

CONCEPTS ET MÉTHODES EN STATISTIQUE

STT 2700

Automne 2012
Professeur : Christian Léger
Tél. : 343-7824

3 crédits
4219, Pavillon André-Aisenstadt
leger@dms.umontreal.ca

Horaire : Mercredi de 11h30 à 12h30 et le vendredi de 10h30 à 12h30 au S-144 du Pavillon Roger-Gaudry ; les travaux pratiques auront lieu le jeudi de 13h30 à 15h30 dans le même local, jusqu'au 21 décembre.

Note : Les cours et les travaux pratiques des 18 et 19 octobre seront interchangeés, alors que le 8 novembre sera une séance de théorie et non pas de travaux pratiques. De plus, les travaux pratiques des dates suivantes seront dans une autre salle : 11 octobre, 1 et 15 novembre (Z-260 du pavillon Claire-McNicoll) et 29 novembre (Z-317, même pavillon).

Des notes de cours et d'autres informations seront disponibles via le site StudiUM du cours : voir <https://studium.umontreal.ca/>

Le cours d'introduction à la statistique (STT 1700) vous a présenté un survol de la discipline. Ceci vous a donné un avant-goût de ce qu'est la statistique sans toutefois pouvoir entrer dans les détails. Le cours de probabilités (MAT 1720) vous a outillé avec les éléments de base de la théorie des probabilités. En particulier, vous avez étudié plusieurs lois de probabilité discrètes et continues, la distribution conjointe de deux variables aléatoires (ou plus), les fonctions génératrices des moments, l'espérance et les moments d'ordre supérieurs, ainsi que des théorèmes sur le comportement des variables aléatoires lorsque la taille de l'échantillon grandit.

Dans ce cours, nous allons étudier d'un point de vue plus formel plusieurs des sujets vus dans le cours d'introduction à la statistique en utilisant les outils de probabilité appris dans le cours MAT 1720. Le niveau mathématique de ce cours sera ainsi beaucoup plus élevé que le cours d'introduction. En plus des outils mathématiques, la simulation jouera parfois un rôle important pour comprendre certains concepts ou voir les limites de certaines méthodes. En effet, même si on réussit à montrer mathématiquement certains résultats sous des hypothèses précises, il est également crucial d'étudier si les résultats continuent de tenir, du moins approximativement, si les hypothèses ne sont pas satisfaites.

Toutefois, via le choix du livre et la façon que ce cours sera donné, la statistique et les données demeureront au coeur de nos préoccupations. Plutôt que d'être une série de recettes *ad hoc*, la statistique est un ensemble de méthodes et de principes qui peuvent être appliqués dans des problèmes très différents. Donc, même si l'éventail des domaines d'application de ces méthodes en statistique qui sera vu dans ce cours est plutôt restreint, il est important de bien mettre l'accent sur les méthodes car celles-ci seront appliquées dans la plupart des autres cours de statistique qui suivront.

Objectifs généraux :

- L'étudiant prendra plaisir à la statistique.
- L'étudiant maîtrisera les aspects essentiels, de même que les limites, des principaux concepts et méthodes de base en statistique, les appliquera dans des situations réelles et saura tirer les conclusions qui s'en suivent.

Objectifs particuliers :

À la fin du cours, l'étudiant devra être en mesure de :

- Estimer des paramètres à l'aide d'une méthode appropriée ;
- Construire un intervalle de confiance et des tests d'hypothèses à l'aide d'une méthode appropriée ;

- Résumer des données à l'aide de graphiques, mesures de localisation et de dispersion ;
- Comparer deux échantillons à l'aide d'une méthode appropriée ;
- Démontrer des résultats de statistique-mathématique ;
- Utiliser un langage de programmation statistique tel R pour analyser des jeux de données ou calculer un estimateur.

Contenu : (Les chapitres font référence au livre de Rice, 2007)

Nous verrons la plupart du contenu des chapitres 6, 8 et 9 et un peu de 10 et 11.

Activités d'enseignement et d'apprentissage :

- Lectures hebdomadaires dans le livre ;
- Cours magistraux avec utilisation d'un projecteur multimédia ainsi que le tableau ;
- Travaux pratiques ;
- Devoirs dont certains nécessiteront un logiciel statistique ; le logiciel R sera privilégié.

Évaluation :

La note sera déterminée à partir de la pondération suivante : 20% Devoirs + 35% Intra (vendredi le 9 novembre de 10h30 à 12h20 au G-815 du Pavillon Roger-Gaudry) + 45% Final (vendredi le 11 janvier de 9h30 à 12h20 dans un local à être déterminé), **si** la moyenne pondérée de l'intra et du final est de 50% ou plus, **sinon** la note des devoirs est plafonnée au minimum de la moyenne des devoirs et de 50% (**menant à un échec**).

Afin d'utiliser efficacement le temps de travail du démonstrateur, il est possible de remettre une copie des devoirs pour **deux** étudiants et seulement un sous-ensemble des questions seront corrigées (le sous-ensemble est déterminé après la remise du devoir).

Au cas où vous vous feriez la réflexion suivante : "C'est quoi ça, on va faire des problèmes qui ne seront même pas corrigés!", voici ma réponse. Comme prof d'université, je fonctionne avec les deux axiomes suivants (Petit Robert : un axiome est une vérité indémontrable mais évidente par quiconque en comprend le sens) : 1) Vous êtes à l'université pour apprendre ; 2) Pour apprendre, il faut faire des devoirs.

Veillez noter que la date limite pour abandonner le cours « sans frais » est le 17 octobre alors « qu'avec frais » (sans être remboursé), la date limite est le 7 décembre. Par la suite, si vous abandonnez, vous aurez un échec. Vous avez également l'obligation de motiver une absence prévisible à l'intra ou au final dès que vous seriez en mesure de constater que vous ne pourriez être présent. Il appartiendra à l'autorité compétente de déterminer si le motif est acceptable.

Plagiat :

L'Université de Montréal a une politique très claire sur le plagiat que vous êtes invités à consulter au www.integrite.umontreal.ca. Elle ne concerne pas que les examens, mais également les devoirs. Ainsi, se mettre en équipe de plusieurs pour résoudre les problèmes du devoir et utiliser exactement le même paragraphe (incluant les fôtes d'otographe!) pour justifier une réponse constitue une forme de plagiat. De la même façon, photocopier un graphique ou utiliser les commandes d'un autre pour produire un graphique constitue également une forme de plagiat.

Bibliographie :

Ouvrage de référence **Obligatoire**

Rice, John A. (2007). *Mathematical Statistics and Data Analysis, 3è édition*. Brooks/Cole, Belmont, CA.

Ouvrage de référence *Recommandé*

Lafaye de Micheaux, Pierre, Drouilhet, Rémy et Liquet, Benoît (2011). *Le logiciel R, Maîtriser le langage – Effectuer des analyses statistiques*. Springer-Verlag, France.

<http://www.springerlink.com/content/978-2-8178-0114-8>

Pour la disponibilité des livres en bibliothèque, contactez le comptoir de prêt

(<http://www.bib.umontreal.ca/nous-joindre/MI.htm>) ou la bibliothécaire Ferroudja Nazef (f.nazef@umontreal.ca)