

Méthodes statistiques en bioinformatique

Groupe 40

Mercredi, de 17h30 à 20h30 PK-4323 (cours)

Responsable(s) du cours

Nom du coordonnateur : WILLEMS, Matthieu**Nom de l'enseignant :** WILLEMS, Matthieu**Local :** PK-4660**Téléphone :** (514) 987-3000 #4803**Courriel :** willems.matthieu@uqam.ca

Auxiliaire : TAHIRI, Nadia

Autre enseignant : LORD, Étienne

Description du cours

Fournir les bases statistiques pour énoncer les hypothèses statistiques et exécuter les tests qui s'appliquent à des problèmes de bioinformatiques.

Préalables académiques :

Rappel des principes de base en probabilité et en statistique. Inférence statistique classique et bayésienne. Processus stochastiques : promenades aléatoires, processus de Poisson, chaînes de Markov. Algorithme BLAST ("Basic Local Alignment Search Tool"). Modèles d'évolution.

Objectifs du cours

- 1) Appliquer et comprendre les bases statistiques en bioinformatique pour énoncer des hypothèses.
- 2) Apprendre le langage R et développer des fonctions d'analyse.

Modalités d'évaluation

Évaluations	Dates	Pondération
Examen de mi-session	1er novembre 2017	30%
T.P. (devoir)	06 décembre 2017	30%
Projet de session	20 décembre 2017	30%
Présentation	13 décembre 2017	10%

Travail pratique (T.P.)

Le T.P. sera une suite de petits exercices qui touchent directement la matière enseignée les semaines précédentes. L'énoncé sera distribué trois semaines avant la date de remise, soit le **15 novembre 2017**. Il s'agit d'un travail **individuel** à remettre au plus tard le **06 décembre 2017**.

Projet de session

Le projet de session est à faire en **équipe de deux ou trois étudiants**. Les équipes devront définir et envoyer un énoncé de projet au plus tard le **15 novembre 2017**. Les étudiants devront rencontrer le professeur responsable de leur projet au moins une fois durant la session pour discuter l'avancement de leur projet.

Le projet réalisé sera livré de deux façons : un exposé en classe de **10-15 minutes** (5 minutes par personne) le **13 décembre 2017**, puis un rapport écrit sous forme d'article à remettre au plus tard le **20 décembre 2017**. Vous devez remettre toutes les annexes que vous avez générées dans un dossier compressé (.zip ou .tar.gz). N'oubliez pas de bien les identifier et de les commenter.

Le rapport principal comportera une dizaine de pages et comprendra les sections suivantes :

- 1) Résumé de votre étude (1/2 page)
- 2) Introduction avec les références pertinentes (2 pages)
- 3) Méthodologie (2 à 3 pages)
- 4) Résultats et discussions (2 à 3 pages)
- 5) Conclusion (1 page)
- 6) Références en écartant les références de *Wikipedia* (1 page maximum)

Exemple de travaux admissibles :

- 1) Révision critique d'un outil bioinformatique (ou package R) : description de l'application et des jeux de données, de la problématique, des tests d'hypothèses et de la méthode statistique. Inclusion d'un exemple pratique.
- 2) Révision des techniques statistiques utilisées dans un sujet d'intérêt en bioinformatique par exemple : la phylogénie, la variation génétique, l'épidémiologie, l'écologie, etc.

Politique d'absence aux examens

L'autorisation de reprendre un examen en cas d'absence est de caractère exceptionnel. Pour obtenir un tel privilège, l'étudiant-e doit avoir des motifs sérieux et bien justifiés.

Il est de la responsabilité de l'étudiant-e de ne pas s'inscrire à des cours qui sont en conflit d'horaire, tant en ce qui concerne les séances de cours ou d'exercices que les examens. **De tels conflits d'horaire ne constituent pas un motif justifiant une demande d'examen de reprise.**

Dans le cas d'une absence pour raison médicale, l'étudiant-e doit joindre un certificat médical original et signé par le médecin décrivant la raison de l'absence à l'examen. Les dates d'invalidité doivent être clairement indiquées sur le certificat. Une vérification de la validité du certificat pourrait être faite. Dans le cas d'une absence pour une raison non médicale, l'étudiant-e doit fournir les documents originaux expliquant et justifiant l'absence à l'examen – par exemple, lettre de la Cour en cas de participation à un jury, copie du certificat de décès en cas de décès d'un proche, etc. Toute demande incomplète sera refusée. Si la direction du programme d'études de l'étudiant-e constate qu'un étudiant a un comportement récurrent d'absence aux examens, l'étudiant-e peut se voir refuser une reprise d'examen.

