

PLAN DE COURS

STT1700 - Introduction à la statistique - 3 crédits (Hiver 2012)

Professeur(e): Pierre Lafaye de Micheaux

Coordonnées: Bureau: 4249

Courriel: lafaye@dms.umontreal.ca

Téléphone: 514-343-6607

Web: <http://www.biostatisticien.eu/moodle>

Disponibilités: Jeudi 10:30-12:30

Assistant(es) d'enseignement: Hervé Tchouake Tchuiquep ; ??

Coordonnées: Bureau(x): 4197;??

Courriel(s): therve@dms.umontreal.ca; @dms.umontreal.ca

Disponibilités: Lundi 10:00-12:00 ; ??

1 Description succincte du cours

Description des données. Production de données. Probabilités. Inférence. Intervalles de confiance et tests d'hypothèses. Données de dénombrement. Tableaux de contingence. Régression linéaire simple.

Remarques : Utilisation du logiciel libre et gratuit R.

2 Horaire et format du cours

Cours théoriques : 34h

Travaux pratiques : 20h/Gr.

Activité	Jour	De - À	Du → Au	Nb séances	Local	Immeuble
Cours	Lundi	13:30-15:30	9 janvier	1	N-515	Pav. Roger-Gaudry
			16 janvier → 6 février	4	3110	Fac. Aménagement
			20 février → 27 février	2		
			19 mars → 2 avril	3		
Cours	Jeudi	09:30-10:30	5 janvier → 1 mars	9	N-515	Pav. Roger-Gaudry
			15 mars → 12 avril	5		
TP Gr.1	Lundi	15:30-17:30	9 janvier → 6 février	5	B-4335	Pav. 3200 J-B
			20 février → 27 février	2	1360	A.-Aisenstadt
			19 mars → 2 avril	3	B-4335	Pav. 3200 J-B
TP Gr.2	Lundi	15:30-17:30	9 janvier → 6 février	5	B-4345	Pav. 3200 J-B
			20 février → 27 février	2		
			19 mars → 2 avril	3		

Chaque séance de cours prendra la forme d'un exposé magistral assuré par le professeur. Ces séances seront parsemées de périodes courtes de résolution d'exercices. Les séances de travaux pratiques (TP) seront assurées par l'assistant d'enseignement. Ces séances permettront à l'étudiant de mettre en pratique, à l'aide d'exercices, les notions théoriques vues en cours.

3 Approche pédagogique et accès au site web du cours

La première méthode pédagogique repose sur l'exposé magistral (exposé professoral hebdomadaire). Accompagnés de périodes courtes de résolution d'exercices en classe et de résolution d'exercices à faire à la maison, les exposés magistraux feront acquérir des connaissances théoriques et empiriques et un bagage conceptuel qui constitueront le coeur de notre enseignement. La deuxième méthode est la mise en pratique des apprentissages par les étudiants, d'une part lors des séances de travaux pratiques, et d'autre part aux examens. Cette mise en pratique impliquera plusieurs niveaux cognitifs visés par notre enseignement : compréhension, application des connaissances acquises, analyse et synthèse, sens critique.

L'étudiant pourra également utiliser certains outils mis à disposition sur le site Internet *moodle* associé à ce cours : forum de discussion, notes de cours, etc. Ainsi, l'étudiant qui désire poser des questions sur certains points du cours, des exercices ou des TPs qu'il n'aurait pas bien compris, devra soit les poser pendant les périodes de disponibilités, soit utiliser le forum *moodle* associé à ce cours. Le professeur et les démonstrateurs **ne répondront à aucune question de ce type par courriel ou par téléphone.**

Notez enfin que pour accéder au site *moodle* de ce cours, **vous devez être inscrit** au cours. L'accès au site se fait via l'adresse www.biostatisticien.eu/moodle et nécessite d'utiliser votre identifiant/mot de passe :
Identifiant : votre code d'accès DGTIC (ex : p0123456)
Mot de passe : changeme (à changer à la première connexion)

Remarque : un certain délai (1 semaine) peut être nécessaire entre le moment où vous êtes inscrit au registrariat et où vous avez accès au site *moodle*. Si vous n'êtes pas capable de vous connecter **après ce délai**, merci de m'envoyer un courriel contenant les informations suivantes : Nom, Prénom, Code d'accès DGTIC, Code permanent (3 premières lettre du nom ...), Sigle du cours (STT1700), adresse courriel **obligatoirement** de l'UdeM (...@umontreal.ca).

