
BIF7105

Méthodes statistiques en bioinformatique

Plan de cours

Responsable(s) du cours

Coordination : MAKARENKOV, Vladimir
PK-4815
poste 3870
makarenkov.vladimir@uqam.ca

Enseignant(s) :

TAHIRI, Nadia
PK-4660
poste 4803
tahiri.nadia@uqam.ca

WILLEMS, Matthieu
PK-4660
poste 4803
willems.matthieu@uqam.ca

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.

Objectif 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 DE REMISE	PONDÉRATION
Examen de mi-session	30 octobre	30%
T.P. (devoir)	27 novembre	30%
Projet de session	18 décembre	30%
Présentation	11 décembre	10%

TRAVAUX PRATIQUES (T.P.)

Le T.P. sera une suite de petits exercices qui touchent directement la matière enseignée les semaines précédentes. Les énoncés seront donnés trois semaines avant la date de remise, soit le 6 NOVEMBRE 2019. Les travaux pratiques doivent être faits individuellement et remis sous format papier.

PROJET DE SESSION

Les projets de session sont à faire en ÉQUIPE DE DEUX OU TROIS ÉTUDIANTS. Les équipes devront définir et envoyer un énoncé de projet au plus tard le 6 novembre 2019. Chaque équipe devra rencontrer le professeur responsable de son 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 11 décembre 2019, puis un rapport écrit sous forme d'article à remettre le 18 décembre 2019. 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 les commenter.

Le rapport principal sera de 10 pages maximum et comprendra les sections suivantes :

- 1) Un résumé de votre étude (1/2 page)
- 2) Une 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 p.ex. la phylogénie, la variation génétique, l'épidémiologie, l'écologie, etc.

Calendrier

Semaine	Date	Activité
1	4 septembre	Présentation du cours, introduction aux statistiques, concept de probabilité, méthode scientifique, mesures de tendance centrale, mesures de dispersion et de position (MW)
2	11 septembre	Distributions de probabilité (loi de Poisson, loi normale, loi de Student), tests d'hypothèses, notion de degrés de liberté (MW)
3	18 septembre	Analyse de variance (ANOVA), test de normalité, analyse factorielle (ANOVA à plusieurs facteurs) (MW)
4	25 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 (NT)
5	2 octobre	Langage R avec RStudio : Opérations sur les listes, structures de contrôle et fonctions, tests statistiques en R (NT)
6	9 octobre	Régressions simples et multiples (MW)
7	16 octobre	Langage R avec RStudio : ANOVA et régression dans R (NT)
8	23 octobre	Chaînes de Markov et théorie de l'échantillonnage, théorème de Bayes et p-value (MW)
9	30 octobre	Examen de mi-session (MW)
10	6 novembre	Langage R avec RStudio : Analyse Bayésienne et analyse par composantes principales (PCoA) (NT)
11	13 novembre	Simulations Monte Carlo, tests de permutation, bootstrap, concept de Likelihood (MW)
12	20 novembre	Langage R avec RStudio : Introduction aux réseaux (Igraph), analyse et statistiques sur les réseaux, analyse des arbres phylogénétiques, distribution du T.P. (NT)

13	27 novembre	Langage R avec RStudio : Simulations de Monte Carlo, échantillonnage aléatoire, bootstrap, réalisation de rapports et présentations avec Knitr (NT)
14	4 décembre	Langage R avec RStudio : Classification hiérarchique (hclust) et par partitionnement (k-means), remise du T.P. (NT)
15	11 décembre	Présentations orales (NT)

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

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 : 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 : info.uqam.ca/politiques.

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-créditation, 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 : www.sciences.uqam.ca/etudiants/integrite-academique.html et www.bibliotheques.uqam.ca/plagiat/le-plagiat-liens-rapides.

Politique no 16 visant à prévenir et combattre le sexisme et les violences à caractère sexuel

Pour consulter la politique no 16 :

instances.uqam.ca/wp-content/uploads/sites/47/2018/05/Politique_no_16.pdf

Services offerts :

Pour obtenir de l'aide, faire une divulgation ou une plainte :
Bureau d'intervention et de prévention en matière de harcèlement
514 987-3000, poste 0886

Pour la liste des services offerts en matière de violence sexuelle à l'UQAM et à l'extérieur de l'UQAM : harcelement.uqam.ca

CALACS Trêve pour Elles – point de services UQAM :

514 987-0348
calacs@uqam.ca
trevepourelles.org

Soutien psychologique (Services à la vie étudiante) :

514 987-3185
Local DS-2110

Service de la prévention et de la sécurité : 514 987-3131

Les étudiants qui ont une lettre signée de leur conseillère ou conseiller de l'Accueil et de soutien aux étudiants en situation de handicap (ASESH), dans laquelle il est fait état de leur inscription au ASESH à titre d'étudiant(e) en situation de handicap, sont invités à remettre ce document à leurs professeur(e)s et chargé(e)s de cours dès le début de la session afin que les aménagements dans le respect des exigences académiques soient déterminées de concert avec chacun des professeur(e)s et chargé(e)s de cours. Les étudiants qui ont une déficience et qui ne seraient pas inscrits au ASESH sont priés de se présenter au AB-2300.

Étudiants avant une déficience de type visuelle, auditive, motrice, trouble d'apprentissage, trouble envahissant du développement et trouble de santé mentale :

Les étudiant(e)s qui ont une lettre d'*Attestation des mesures d'aménagements académiques* obtenue auprès d'une conseillère, d'un conseiller de l'ACCUEIL ET SOUTIEN AUX ÉTUDIANTS EN SITUATION DE HANDICAP (ASESH) doivent rencontrer leurs enseignant(e)s au début de la session afin que des mesures d'aménagement en classe ou lors des évaluations puissent être mises en place. Ceux et celles qui ont une déficience ou une incapacité mais qui n'ont pas cette lettre doivent contacter l'ASESH au (514) 987-3148 ou se présenter au AB-2300 le plus tôt possible.