L'étudiant-e absent-e lors d'un examen doit, dans les cinq (5) jours ouvrables suivant la date de l'examen, présenter une demande de reprise en utilisant le formulaire prévu, disponible sur le site Web du département à l'adresse suivante : <http://info.uqam.ca/politiques/>

L'étudiant-e doit déposer le formulaire dûment complété au secrétariat de la direction de son programme d'études : PK-3150 pour les programmes de premier cycle, PK-4150 pour les programmes de cycles supérieurs. Pour plus de détails sur la politique d'absence aux examens du Département d'informatique, consultez le site web suivant : <http://info.uqam.ca/politiques>

Calendrier détaillé du cours

Dans le tableau ci-dessous, les noms des enseignant-e-s sont indiqués par leurs initiales.

Semaine	Date	Activité
1	06 septembre	Présentation du cours, introduction aux statistiques, concept de probabilité, méthode scientifique, mesures de tendance centrale, mesures de dispersion et de position (MW)
2	13 septembre	Introduction au langage R avec RStudio, en R : Introduction aux structures de données : vecteurs, « data.frame » et matrices, lecture de fichiers, calculs simples et sorties graphiques (EL)
3	20 septembre	Distributions de probabilité (loi de Poisson, loi normale, loi de Student), tests d'hypothèses, notion de degrés de liberté (MW)
4	27 septembre	Langage R avec RStudio : Opérations sur les listes, structures de contrôle et fonctions, tests statistiques en R (EL)
5	04 octobre	Analyse de variance (ANOVA), test de normalité, analyse factorielle (ANOVA à plusieurs facteurs) (MW)
6	11 octobre	Régressions simples et multiples (MW)
7	18 octobre	Langage R avec RStudio : ANOVA et régression dans R (EL)
8	25 octobre	Chaînes de Markov et théorie de l'échantillonnage, théorème de Bayes et p-value (MW)
9	1er novembre	Examen de mi-session (MW)
10	08 novembre	Langage R avec RStudio : Analyse Bayésienne et analyse par composantes principales (PCoA) (EL)
11	15 novembre	Langage R avec RStudio : Introduction aux réseaux (Igraph), analyse et statistiques sur les réseaux, analyse des arbres phylogénétiques, distribution du T.P. (EL)
12	22 novembre	Simulations Monte Carlo, tests de permutation, bootstrap, concept de Likelihood (MW)
13	29 novembre	Langage R avec RStudio : Simulations de Monte Carlo, échantillonnage aléatoire, bootstrap, réalisation de rapports et présentations avec Knitr (NT)
14	06 décembre	Langage R avec RStudio : Classification hiérarchique (hclust) et par partitionnement (k-means), remise du T.P. (NT)
15	13 décembre	Présentations orales

Intégrité académique

PLAGIAT Règlement no 18 sur les infractions de nature académique. (extraits)

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes;
- l'utilisation totale ou partielle du texte d'autrui en la faisant passer pour sien ou sans indication de référence;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant;
- l'obtention par vol, manoeuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-crédation, un rapport de stage ou un rapport de recherche;
- Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, veuillez consulter les sites suivants : <http://www.sciences.uqam.ca/etudiants/integrite-academique.html> et <http://www.bibliotheques.uqam.ca/recherche/plagiat/index.html>

Médiagraphie

Logiciel

RStudio: <https://www.rstudio.com/products/rstudio/download/> LO

Livres et aide mémoire

Kauffman, M. (2009) AIDE MÉMOIRE R Référence des fonctions de R les plus courantes. (https://cran.r-project.org/doc/contrib/Kauffmann_aide_memoire_R.pdf) VO

Krijnen, W. P. (2009). Applied statistics for bioinformatics using R. Institute for Life Science and Technology, Hanze University. (<https://cran.r-project.org/doc/contrib/Krijnen-IntroBioInfStatistics.pdf>) VO

Laude, H. (2016). Data Scientist et langage R - Guide d'autoformation à l'exploitation des Big Data. ENI. VR

Paradis, E. (2005) R pour les débutants (https://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf) VO