4 Évaluation des apprentissages

Examen intra 1 : 25% (le lundi 13 février de 14:30 à 16:30 au B-2285, Pav. 3200 J-B).

Examen intra 2 : 25% (le lundi 12 mars de 14:30 à 16:30 au B-2285, Pav. 3200 J-B).

Examen final : 50% (*probablement* le lundi 16 avril de 13:30 à 16:30 au B-2245, Pav. 3200 J-B).

Le détail des épreuves :

Examens en classe : Les examens *intra* et l'examen *final* se dérouleront en classe et porteront sur les exposés professoraux, les séances de travaux pratiques et les exercices à faire à la maison. Les étudiants auront droit à une page aide mémoire (1 page au format letter 8.5x11 pouces, recto-verso où ne figure aucun exercice ou exemple résolu) qui sera donnée (via le site web) par le professeur la semaine précédant l'examen et qui sera également distribuée le jour de l'examen. Il est indispensable que l'étudiant amène sa calculatrice durant les examens. **IMPORTANT :** Notez que la résolution des exercices à faire pour les TPs pourra vous assurer la maîtrise et la rapidité d'exécution de certains calculs simples et répétitifs, mais il s'agit d'une activité insuffisante pour vous assurer une note suffisante à l'examen. Il vous faut donc impérativement travailler les exercices supplémentaires, et avoir compris en profondeur les notes de cours et les exemples qui y sont présentés.

Veillez noter que la date limite pour abandonner le cours «sans frais» est le **19 janvier**. Vous pouvez abandonner le cours «avec frais» (sans être remboursé) mais sans échec inscrit au bulletin jusqu'au **11 mars**. 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 serez en mesure de constater que vous ne pourrez être présent. Il appartiendra à l'autorité compétente de déterminer si le motif est acceptable.

Critères généraux d'évaluation :

- *Contenu* : acquisition des connaissances et de leur application adéquate, capacité d'analyse et de synthèse, cohérence de l'argumentaire, capacité de relier les sujets entre eux, capacité d'évaluer, de faire preuve d'un sens critique appuyé par les connaissances ;
- *Forme* : respect des directives afférentes aux diverses épreuves, respect des règles de présentation matérielle (page de couverture, prescriptions de longueur, etc...);
- *Qualité rédactionnelle* : orthographe, grammaire, ponctuation, clarté dans l'expression ;
- *Ponctualité* : le respect des délais prescrits.

Remarques importantes :

- La présence en classe est **très fortement recommandée** pour bien réussir ce cours !
- Tout plagiat¹ entraînera de graves sanctions.
- Seules les notes des deux intras seront affichées sur le site du cours.

5 Objectifs

Le monde qui nous entoure est rempli de phénomènes aléatoires, que ce soit la durée de vie d'une personne, le nombre de processus qui roulent en même temps sur un ordinateur à un moment donné, le nombre de bières consommées par des étudiants un vendredi soir, etc. Certaines variables peuvent avoir un effet sur ces phénomènes. Par exemple, la consommation de bières est-elle la même chez les étudiantes que chez les étudiants ? La statistique est la science qui permet de prendre des décisions face à l'incertitude. Pour y arriver, il faut recueillir des données provenant du phénomène aléatoire et analyser ces données à partir d'un modèle stochastique qui représente le phénomène. La science de la statistique étudie les façons les plus efficaces d'effectuer la cueillette, la modélisation et l'analyse des données afin d'extraire le maximum d'information de celles-ci. Ce cours offre un survol des principales questions auxquelles la statistique tente de répondre. L'accent sera mis davantage sur la compréhension des concepts statistiques que sur le développement d'un livre de recettes statistiques. Pour ceux dont ce sera le seul cours de statistique, cette introduction vous donnera une idée des principaux enjeux de la discipline et vous permettra d'être de meilleurs consommateurs de statistiques publiées dans les journaux ou les revues scientifiques. Les notions vues dans ce premier cours seront approfondies dans plusieurs cours subséquents. Cette introduction permettra à ceux qui poursuivront en statistique de se faire une bonne idée globale de la discipline. Des notes de cours seront disponibles sur ma page web. Il peut être très utile de les télécharger, de les imprimer et de les amener en cours car je suivrai leur contenu religieusement.

