 reprises et qu'il vous arrive un accident, je n'accorderai pas d'extension.

	Semaine	Sessions		Problèmes (23:59:59 EST)						
		Lundi	Jeudi	L	M	M	J	V		
Module 1: Introduction	01: 10-Jan	1.1, 1.1.1 – 1.1.4		1.2, 1.2.1 – 1.2.5						
	02: 17-Jan	1.2.6	2.1	2.1, 2.2						
Module 2: Échantillonnage aléatoire simple	03: 24-Jan	2.3, 2.3.1		2.3.2 – 2.3.3, 2.4, 2.4.1		1	2	3	4	8
	04: 31-Jan	2.4.2 – 2.4.3	3.1	3.2, 3.2.1		5	6	10	11	12
Module 3: Échantillonnage aléatoire stratifié	05: 07-Fév	3.2.2 – 3.2.3		3.3, 3.3.1		9	13	14	15	17
	06: 14-Fév	3.3.2, 3.4	4.1, 4.2, 4.2.1	4.2.2 – 4.2.4		7	18	22	23	24
Module 4: Estimation par le quotient, par la régression, et par la différence	07: 21-Fév	Congé: 19-Fév au 27-Fév				19	20	21	25	26
	08: 28-Fév	4.2.5 – 4.2.6, 4.3, 4.3.1 – 4.3.3		4.3.4 – 4.3.5, 4.5.1 – 4.5.2		27	28	29	30	41
Module 5: Conception de questionnaires et collecte automatisée	09: 07-Mar	4.4, 4.5.3 – 4.5.4		5.1, 5.1.1 – 5.1.4, 5.2		31	33	35	36	37
	10: 14-Mar	5.2.1 – 5.2.5		6.1, 6.2, 6.2.1 – 6.2.2		32	34	38	39	40
Module 6: Échantillonnage par grappes	11: 21-Mar	6.2.3		6.2.4, 6.3, 6.3.1 – 6.3.3		44	46	47	48	49
	12: 28-Mar	7.1, 7.2, 7.2.1 – 7.2.3, 7.3		8.1, 8.1.1 – 8.1.2, 8.2, 8.2.1 – 8.2.2		50	51	56	57	58
Module 7: Échantillonnage systématique	13: 04-Avr	8.3, 8.3.1 – 8.3.2, 8.4.1 – 8.4.2		8.4.3 – 8.4.5		59	60	61	63	64
	14: 11-Avr					65	66	67	69	70

# Échelle de notation

Pour chaque problème, les notes possibles sont:

- 3: parfait ou presque
- 2: erreur conceptuelle mineure
- 1: deux erreurs conceptuelles mineures ou plus, ou erreur conceptuelle majeure
- 0: tout le reste

# Rôles

## Enseignant:

- préparation et **présentation du matériel** (diapos, exercices, solutions)
- consultation (Zoom, Slack, en personne) – **explications**

## Étudiant.e.s:

- poser des questions sur Slack, en classe, pendant les heures de consultation
- faire des exercices dont vous ne possédez pas les solutions
- accepter de faire des fautes et apprendre de vos erreurs, et
- ne pas vous inquiéter outre-mesure au sujet de votre note.

# Attentes

**Engagement:** l'Université s'attend à ce que vous passiez 135-150 heures sur ce cours [8-10 heures par semaine], dans la salle de classe, en lecture, sur des problèmes, à l'étude, etc. [3 heures de classe, 5 heures de problèmes]

**IMPORTANT:** si vous manquez une session, vous pouvez venir consulter les diapos à mon bureau, mais je ne distribuerai pas de PDF supplémentaires.

Je vous invite à prendre toutes les précautions raisonnables afin d'optimiser votre santé physique et mentale. Prenez-soin de vous, et de vos proches. Aucun problème n'est trop petit (ou trop gros) pour être ignoré. N'hésitez pas à me contacter ou à consulter les ressources suivantes, au besoin:

- <https://www.uottawa.ca/sante/services/sante-mentale>

# Notes

## Intégrité académique

- <https://www.uottawa.ca/administration-et-gouvernance/reglement-scolaire-14-autres-informations-importantes>

## Accommodements académiques

- <https://www.uottawa.ca/administration-et-gouvernance/reglement-academique-i-16-accommodements-academiques>

## Dates importantes

- <https://www.uottawa.ca/dates-importantes-et-echeances-scolaires>

# Notes

**Question:** J'enseigne le cours d'échantillonnage l'an prochain, et on se demande s'il serait préférable de montrer à nos étudiant.e.s comment s'y prendre avec R ou SAS. Vous en êtes où à StatCan, là-dessus?

**Réponse:** Tu me poses une bonne question et la réponse n'est pas simple étant donné que StatCan est en période de lente transition.

On nous encourage fortement à explorer d'autres avenues que SAS, et R est l'alternative préférée. Mais cette transition est lente et parfois impossible. Par exemple, l'Agence est mieux outillée pour SAS (avec des options, telle que la Grille SAS, qui permettent de rouler d'énormes fichiers avec SAS).

# Notes

**Réponse:** (cont.) R n'a pas cette contrepartie et quand on travaille avec de gros fichiers (comme nous, au Recensement), R n'est pas une option parce que c'est juste trop lent (à moins d'avoir des serveurs très puissants). C'est dommage parce qu'il y a des packages dans R qu'on aurait bien voulu utiliser. Mais ceci n'empêche pas R de monter en popularité au sein de StatCan.

Donc pour résumer ma perception des choses : R est le futur mais SAS est encore beaucoup notre présent. Je pense qu'un mix des deux, dans le cours d'échantillonnage, serait une combinaison formidable pour les étudiants. Je pense qu'en période de transition, ce sont les employés qui connaissent les deux qui sont les plus avantagés. Si cela n'est pas possible, j'opterais alors peut-être pour R.

# Notes

Je vais donc utiliser R lors de mes exemples et dans mes solutions, mais je le ferai sans utiliser de librairies spécialisées, telle que `survey`. Libre à vous de vous servir des logiciels qui vous conviennent.

Contactez-moi si vous faites affaire avec le **Service d'accommodements scolaires**.

---

Ne vous laissez intimider ni par le matériel, ni par le professeur!

Profitez-des heures de disponibilité, retroussez vos manches, et bon succès!