6 Plan du cours

6.1 Statistiques descriptives

Variabes et distributions. Types d'une variable. Diagramme en tuyaux d'orgues. Diagramme circulaire. Diagramme en bâtons. Histogramme. Tiges et feuilles. Graphique temporel. *Box plot*. Mesures de tendance centrale. Mesures de dispersion. Couples de variables. Distribution conjointe, marginale et conditionnelle. Indépendance. Nuage de points. Covariance, corrélation et droite des moindres carrés.

6.2 Probabilités et outils biostatistiques

Définitions. Axiomes et propriétés. Opérations ensemblistes. Événements impossibles, certain, disjoints. Probabilité conditionnelle. Théorème de Bayes. Événements indépendants. Vrais positifs, faux positifs, faux négatifs, vrais négatifs. Sensibilité et spécificité. Valeurs prédictives positive et négative, risque relatif, prévalence. Espace ROC.

¹Voir <http://www.fas.umontreal.ca/plagiat>

6.3 Variables aléatoires

Définition. Variables aléatoires discrètes. Fonction de masse. Fonction de répartition. Espérance et variance d'une (somme, produit de) variable(s) aléatoire(s). Loi Binomiale. Loi multinomiale.

6.4 Lois continues

Variables aléatoires continues. Fonction de densité et fonction de répartition. Espérance et variance. Loi normale et propriétés. Théorème limite central. Approximation normale de la loi binomiale. Lois du chi-deux. Lois de Student.

6.5 Estimation ponctuelle

Population et échantillon. Échantillon aléatoire simple. Statistique et estimateur. Estimateur sans biais. Estimateur sans biais d'une moyenne, d'une variance et d'une proportion. Variance d'un estimateur.

6.6 Intervalles de confiance

Définition et interprétation. Intervalle de confiance pour une moyenne, pour la différence de deux moyennes. Intervalle de confiance pour une proportion, pour la différence de deux proportions.

6.7 Tests d'hypothèses pour de grands échantillons

Définitions. Composantes d'un test d'hypothèse : hypothèse nulle et alternative, région critique, erreurs de première et seconde espèce, valeur- p , statistique. Test unilatéral, bilatéral. Puissance d'un test. Tests d'hypothèses pour une moyenne, pour la différence de deux moyennes. Courbes ROC. Tests d'hypothèses pour donnée appariées. Tests d'hypothèses pour une proportion, pour la différence de deux proportions. Randomisation.

6.8 Inférence pour de petits échantillons

Intervalles de confiance et tests d'hypothèses pour une moyenne, pour la différence de deux moyennes.

6.9 Tests du khi-deux

Tests d'ajustement. Tests d'indépendance. Causalité et dépendance.

7 Livre support obligatoire

Mendenhall, W., Beaver, R.J., Beaver, B.M. et Ahmed, S.E. (2011). *Introduction to Probability and Statistics*, Second Canadian Edition. Nelson, Canada.

8 Livres recommandés

Ces ouvrages sont disponibles à la bibliothèque Maths-Info : <http://www.bib.umontreal.ca/MI>

- Lafaye de Micheaux, P., Drouilhet, R. et Liquet, B. (2010). *Le logiciel R - Maîtriser le langage - Effectuer des analyses statistiques*, Springer.
<http://www.springerlink.com/content/978-2-8178-0114-8>
- Moore, D.S. et McCabe, J.A. (2006). *Introduction to the Practice of Statistics*, 5th édition. W. H. Freeman and Company, New York.

Dernière mise à jour: 3 janvier 2012.